International Journal on Advances in Internet Technology

IARIA

Includes a special issue on Wireless Mesh Networks

2010 vol. 3 nr. 1&2

The *International Journal on Advances in Internet Technology* is published by IARIA.
ISSN: 1942-2652
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Internet Technology
Volume 3, Number 1 & 2, 2010

Communication Theory, QoS and Reliability

- Adrian Andronache, University of Luxembourg, Luxembourg
- Shingo Ata, Osaka City University, Japan
- Eugen Borcoci, University "Politehnica" of Bucharest (UPB), Romania
- Michel Diaz, LAAS, France
- Michael Menth, University of Wuerzburg, Germany
- Michal Pioro, University of Warsaw, Poland
- Joel Rodriques, University of Beira Interior, Portugal
- Zary Segall, University of Maryland, USA

Ubiquitous Systems and Technologies

- Sergey Balandin, Nokia, Finland
- Matthias Bohmer, Munster University of Applied Sciences, Germany
- David Esteban Ines, Nara Institute of Science and Technology, Japan
- Dominic Greenwood, Whitestein Technologies AG, Switzerland
- Arthur Herzog, Technische Universitat Darmstadt, Germany
- Malohat Ibrohimova, Delft University of Technology, The Netherlands
- Reinhard Klemm, Avaya Labs Research-Basking Ridge, USA
- Joseph A. Meloche, University of Wollongong, Australia
- Ali Miri, University of Ottawa, Canada
- Vladimir Stantchev, Berlin Institute of Technology, Germany
- Said Tazi, LAAS-CNRS, Universite Toulouse 1, France

Systems and Network Communications

- Eugen Borcoci, University 'Politechncia' Bucharest, Romania
- Anne-Marie Bosneag, Ericsson Ireland Research Centre, Ireland
- Jan de Meer, smartspace®lab.eu GmbH, Germany
- Michel Diaz, LAAS, France
- Tarek El-Bawab, Jackson State University, USA
- Mario Freire, University of Beria Interior, Portugal / IEEE Portugal Chapter
- Sorin Georgescu, Ericsson Research - Montreal, Canada
- Huaqun Guo, Institute for Infocomm Research, A*STAR, Singapore
- Jong-Hyouk Lee, INRIA, France
- Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway
- Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
- Sjouke Mauw, University of Luxembourg, Luxembourg
- Reijo Savola, VTT, Finland

Future Internet

- Thomas Michal Bohnert, SAP Research, Switzerland
- Fernando Boronat, Integrated Management Coastal Research Institute, Spain

- Chin-Chen Chang, Feng Chia University - Chiayi, Taiwan
- Bill Grosky, University of Michigan-Dearborn, USA
- Sethuraman (Panch) Panchanathan, Arizona State University - Tempe, USA
- Wei Qu, Siemens Medical Solutions - Hoffman Estates, USA
- Thomas C. Schmidt, University of Applied Sciences – Hamburg, Germany

Challenges in Internet

- Olivier Audouin, Alcatel-Lucent Bell Labs - Nozay, France
- Eugen Borcoci, University "Politehnica" Bucharest, Romania
- Evangelos Kranakis, Carleton University, Canada
- Shawn McKee, University of Michigan, USA
- Yong Man Ro, Information and Communication University - Daejon, South Korea
- Francis Rousseaux, IRCAM, France
- Zhichen Xu, Yahoo! Inc., USA

Advanced P2P Systems

- Nikos Antonopoulos, University of Surrey, UK
- Filip De Turck, Ghent University – IBBT, Belgium
- Anders Fongen, Norwegian Defence Research Establishment, Norway
- Stephen Jarvis, University of Warwick, UK
- Yevgeni Koucheryavy, Tampere University of Technology, Finland
- Maozhen Li, Brunel University, UK
- Jorge Sa Silva, University of Coimbra, Portugal
- Lisandro Zambenedetti Granville, Federal University of Rio Grande do Sul, Brazil

Additional reviews

- Kaan Bür, Lund University, Sweden

International Journal on Advances in Internet Technology
Volume 3, Numbers 1 & 2, 2010

CONTENTS

Section from page 1 to page 87 is a special issue on Wireless Mesh Networks.

Collision Reduction in Cognitive Wireless Local Area Network over Fibre **1 - 12**
Haoming Li, The University of British Columbia, Canada
Alireza Attar, The University of British Columbia, Canada
Victor C. M. Leung, The University of British Columbia, Canada
Qixiang Pang, General Dynamics Canada, Canada

On Optimization of Wireless Mesh Networks using Genetic Algorithms **13 - 28**
Rastin Pries, University of Würzburg, Germany
Dirk Staehle, University of Würzburg, Germany
Barbara Staehle, University of Würzburg, Germany
Phuoc Tran-Gia, University of Würzburg, Germany

The design and implementation of the Cyclic Scheduling Algorithm: A multi-channel MAC protocol **29 - 42**
Mthulisi Velempini, University of Cape Town, South Africa
Mqhele. E. Dlodlo, University of Cape Town, South Africa

Service area deployment of IEEE 802.16j wireless relay networks: service area coverage, energy consumption, and resource utilization efficiency **43 - 52**
Shoichi Takemori, Osaka University, Japan
Go Hasegawa, Osaka University, Japan
Yoshiaki Taniguchi, Osaka University, Japan
Hirotaka Nakano, Osaka University, Japan

MIMO Capacity of Wireless Mesh Networks **53 - 64**
Sebastian Max, RWTH Aachen University, Germany
Bernhard Walke, RWTH Aachen University, Germany

Behind-the-Scenes of IEEE 802.11a based Multi-Radio Mesh Networks: A Measurement driven Evaluation of Inter-Channel Interference **65 - 76**
Sebastian Robitzsch, University College Dublin, Ireland
John Fitzpatrick, University College Dublin, Ireland
Seán Murphy, University College Dublin, Ireland
Liam Murphy, University College Dublin, Ireland

IEEE 802.16 Wireless Mesh Networks Capacity Assessment Using Collision Domains **77 - 87**
Rafal Krenz, Poznan University of Technology, Poland

Simulation of Multihop Energy-Aware Routing Protocols in Wireless Sensor Networks **88 - 103**
Adrian Fr. Kacsó, University of Siegen, Germany

TeraPaths: End-to-End Network Resource Scheduling in High-Impact Network Domains **104 - 117**
Dimitrios Katramatos, Brookhaven National Laboratory, USA
Xin Liu, Brookhaven National Laboratory, USA
Kunal Shroff, Brookhaven National Laboratory, USA
Dantong Yu, Brookhaven National Laboratory, USA
Shawn McKee, University of Michigan, USA
Thomas Robertazzi, Stony Brook University, USA

State of the Art and Innovative Communications and Networking Solutions for a Reliable and Efficient Interplanetary Internet **118 - 127**
Giuseppe Araniti, University "Mediterranea" of Reggio Calabria, Italy
Igor Bisio, University of Genoa, Italy
Mauro De Sanctis, University of Rome "Tor Vergata", Italy

Circuit Analysis and Simulations through Internet **128 - 136**
Jiří Hospodka, Czech Technical University in Prague, Czech Republic
Jan Bičák, ASICentrum, Czech Republic

Performance Analysis of Scheduling and Dropping Policies in Vehicular Delay-Tolerant Networks **137 - 145**
Vasco N. G. J. Soares, University of Beira Interior, Portugal
Farid Farahmand, Sonoma State University, USA
Joel J. P. C. Rodrigues, University of Beira Interior, Portugal

Cost-Optimal and Cost-Aware Tree-Based Explicit Multicast Routing **146 - 158**
Miklós Molnár, IRISA, France

Comparison of Packet Switch Architectures and Pacing Algorithms for Very Small Optical RAM **159 - 169**
Onur Alparslan, Osaka University, Japan
Shin'ichi Arakawa, Osaka University, Japan
Masayuki Murata, Osaka University, Japan

On Designing Semantic Lexicon-Based Architectures for Web Information Retrieval **170 - 183**
Vincenzo Di Lecce, Politecnico di Bari, Italy
Marco Calabrese, Politecnico di Bari, Italy
Domenico Soldo, myHermes S.r.l., Italy

Editorial

Wireless Mesh Networks (WMNs) have attracted significant interest from both academia and industry because of the challenging research questions they pose and the novel applications and services they can provide. In Wireless Mesh Networks, mesh routers relay packets wirelessly towards the destination or an internet gateway. Compared to Ad-Hoc or Sensor Networks, routers in Wireless Mesh Networks have constant power supply and very limited mobility, which limits the problem scope and allows more practical solutions to be developed. As a result, several vendors are actively developing products based on mesh technology The low cost of devices combined with the distributed and self-organised management possibilities enable rapid deployment and simple maintenance of an operational network. WMN based networks provide novel solutions for wireless last mile access, wireless broadband home networks, wireless enterprise backbone networks, user community wireless networks or wireless municipality networks. Also, such mesh networks enable interesting services such as content sharing, multicast video delivery, sensor network backhaul, and vehicular network infrastructure support in addition to wireless Internet access.

This Special Issue focuses on recent advancements in the area of Wireless Mesh Networks. It brings together a set of articles that provide a good balance between experimental and more theoretical aspects of various issues related to Wireless Mesh Networks. For this issue, five contributions have been selected based on a rigorous review process, which involved mostly three but at minimum two reviews per paper. Every review was cross-checked by the editor in order to ensure a high quality.

The first paper in this special issue is Collision Reduction in Cognitive Wireless Local Area Network over Fibre by Haoming Li, Alireza Attar, Victor Leung and Qixiang Pang. The work proposes to use remote antenna units (RAU) connected to a centralized cognitive access point using optical fiber. The architecture supports multiple independent channels at each RAU and transmitter and receiver diversity. Due to the reduction of collisions, substantial improvement in throughput and reduction of packet loss under dynamic traffic conditions is achievable.

The second paper in this special issue is On Optimization of Wireless Mesh Networks using Genetic Algorithms by Rastin Pries, Dirk Staehle, Barabara Staehle and Phuoc Tran-Gia. The work investigates the usage of genetic algorithms to optimize large wireless mesh network deployments. Based on the collision domain concept, different fitness functions are evaluated which allow optimizing the network for different purposes. While genetic algorithms might solve the complex structure of WMNs in small computation time, the parameters of the genetic algorithm have to be carefully chosen and adapted to the given topology.

The third paper in this special issue is A multiple channel selection and coordination MAC Scheme by Mthulisi Velempini and Mqhele. E. Dlodlo. While multichannel MAC protocols based on dedicated control channel approach are interesting solutions to increase capacity, the common control channel could be the bottleneck. In this work, authors investigate this in more detail and propose a new cyclic scheduling scheme, which schedules data transmission in phases. As a result, signalling overhead in the control channel can be reduced which leads to higher capacity for data transmissions.

The forth paper in this special issue is Service area deployment of IEEE 802.16j wireless relay networks: service area coverage, energy consumption, and resource utilization efficiency by Shoichi Takemori, Go Hasegawa, Yoshiaki Taniguchi and Hirotaka Nakano. The work presents three methods which allow

determining the service area of IEEE 802.16j wireless relay networks. The methods differ in the amount of knowledge required regarding neighboring nodes. Based on extensive simulation experiments authors determine different parameters such as coverage ratio, service area overlap characteristics, energy consumption, and utilization efficiency of wireless network resources.

The fifth paper in this special issue is Capacity Improvement by MIMO in Wireless Mesh Networks by Sebastian Max and Bernhard Walke. The work combines a realistic MIMO model with a capacity calculation framework to estimate the capacity of a meshed deployment of MIMO based mesh routers. The results show that while the link capacity increase of MIMO cannot be exploited fully in the mesh context, WMNs still benefit from the MIMO gain. However, it remains unclear what real gains can be achieved if a realistic MAC protocol is run on top of such MIMO system.

The sixth paper in this special issue is Behind-the-Scenes of IEEE 802.11a based Multi-Radio Mesh Networks: A Measurement driven Evaluation of Inter-Channel Interference by Sebastian Robitzsch, John Fitzpatrick, Seán Murphy and Liam Murphy. The work explores interference problems of multi-radio meshed systems which occur if two or more radios within close vicinity operate in parallel on adjacent channels (ACI). Based on extensive measurement results, the paper shows that such ACI may severely limit the performance of multi-radio mesh network deployments. Authors conclude that achievable performance depends on many effects such as transmission power, modulation coding scheme, channel separation and physical layer effects such as carrier sensing, retransmissions and packet distortion.

The seventh paper in this special issue is IEEE 802.16 Wireless Mesh Networks Capacity Assessment Using Collision Domains by Rafal Krenz. The work applies the concept of collision domain modelling to IEEE 802.16 based WMNs. While the paper considers only a simple chain topology, the method can be extended to arbitrary topologies and real world impairments such as interference or fading can be easily incorporated.

Finally, we would like to thank the authors who have submitted their work to this Special Issue. We also would like to thank the reviewers for their invaluable contributions to the review process. We hope that you will enjoy the contents of this special issue and find it useful for discovering more about the challenges and approaches for Wireless Mesh Networks, and that you will be inspired to contribute to IARIA's conferences that include topics relevant to this journal.

Andreas J. Kassler, Editor-in-Chief

Collision Reduction in Cognitive Wireless Local Area Network over Fibre

Haoming Li, Alireza Attar, Victor C. M. Leung
Department of Electrical and Computer Engineering
The University of British Columbia
2332 Main Mall, Vancouver, BC, Canada
{hlih, attar, vleung}@ece.ubc.ca

Qixiang Pang
General Dynamics Canada
Calgary, AB, Canada
Kevin.Pang@gdcanada.com

***Abstract*—Cognitive wireless local area network over fiber (CWLANoF), which employs remote antenna units (RAUs) connected to a central cognitive access point through optical fibres, can provide a cost-effective and efficient architecture for devices to equally share the industrial, scientific, and medical band by taking advantage of cognitive radio capabilities. Based on the CWLANoF architecture, we propose two methods to reduce collisions among stations, with multiple independent channels operating at each RAU, and transmitter and receiver diversity through cooperation of adjacent RAUs. Multi-channel-operation method is enabled by wide-band optical fibres and diversity method is enabled by distributed RAUs in the CWLANoF architecture. Extensive simulations show substantial improvements in Transmission Control Protocol throughput and packet error rate reduction of constant-bit-rate traffic streams, especially under dynamic traffic conditions.**

Keywords- cognitive radio; radio over fibre; WLAN; diversity; capture effect

I. INTRODUCTION

Wireless local area networks (WLANs) are widely used for connecting computing equipment in homes and offices to the Internet. Cognitive wireless local area network over fiber (CWLANoF) is a new architecture [1] that applies advanced cognitive radio [2] and broadband radio over fiber (RoF) [3] technologies to infrastructure-based IEEE 802.11 WLAN Extended Service Sets (ESSs) comprised of multiple access points (APs), each forming its own Basic Service Set (BSS). This architecture is intended to provide centralized radio resource management and access control through cooperative spectrum sensing.

Since WLANs share the industrial, scientific, and medical (ISM) band with other independently-operated license-free devices such as Bluetooth radios and microwave ovens, they must tolerate interference from these devices. Cognitive radio techniques have been recently proposed for secondary users to exploit spectrum holes left unused in licensed frequency bands by primary users of the allocated spectrum. In this paper, we exploit cognitive radio techniques in the CWLANoF architecture for equal access in the license-free ISM band to enhance system performance via spectrum sensing, interference avoidance and coexistence.

Successful simultaneous transmissions of multiple WLAN channels over low-cost multi-mode optical fibres [4]-[7] and clarification of WLAN medium access control (MAC) operation in RoF structures [8]-[10] motivate the proposal of CWLANoF as an architecture that offers huge potentials to increase system capacity and improve quality-of-service. The ever-decreasing cost of optical fibers and wavelength-division multiplexing components has resulted in commercial fiber-based indoor wireless networks being deployed to penetrate large buildings such as stadiums [11], hospitals [12], business buildings and shopping malls [13]. It would be expensive to cover these buildings with cable-based networks due to the ever-increasing cable cost. It is also difficult to monitor and manage the radio environment within such large buildings if antennas cannot be efficiently coordinated. The success of these commercial indoor wireless networks further demonstrated potential markets for CWLANoF networks.

In this work, we focus on how to reduce access collisions in a WLAN through methods made possible by the CWLANoF architecture, which would be difficult if not impossible to realize in a conventional WLAN.

In a conventional WLAN, each AP performs carrier sensing independently and only over the channel it operates on. In contrast, CWLANoF applies cognitive radio techniques to enable a WLAN ESS to more efficiently utilize the ISM band. The centralized architecture of CWLANoF systems enables cooperative sensing and consequently reduces the interference detection time while improving the detection accuracy. Moreover, the multi-channel carrying capability of advanced broadband RoF systems can significantly increase available radio resources at each WLAN AP. By implementing dynamic radio resource management based on accurate spectrum sensing, interference avoidance or mitigation can be easily accomplished. Effectively, the CWLANoF architecture enables the new concept of applying cognitive radio techniques for equal spectrum access in the ISM band.

In a conventional WLAN, each AP has an 802.11 radio modem and is digitally bridged to the distribution system, usually an 802.3 Ethernet. In a CWLANoF, radio modems and bridges in the APs are moved to a centralized unit referred as the *cognitive access point* (CogAP); the resulting simplified APs are now named as *remote antenna units* (RAUs). RAUs are connected to the CogAP via analog radio frequency signals transmitted over optical fibers. By centrally processing

broadband RF signals received from the RAUs, the CogAP has a complete picture of the radio spectrum usage in the coverage area of the WLAN ESS. The Distributed Coordinated Function of 802.11 MAC employing carrier-sensed multiple access with collision avoidance (CSMA/CA) is carried out at the CogAP instead of at individual APs as in a conventional WLAN. These changes enable a CWLANoF to more effectively combat packet collisions that inevitably occur over a random-access channel. A typical CWLANoF system and the structure of the CogAP are illustrated in Fig. 1.

The paper is organized as follows. In Section II, we review related work on how to improve WLAN system capacity. In Section III, two methods are proposed to reduce collisions among WLAN stations: *load balancing* to reduce collisions caused by heavy traffic, and *transmitter and receiver diversity* to reduce the effects of collisions. The performance of proposed methods is evaluated through extensive Monte-Carlo simulations in Section IV. We conclude by summarizing the paper and discussing future work in Section V.

II. Related Work on WLAN

Much recent research on WLANs aims to increase system capacity of individual WLAN BSSs, and reduce co-channel interference (CCI) and adjacent-channel interference (ACI) among BSSs in WLAN ESSs.

The system capacity of a WLAN BSS can be increased through three methods: enhancing existing MAC protocols by either adjusting parameters or adding new MAC flavors to achieve a higher MAC efficiency, exploiting capture effects, and introducing multiple-input, multiple-output (MIMO) to exploit spatial multiplexing. Enhancing the WLAN MAC protocol usually requires an update of station hardware or firmware. We therefore mainly review recent advances on exploiting capture effects and employing MIMO in WLAN.

The capture effect has been initially studied within the context of an ALOHA network [14]. It refers to the fact that when two packets arrive at one station at the same time, the packet with stronger signal strength will be synchronized and "captured" by the station. Luo and Ephremides [15] showed that with the capture effect, system throughput is maximized when all nodes transmit at maximum power. This conclusion, however, is based on an optimistic assumption that any packet can be successfully received as long as it has the highest power level at the receiver, regardless of how many overlapping packets are being received at lower power levels. After taking interference into account, Hadzi-Velkov and Spasenovski investigated the capture effect and its interaction with RTS/CTS (request to send/clear to send) in 802.11b networks [16]. Kochut *et al*. studied the capture effect by comparing system throughput at the physical and transport layers in 802.11b networks [17]. Their comparisons showed that capture effect is magnified through variations of contention window size in the MAC layer and congestion window size in the Transmission Control Protocol (TCP) layer. Based on Bianchi's model [18], WLAN performance is derived in [19] by considering the capture effect. Capture effect and successive interference cancellation were later studied in a direct sequence spread spectrum (DSSS)-based ZigBee network [20].

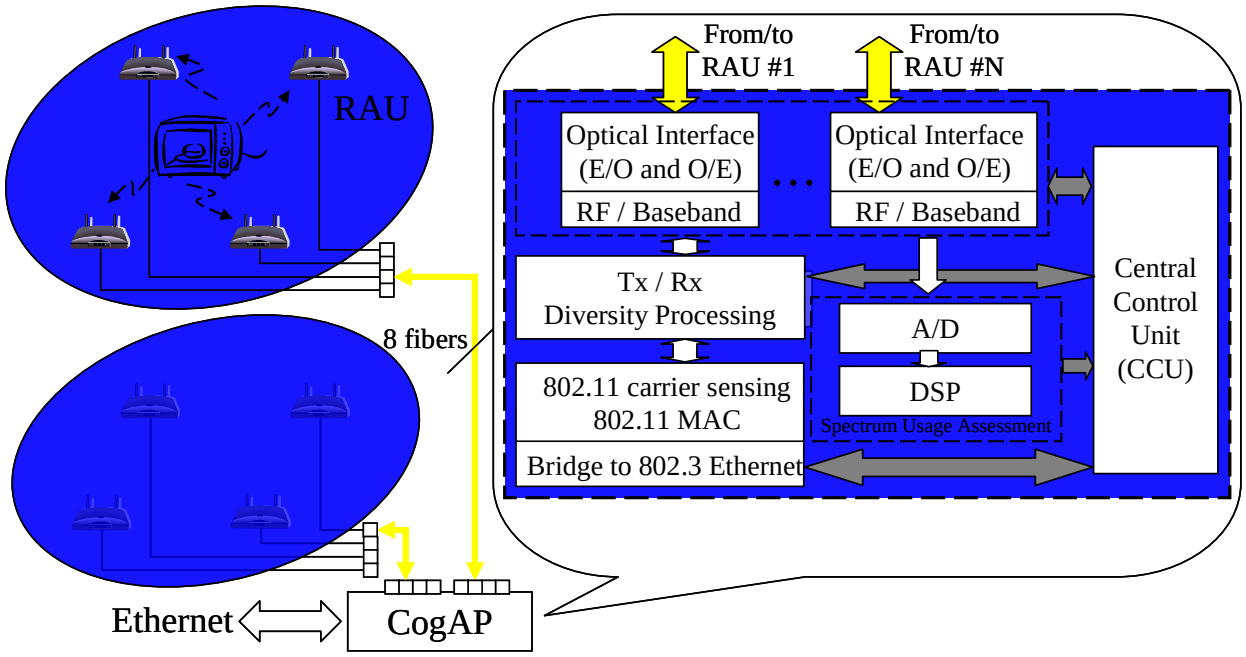


Fig. 1 A typical CWLANoF system and CogAP structure. E/O: electrical-optical converter. O/E: optical-electrical converter.

Capture effect in 802.11a networks was studied in [21][22] through real-world experiments using commercial WLAN devices. It was shown that with an arrival time difference of up to 50 µs, the stronger 802.11a packet can still be captured. Different from previous 802.11b capture effect studies where the stronger frame has to arrive within the preamble time of the weaker frame, this observation suggests that even when the arrival time difference of two packets is larger than the preamble length of the first packet, the stronger packet could still be captured. Such phenomena have been observed in commercial 802.11a/b/g adapters working in either DSSS mode or orthogonal frequency division modulation mode.

The capture effect was further exploited in the form of "message in message" (MIM) to increase system throughput [23][24]. The AP sends the message with smaller channel gain first and the message with larger channel gain later such that the weaker packet's preamble can be successfully locked by one recipient and the stronger packet can also be locked by another recipient. The AP abuses CSMA rule and stations use delayed ACKs. Using MIM requires the AP to update the system interference map periodically.

Exploiting diversity in WLAN is classified into micro-diversity and macro-diversity. The IEEE 802.11n standard is developed to enable micro-diversity in WLANs using MIMO. Previous work on macro-diversity includes the concept of distributed radio bridges proposed in [25] and their subsequent applications in WLAN [26][27].

A WLAN ESS is a multi-cell WLAN system in which the WLAN controller assigns channels and sets maximum AP transmit power to different BSSs to reduce CCI and ACI among them. Sub-optimal radio resource management algorithms have been extensively studied for this purpose. These algorithms address three basic problems: channel allocation across APs [28], user association (or load balancing) [29], and AP transmit power control [30]. The conflict set coloring method jointly optimizes channel allocation and load balancing [31]. Measurement-driven guidelines in [32] provide a heuristic method to jointly address the three basic problems. However, due to the limited number (usually one) of channels that each BSS can support, these algorithms have limited abilities to handle dynamic traffic, and become extremely complicated when channel allocation, load balancing and AP transmit power control are jointly considered. Authors of [33] investigated how to coordinate MAC mechanisms across multiple APs in the ESS by switching from contention-based access to time-slotted access when the ESS is heavily loaded with audio and video streams. The MAC switching reduces packet collisions and thus provides better quality-of-service for multimedia streams. However, the signaling protocol required by the AP coordination was not given in [33].

III. COLLISION REDUCTION

A collision happens when two stations access the channel at the same time, or when one station fails to sense an on-going packet transmission due to fading or hidden terminal problem and starts a new transmission. Based on the CWLANoF architecture, in this paper we propose a load-balancing method to reduce collisions caused by heavy traffic, and a transmitter and receiver diversity method to reduce the impact of packet collisions by increasing the chance of successful reception. The two methods used to reduce collisions in CWLANoF are illustrated in Fig. 2.

A. Load Balancing Method

A practical load balancing technique facilitated by the CWLANoF architecture is to distribute the total traffic load in the frequency domain. The broadband RoF connection between each RAU and the CogAP allows more than one channels to be allocated to any RAU. Consider the case of two RAUs covering a given area: RAU1 operates on the channel f_1 and RAU2 operates on f_2. When the collision rate on f_1 is higher than a target threshold, the CogAP can use the "disassociation" process to force some of the stations to be dissociated from this channel, while simultaneously sending beacons on a different channel f_3. Stations dissociated from f_1 will then have two options. If a dissociated station receives beacons on channel f_2 from RAU2, it can request to associate with RAU2 on this channel. This effectively transfers a portion of the traffic load of RAU1 to RAU2, creating a distributed load balancing solution among RAUs. Alternately, a dissociated station will receive beacons on channel f_3 from RAU1, and request to associate with RAU1 over this channel. In this case, load balancing occurs over the frequency domain within the same RAU, where a portion of the traffic at RAU1 is switched from overloaded channel f_1 to channel f_3. The second case is particularly made possible by the broadband RoF connections between RAUs and CogAP. In contrast, conventional WLAN APs are generally not equipped for multi-channel operations.

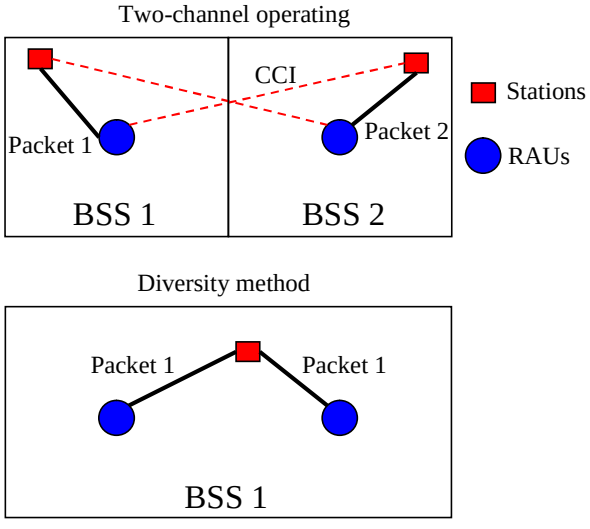


Fig. 2 Collision reduction: diversity and two-channel-operation

The gain in system throughput in the above example is two-fold: one from increased MAC efficiency due to decreased contention among stations accessing the same channel, and another from the use of three channels instead of two. We are more interested in the latter owing to its potential of linearly increasing system throughput. However, to fairly compare a CWLANoF with a conventional WLAN, we investigate the worst case where the new channel assigned to RAU1 is the same as that assigned to RAUs, i.e., f_2. We shall examine the throughput gain that can be achieved in the presence of co-channel-interference on f_2.

WLAN operations on f_1 and f_2 can be independent and as such we refer to this load-balancing method as *multiple-independent-channel-operation*. Let us compare a two-AP conventional WLAN, where AP1 operates on f_1 and AP2 operates on f_2, with a two-RAU CWLANoF, where RAU1 operates on f_1 and f_2 and RAU2 operates on f_2. We can certainly focus on the throughput on f_2. It is clear that the CWLANoF provides the worst throughput on f_2 when the spatial frequency re-use is impossible, i.e., when all stations associated on f_2 can perfectly hear each other. We now argue that even in such a situation, the CWLANoF could provide a higher throughout than a conventional WLAN. For simplicity, we only consider downlink traffic. In the conventional WLAN, AP2 can only send one data packet at a time. In the CWLANoF with RAU1 and RAU2 independently operated, they might simultaneously send two data packets on f_2. Owing to capture effects, the two packets may both survive from the collision, thus generating a throughput gain.

B. Transmiter and Receiver Diversity Method

Besides operating channels independently, the CogAP can also manage channels to exploit macro-diversity since signals received from widely separated RAUs tend to be uncorrelated. If each RAU is also equipped with multiple antennas, we can further implement micro-diversity in conjunction with macro-diversity. However, if we keep the number of fibers between each RAU and the CogAP the same, i.e., one to transmit and one to receive, wavelength division multiplexing would then be required to deliver RF signals from/to different antennas attached to the same RAU. Here we focus on macro-diversity enabled by distributed RAUs.

1) Receiver Diversity

Consider an area covered with two RAUs, using the same set of frequencies to serve a group of stations. If maximum-ratio combining (MRC) is used at the CogAP for uplink signals, not only do we achieve an array gain of 3 dB due to the increased receive antenna gain, but also obtain a diversity gain if the two paths from the station to the two RAUs experience independent fading. Both gains will help reduce the effects of packet collisions, resulting in increased throughput and reduced packet error rate (PER). When the number of RAUs increases to four, we expect a higher performance improvement due to 3 dB more in array gain and a higher diversity gain.

An immediate effect of receiver diversity is an improvement in sensing capability at the CogAP, and hence a reduction in WLAN packet collisions between downlink packets and uplink packets. Another effect of diversity gain is to reduce unfairness among stations in terms of their chances to access the channel due to their different distances from the RAUs.

2) Transmitter Diversity

For the downlink, we can use transmitter diversity to improve signal-to-noise-ratio (SNR) at the stations without requiring them to have additional capabilities. Multiple copies of each packet are distributed to RAUs and then to the destination such that when some copies are largely attenuated due to poor channel conditions, other copies can still reach the destination; hence, transmitter diversity. By reciprocity of the channel, transmitter diversity at the CogAP through multiple RAUs achieves the same SNR gain as receiver diversity, subject to a total transmit power constraint on all RAUs.

We investigate equal-gain combing (EGC) and MRC using transmitter diversity. In EGC scheme, each RAU is subject to a given per-RAU transmit power constraint, which reduces distortions due to nonlinearity at optical-electrical converters of RAUs. In MRC scheme, RAUs are only subject to a total transmission power constraint, and therefore have a larger freedom on transmission power allocation across RAUs, providing a larger SNR gain than EGC.

Both EGC and MRC require that signals from different RAUs can be added coherently at the receiving station. Therefore, the CogAP must have exact channel state information (CSI) from all participating RAUs to the receiving station right before a packet is sent, such that signal phases can be properly shifted at the different RAUs. This makes CSI estimation for transmitter diversity more difficult than receiver diversity, where the CogAP can always rely on the physical-layer header of WLAN packets to estimate CSI.

IV. Performance Evaluations

We utilize the NS-2.33 simulator [34] with its dei80211mr WLAN rate adapter package [35] to evaluate the performance of the proposed methods. The interference-recorded channel model incorporated in this package greatly enhances the accuracy of simulations involving channel capturing.

A. Simulation Model

The simulation model includes two RAUs connected to one CogAP, which is then connected to a fixed host computer. Stations are either uniformly or non-uniformly placed in a 30-by-60 m^2 area. When no diversity is used, the CogAP communicates with stations through their closest RAUs. Traffic streams only flow between stations and the fixed host. The wireless propagation model is a simplified pathloss model [36] with shadowing and Rayleigh fading.

The WLAN uses 802.11g and two non-overlapping channels are used. Each AP in the baseline conventional

WLAN operates on one channel only, while the CogAP in CWLANoF operates on both channels through the two RAUs either co-operatively for macro-diversity or independently. Data mode used by each station and the CogAP is determined by the signal-to-noise-ratio-based dynamic rate adaptor in dei80211mr package. No RTS/CTS is used. Perfect CSI is assumed to be available at the CogAP.

The frequency plan used in the simulations is shown in Fig. 3 and simulation parameters are listed in Table 1. The synchronization interval (SI) is used to model the capturing process. When the arrival times of two packets are within the SI, it is assumed that the receiver is able to synchronize to the packet with the stronger receive power.

The simulations employ two types of traffic that represent increasingly popular Internet applications: File Transfer Protocol (FTP) over TCP in downlink representing traffic from file downloading applications, and constant bit-rate (CBR) traffic in both uplink and downlink representing Voice over Internet Protocol (VoIP) and IP television (IPTV). FTP over TCP traffic is saturated, i.e., stations always have packets to send. VoIP and IPTV, as multimedia traffic, have the same fixed packet interval yet different packet length due to different amount of information contained in their packets. We are mainly interested in file downloading speed and voice and video quality; thus, TCP downlink throughput and CBR PER are chosen as our main performance evaluation criteria. To evaluate proposed methods for file sharing applications, we also evaluate TCP uplink throughput in some simulation scenarios.

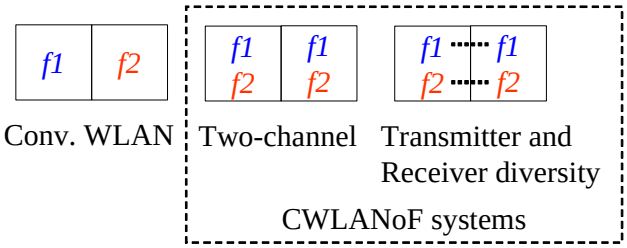


Fig. 3 Frequency plan in simulations

TABLE 1. SIMULATION PARAMETERS

Propagation	Pathloss exponent = 2.5
	Reference distance, d_0 = 2 m
	Standard deviation of shadowing = 3.5 dB
PHY	Transmission power, P_t = 10 mW
	Carrier-sensing threshold = -70 dBm
MAC	Synchronization interval = 5 μs
	aSlotTime = 20 μs
	CWMin/Max = 31 / 1023
FTP traffic	TCP/Reno. Packet size = 1000 bytes.
CBR traffic	Packet interval 20 ms. 1000-byte packets for IPTV; 40-byte packets for VoIP.

B. *Effects of Receiver Diversity*

We first investigate the effect of receiver diversity, i.e., receiving packets transmitted from one station using more than one RAU. Assuming the channel gain of each path is Rayleigh distributed, we know that the received signal power at a given RAU$_i$, $P_r^i(x,y)$, is an exponentially distributed random variable with the probability density function (p.d.f.)

$$f_{i,x,y}(P_r^i) = 1/P_{avg}^i(x,y) \cdot e^{-P_r^i / P_{avg}^i(x,y)}, \qquad (1)$$

where $P_{avg}^i(x,y)$ is the averaged power of signals received by RAU$_i$ from a station located at (x,y) and reflects the pathloss between them. Thus, at the CogAP that receives the signals from both RAU$_i$ and RAU$_k$, the p.d.f. of the total received signal power is given by

$$f_{cap,x,y}(P_r^{cap}) = \frac{(e^{-\frac{P_r^{cap}}{P_{avg}^i(x,y)}} - e^{-\frac{P_r^{cap}}{P_{avg}^k(x,y)}})}{P_{avg}^i(x,y) - P_{avg}^k(x,y)}, \qquad (2)$$

where P_r^{cap} is the power of the combined signal at the CogAP (by combining signals from RAU$_i$ and RAU$_k$) and the notation $f_{cap,x,y}(P_r^{cap})$ implies the p.d.f. of P_r^{cap} is a function of (x,y), the geographical coordinate of the transmitter. When the station has the same average pathloss to the two RAUs, $f_{cap,x,y}(P_r^{cap})$ becomes an Erlang distribution with the shape factor N=2. Given different locations of stations, the CogAP and individual RAUs exhibit different outage probabilities Prob($P_r \le$ threshold) over the whole area, as shown in Fig. 4. Fig. 5 shows the reduction on the outage probability when MRC or EGC is used, compared to the case where diversity is not employed. In both figures the blue surfaces correspond to EGC and the red ones correspond to MRC. The black surface corresponds to conventional WLAN in Fig. 4 and serves as the zero-reference plane in Fig. 5. Results are based on simulations using parameters in Table 1. Two RAUs (or APs in the conventional WLAN) are fixed at locations (15, 15) and (15, 45) in units of meters.

We observe that MRC always reduce outage probabilities more than EGC, especially for stations at corner areas where the pathlosses to the two RAUs largely differ. From Fig. 4, we also observe that compared with EGC, MRC has a flatter distribution of outage probabilities across the area, so that stations located farther away from the RAUs will still be heard by the CogAP. Therefore, their uplink packets will less likely collide with downlink packets, and consequently their contention windows will not suffer as much from exponential increases. With receiver diversity at the CogAP, stations farther away from the RAUs still have good chances to access the channel compared with those closer to the RAUs.

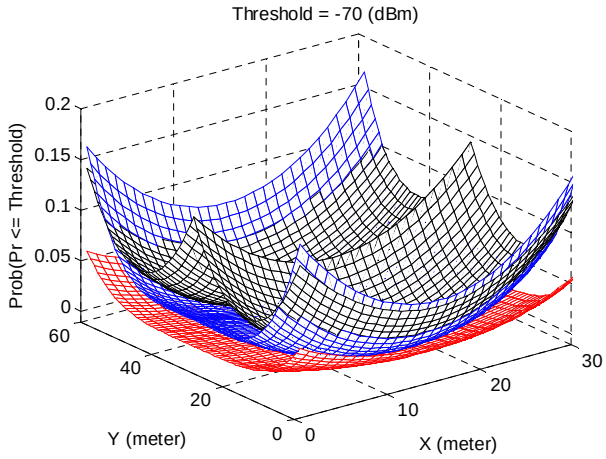


Fig. 4 Outage probabilities across a WLAN ESS

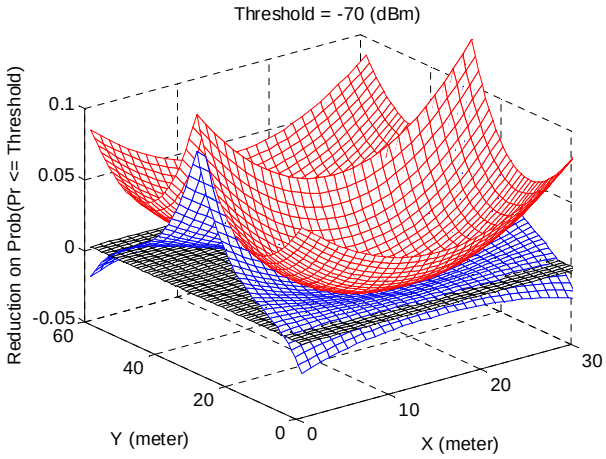


Fig. 5 Reduction in outage probability with MRC and EGC

C. *Spatially Uniformly Distributed Traffic*

Stations are uniformly placed over the whole area to represent spatially uniformly distributed traffic. For a given number of active stations, 20 different station locations are randomly generated. Simulated results under these scenarios are then averaged for evaluations. From Fig. 6 and Fig. 7, we observe that under two types of spatially uniformly distributed traffic, operating two channels in both RAUs increases TCP throughput by 14%~18% when only 8 stations are active. When the number of active stations increases, the TCP throughput gain also increases, reaching 36% for VoIP uplink/downlink plus FTP downlink traffic and 20% for IPTV downlink plus FTP downlink traffic when the number of active stations is 32. Downlink PER for CBR traffic is also reduced by 50%. Uplink PER for CBR is zero and thus not presented here. The reason is that uplink is not heavily loaded without presence of saturated FTP over TCP traffic. The performance gain of two-channel-operating can be attributed to channel capturing effects on downlink data packets in CWLANoF (see Section III.A).

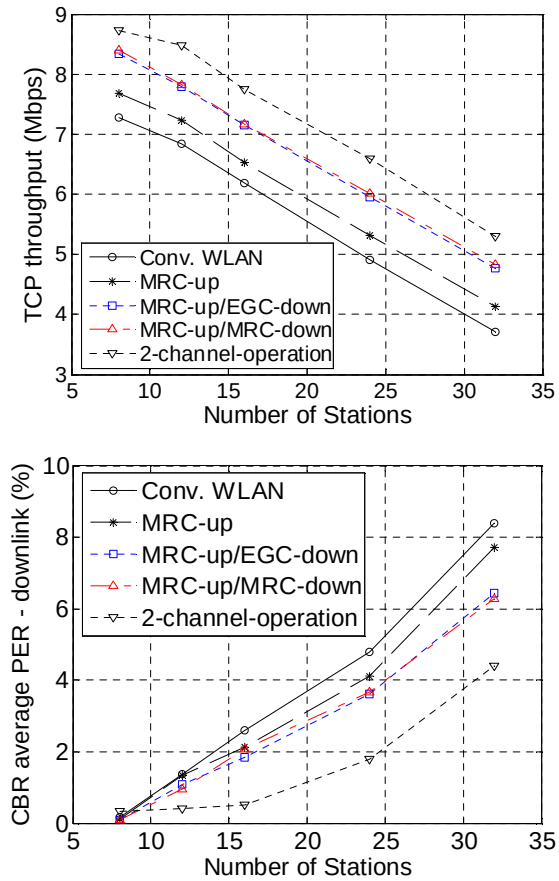


Fig. 6 TCP throughput and CBR PER vs. Number of stations. Traffic: VoIP uplink/downlink + FTP downlink. MRC-up: MRC is used for uplink diversity. EGC-down: EGC is used for downlink diversity.

Note that we use the number of stations to indicate the intensity of traffic since all of stations have always-on CBR and FTP/TCP traffic.

The confidence interval of obtained average PERs can be estimated by *berconfint* function in MATLAB, provided that the number of packet errors follows binomial distribution. Each simulation run lasts 120 seconds with 20 ms packet interval. Therefore, in 12-station case, 36,000 downlink and 36,000 uplink packets are generated, resulting a 95% confidence interval [0.93%, 1.08%] at 1% average PER, and [0.08%, 0.13%] at 0.1% average PER. A higher number of stations or a higher average PER generates a tighter confidence interval. To avoid clutter, we do not superpose the confidence intervals on PER figures.

Comparing the diversity methods we have investigated, using only MRC for uplink diversity slightly increases TCP throughput and reduces downlink PER for CBR traffic, while engaging additionally downlink EGC or MRC transmit diversity further improves performance by providing a higher TCP throughput gain and lower PER for CBR traffic. The results also show that two-channel-operation always outperforms the diversity methods in either TCP throughout or downlink PER for CBR traffic. The advantage of multi-

channel operation originates from additional operation channels, which can linearly increase system capacity (assuming no CCI), while diversity methods we investigate here only logarithmically increase system capacity. When the always-on VoIP traffic is not present, our simulation results showed similar TCP throughput gains from both multi-channel operation and diversity methods. The results are not presented here to avoid repetition.

As shown in Fig. 8, heavy traffic streams like VoIP uplink/downlink and FTP uplink/downlink largely increase CBR PER. We observe that diversity methods still consistently improve TCP throughput when the number of active stations changes. CBR PER, however, is only slightly affected, and we regard the small difference of CBR PER between conventional WLAN and diversity methods as random effects in the simulations. In fact, our simulations show little difference among diversity methods; therefore, only MRC-uplink/MRC-downlink method is plotted in the CBR PER figure to avoid clutter. The two-channel-operation method improves TCP throughput and outperforms diversity methods when the number of active stations is less than 10. When the network contains more than 10 active stations, TCP throughput of two-channel-operation method decreases and even becomes worse than conventional WLAN when there are 24 active stations.

To identify the reason of TCP throughput degradation of two-channel-operation, we plotted the TCP uplink and downlink throughput separately in Fig. 9. We observe that two-channel-operation generates the highest TCP downlink throughput but the lowest TCP uplink throughput, which taken together causes the lowest total TCP throughput. To explain the reason of TCP uplink throughput degradation, we notice that when two-channel-operation method is used, stations being served in one channel are found in areas twice as large as those in the conventional WLAN, and therefore suffer more packet collisions due to the hidden terminal problem. The above observation suggests that when there are too many active stations in the CWLANoF ESS, enough number of channels should be operated to ensure proper file sharing efficiency.

Compared with diversity methods and conventional WLAN, the two-channel-operation method generates the highest CBR PER when the number of active stations is less than 12, and the lowest CBR PER when the number of active stations is larger than 12. This phenomenon is not easy to see in scenarios under VoIP traffic. To see it more clearly, observe scenarios under IPTV downlink and FTP downlink traffic, enlarged in Fig. 10. When two-channel-operation method is used in lightly loaded networks, CBR PER is increased because capturing a new packet causes a loss of

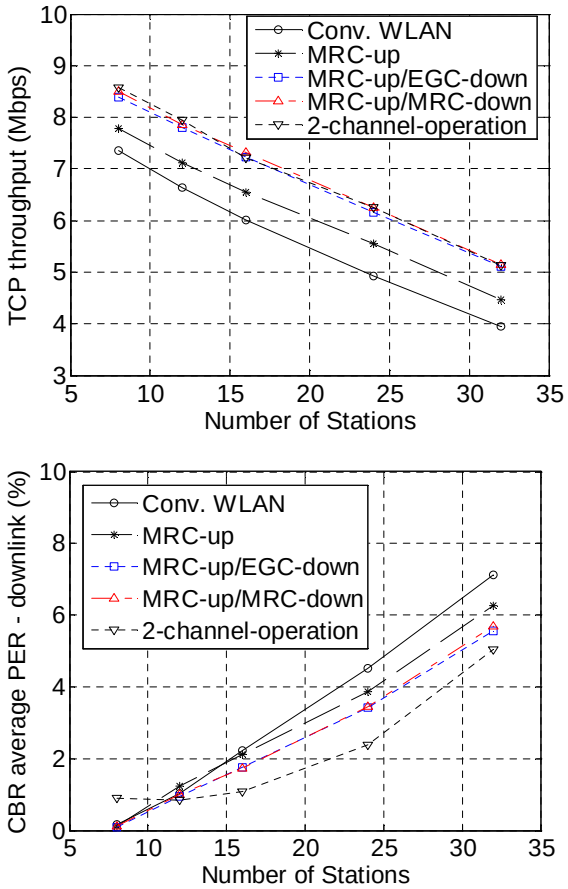


Fig. 7 TCP throughput and CBR PER vs. Number of stations. Traffic: IPTV downlink + FTP downlink.

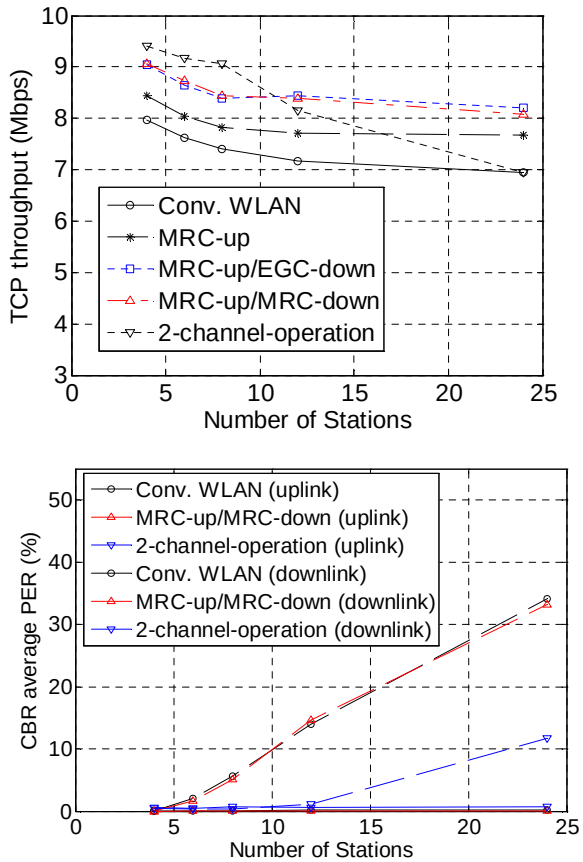


Fig. 8 TCP throughput and CBR PER vs. Number of stations. Traffic: VoIP uplink/downlink + FTP uplink/downlink.

previously being received packet; in heavily loaded networks, however, the reduced packet collisions due to extra channels outweighs the disadvantage of CBR packet loss due to channel capturing, causing lower overall CBR PER than diversity methods and conventional WLAN. Although two-channel-operation caused a bit higher CBR PER for uplink packets, CBR PER in downlink is largely decreased. For VoIP application, such balanced CBR PERs provided by two-channel-operation would be very useful.

1) Explanations on the performance improvement

TCP throughput gain and CBR PER reductions of diversity-based CWLANoF systems come from independent channel fading and more antennas involved at the receiver or transmitter. Gains of two-channel-operating CWLANoF systems come from channel capturing and channel fading effects. We now further explain where these gains come from.

Suppose a conventional WLAN ESS serves 10 stations on channel 1 of AP1 and another 10 stations on channel 2 of AP2. Assume these stations are associated to the AP closer to them. Two-channel-operating actually splits the 20 stations into 4 groups, each assigned to one channel through one RAU. The resulting CWLANoF system can be viewed as four independently-operated conventional BSSs. Although CCIs exist between these BSSs, we still gain system capacity due to the linearly increased bandwidth while the signal-to-interference-noise ratio is only logarithmically degraded. On the other hand, a diversity-based CWLANoF controls RAUs and forms BSSs with distributed antennas, serving stations that spread out in the whole area.

An intuitive example can observed from Fig. 2: apparently stations located in the middle of the area will favor diversity technique since they have similar average pathlosses to the two RAUs, while stations at corner areas will favor multiple-independent-channel-operating method since there will be less CCIs between corners. This observation reminds us that when the location information of stations is available at the CogAP, advanced location-aware channel management techniques can provide even higher system capacity. How to achieve a balance between multiple-independent-channel-operating method and diversity techniques to better serve stations with dynamic traffic would be our next research topic.

It should be noted that when uplink or downlink diversity is used, if two or more packets collide, we discard both of them in simulations. Therefore, the performance shown is the lower bound of MRC or EGC performance in practice. And once multi-user reception techniques such as sequential interference cancellation are applied, diversity could obtain higher TCP throughput and decrease the CBR PER more. However, in multi-user access, the MAC layer needs a careful joint-design with sophisticated signal processing at the physical layer.

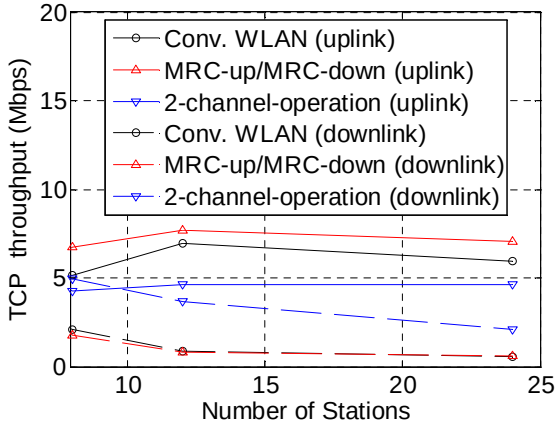


Fig. 9 TCP throughput degradation of two-channel-operation method in heavily loaded networks. Traffic: VoIP uplink/downlink + FTP uplink/downlink.

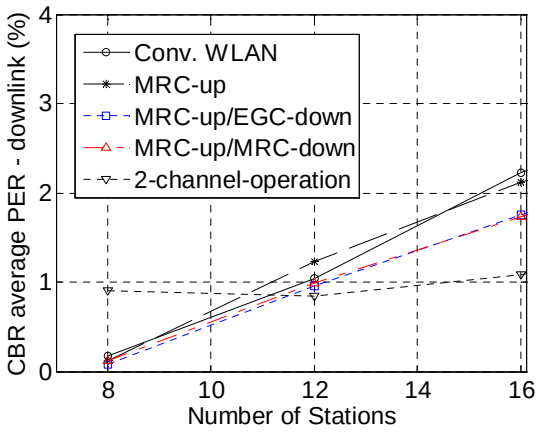
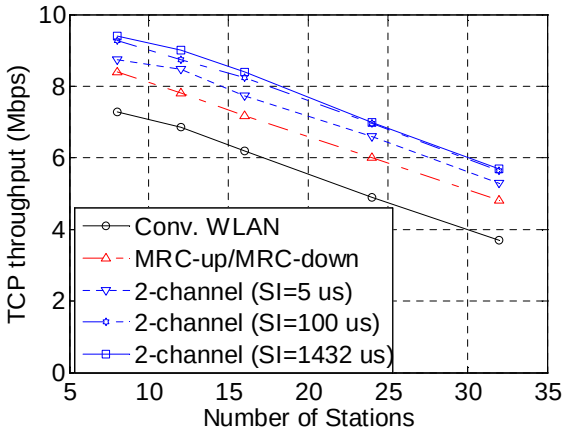


Fig. 10 CBR PER degradation of two-channel-operation in lightly loaded networks. Traffic: IPTV downlink + FTP downlink.

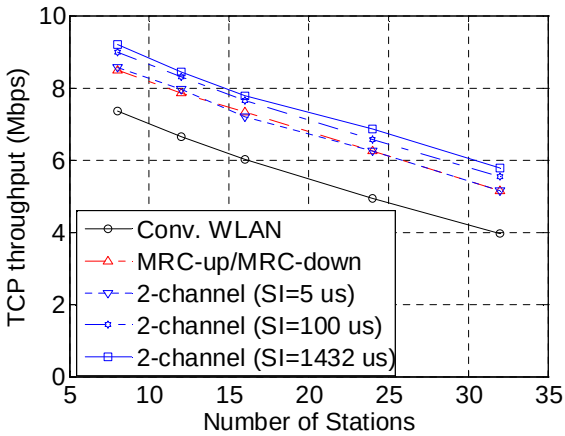
2) Effects of SI values

Results in [1] only showed scenarios using SI = 5 µs. We examine effects of different SI values on TCP throughput under three types of traffic, as shown in Fig. 11. Effects of SI on CBR PER is very small and thus omitted. Since different SI values only affect two-channel-operation; only one diversity method (MRC in uplink and downlink) is plotted for comparison purpose.

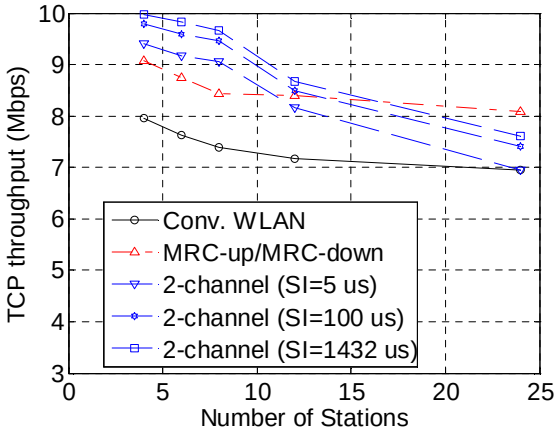
We observe that by using SI = 1432 µs, TCP throughput can be increased by 5.6%~10% when compared with SI = 5 µs, owning to the fact that larger SI values cause more packet capturing than smaller SI values. However, we also notice that there is little difference between SI = 1432 µs and SI = 100 µs, indicating that by only looking for the strongest signal during SI = 100 µs, a WLAN receiver can achieve most of throughput gain due to capture effect. For the rest of this work, 100 µs is used as SI value.



(a) Traffic: VoIP uplink/downlink + FTP downlink



(b) Traffic: IPTV downlink + FTP downlink



(c) Traffic: VoIP uplink/downlink + FTP uplink/downlink

Fig. 11 Effects of SI values

D. *Spatially Non-uniformly Distributed Traffic*

When a hotspot area has much larger traffic demand than other areas, we face a spatially non-uniformly distributed traffic. We split the 30-by-60 m^2 area into 3-by-6 sub-areas and place the hotspot into one of these sub-areas to simulate non-uniformly distributed traffic. Totally 8 stations are used for background traffic and 4 other stations are placed in certain hotspot location, as shown in Fig. 12. By geometric symmetry, we only need study hotspot locations from 1 to 6. To concentrate on studying the effects of dynamic traffic, we fix the locations of background-traffic stations at the centers of sub-areas 2, 4, 6, etc. Stations that generate hotspot traffic are also fixed in the center part of their respective sub-area. Only one set of station locations is used for simulations followed.

As shown in Fig. 13 and Fig. 14, compared with conventional WLAN, diversity methods in CWLANoF achieve 10%~62% higher TCP throughput and two-channel-operation achieves 17%~48% higher TCP throughput, whereas only 14%~36% gain is achieved when the traffic is spatially uniformly distributed (comparing Fig. 6 to Fig. 8). CBR downlink PER is also largely reduced. This demonstrates CWLANoF's enhanced capability to handle dynamic traffic.

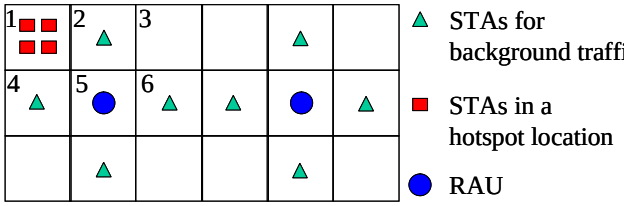


Fig. 12 Spatially non-uniformly distributed traffic. Hotspot locations are numbered from 1 to 6.

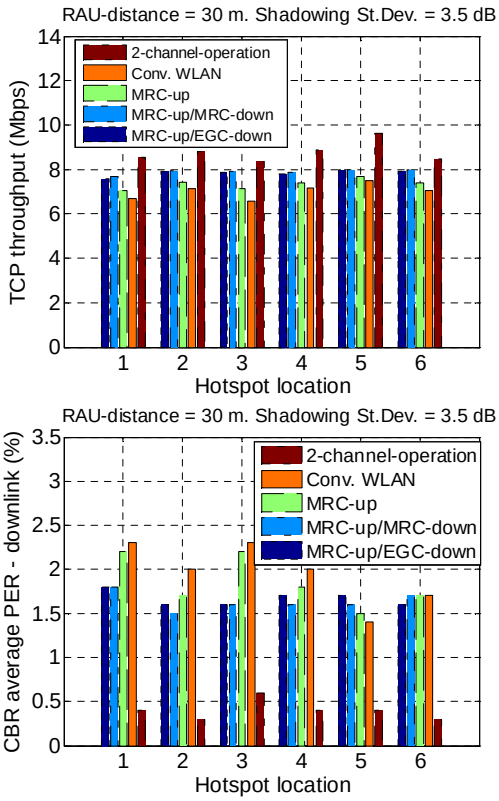


Fig. 13 TCP throughput and CBR PER vs. Hotspot location. (VoIP uplink/downlink + FTP downlink). RAU-distance = 30 m. Shadowing St.Dev. = 3.5 dB.

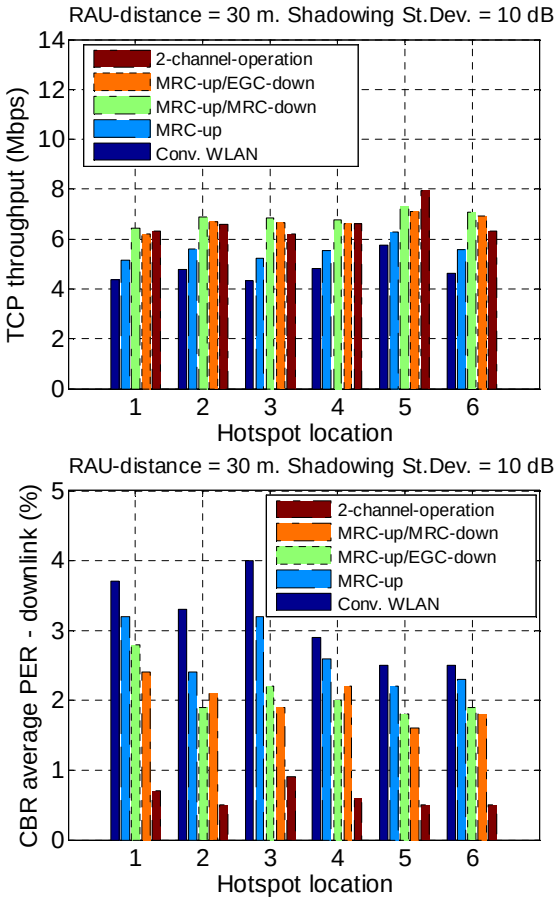


Fig. 14 TCP throughput and CBR PER vs. Hotspot location. (VoIP uplink/downlink + FTP downlink). RAU-distance = 30 m. Shadowing St.Dev. = 10 dB.

Not surprisingly, two-channel-operating achieves a larger throughput gain when the hotspot is in the corners (e.g., location 1), while diversity methods achieve larger gains when the hotspot is in the overlapping area of RAU1 and RAU2 (e.g., location 3 and 6). In fact in such areas, MRC in both uplink and downlink achieves higher TCP throughput than two-channel-operating when the standard deviation of shadowing increases to 10 dB.

When the hotspot moves to location 5, stations at the hotspot are closer to RAU1. Therefore, CCI from stations in BSS2 to those in BSS1 is less likely due to the capture effect. Thus, we observe a larger TCP throughput gain in the tow-channel-operation method, as shown in Fig. 13 and Fig. 14.

Focusing on spatially non-uniformly distributed traffic, we further study the effects of RAU-distance (i.e., the size of BSS) and the maximum transmission power of cooperating RAUs.

1) Effects of RAU-distance

The RAU-distance is also the physical size of a BSS in our simulations. Comparing Fig. 15 with Fig. 14, we observe that both diversity and two-channel-operation methods provide larger TCP throughput gains when the RAU-distance is increased to 45 or 60 meters. SNR gains generated by diversity methods have larger effects on throughput due to increased RAU-distance and consequently increased pathloss. Two-channel-operation method provides higher throughput gains due to increased RAU-distance and consequently reduced CCI, especially at hotspot location 5 where stations are less susceptible to CCIs from stations being served by the neighboring RAU.

2) Effects of total transmission power constraint of cooperating RAUs

In previous simulations, the maximum transmission power of each conventional AP is 10 mW. When MRC or EGC is employed at an RAU, the maximum transmission power of each RAU is set as 5 mW for a fair comparison with conventional WLAN. However, an RAU is covering the same area as a conventional AP. Therefore, an RAU can transmit at 10 mW without causing excessive interference to other ISM devices or health concerns on human bodies.

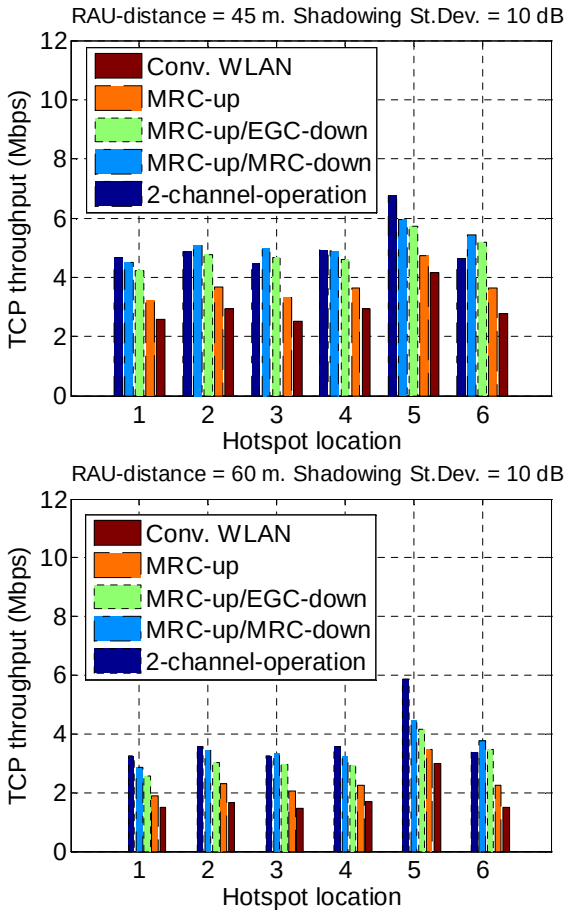


Fig. 15 Effects of RAU-distance. Traffic: VoIP uplink/downlink + FTP downlink. RAU-distance = 45 m or 60 m. Shadowing St.Dev. = 10 dB.

Effects of RAU-distance and RAU transmission power constraints should be examined jointly since both parameters affect system performance by changing SNR at receiver. Our results in Fig. 16 indicate that for usual office WLAN deployment (RAU-distance = 30~60 meters), both proposed methods improve TCP throughput. Compared with diversity methods subject to normal transmission power constraint, when the RAU-distance is small, increasing RAU transmission power constraint has negligible effects on TCP throughput.

V. Conclusion and Future Work

CWLANoF systems can provide a cost-effective and efficient method for devices to equally share the ISM band by taking advantage of cognitive radio capabilities. In this paper, we have proposed two methods that utilize the specialized capabilities of the CWLANoF architecture to improve system capacity by reducing packet collisions through load balancing and employing diversity to reduce the effects of packet collisions. By exploiting the wideband RoF connections between RAUs and the CogAP in a CWLANoF, multiple-impendent-channel-operation at each RAU has been proposed to reduce the collision probability in each channel by moving stations to different channels. By configuring RAUs as distributed antennas in a CWLANoF, we have demonstrated the use of macro-diversity to increase the sensing capability of the CogAP. Simulation results show that both methods can achieve 14%~18% TCP throughput gain and 10%~50% CBR PER reductions for spatially uniform traffic in an IEEE 802.11g network, and up to 62% TCP throughput gain when hotspots exist. We also studied effects of different SI values, RAU-distance, and total transmission power constraint of cooperating RAUs. Similar TCP throughput gain and CBR PER reduction are observed in all scenarios.

While the CWLANoF architecture raises a rich set of research problems, there is one promising direction to further reduce WLAN packet collisions, using location-aware radio resource management at the CogAP while employing collision reduction among stations. By taking advantage of its centralized signal processing capability, the CogAP can determine the locations of the stations. Obviously, heavy traffic from stations located in the overlapping area of RAUs are better served by applying diversity methods, whereas stations that do not benefit much from diversity can be off-loaded to new channels. This strategy is a simple example of how to dynamically exploit the performance gains between macro-diversity and multiple-independent-channel-operation. Further research is needed to jointly design signal processing techniques in physical layer and the MAC in CWLANoF systems such that multiple collided packets can be successfully received and acknowledged in time.

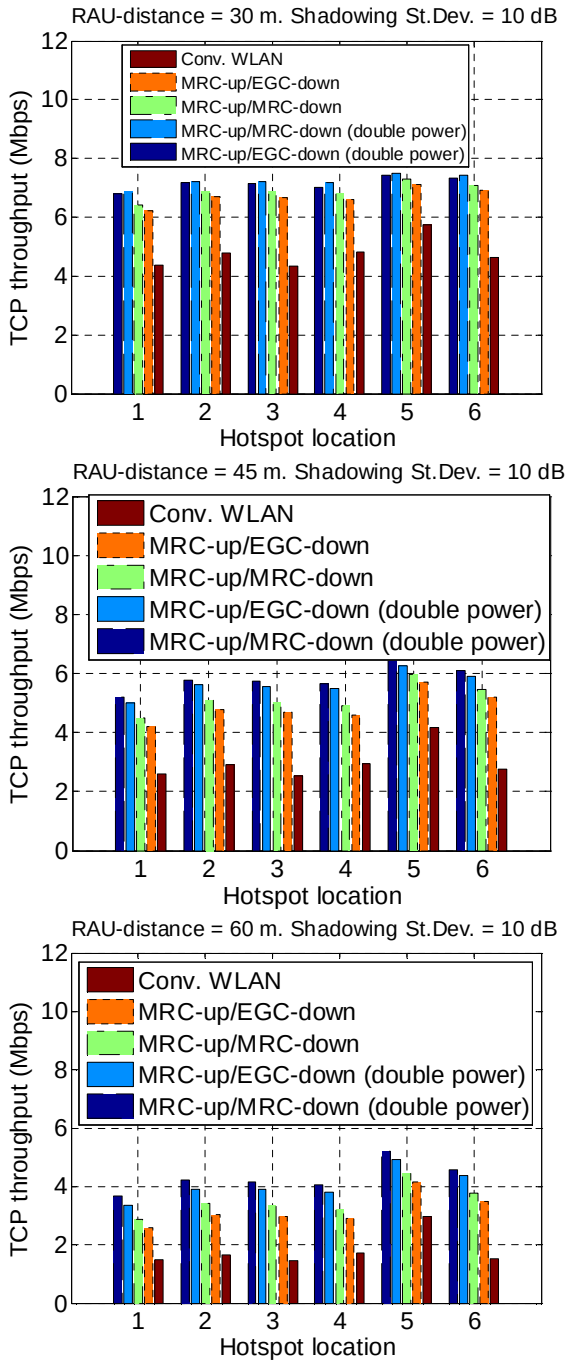


Fig. 16 Effects of total transmission power constraint of cooperating RAUs. Traffic: VoIP uplink/downlink + FTP downlink. RAU-distance = 30 m, 45 m or 60 m. Shadowing St.Dev. = 10 dB. Double power: the total transmission power constraint of cooperating RAUs is 20 mW.

Acknowledgment

This work was supported in part by the Canadian Natural Sciences and Engineering Research Council.

References

[1] H. Li, A. Attar, Q. Pang, and V. C. M. Leung, “Collision avoidance and mitigation in cognitive wireless local area network over fibre”, in *Proc. IEEE First International Conference on Evolving Internet (INTERNET’09)*, IEEE Press, Apr. 2009, pp. 133-138

[2] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13-18, Aug. 1999

[3] T.-S. Chu and M. J. Gans, "Fiber optic microcellular radio," *IEEE Trans. Veh. Tech.*, vol. 40, no. 3, pp. 599-606, 1991

[4] M. Sauer, A. Kobyakov, and J. George, "Radio Over Fiber for Picocellular Network Architectures," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3301-3320, 2007

[5] A. Nkansah, A. Das, C. Lethien, J.-P. Vilcot, N. J. Gomes, I. J. Garcia, J. C. Batchelor, and D. Wake, "Simultaneous dual band transmission over multimode fiber-fed indoor wireless network," *IEEE Microw. Wireless Compon. Lett.*, vol. 16, no. 11, pp. 627–629, 2006

[6] T. Niiho, M. Nakaso, K. Masuda, H. Sasai, K. Utsumi, and M. Fuse, "Transmission performance of multichannel wireless LAN system based on radio-over-fiber techniques," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 2, pp. 980-989, 2006

[7] A. Das, A. Nkansah, N. J. Gomes, I. J. Garcia, J. C. Batchelor, and D. Wake, "Design of low-cost multimode fiber-fed indoor wireless networks," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 8, pp. 3426-3432, 2006

[8] K. K. Leung, B. McNair, L. J. Cimini Jr., and J. H. Winters, "Outdoor IEEE 802.11 cellular networks: MAC protocol design and performance," in *Proc. IEEE ICC 2002*, vol. 1, 2002.

[9] M. G. Larrodé, A. M. J. Koonen, and P. F. M. Smulders, "Impact of radio-over-fibre links on the wireless access protocols," in *Proc. NEFERTITI Workshop*, Brussels, Belgium, Jan. 2005

[10] A. Das, M. Mjeku, A. Nkansah, and N. J. Gomes, "Effects on IEEE 802.11 MAC Throughput in Wireless LAN Over Fiber Systems," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 1-8, 2007

[11] [Online]. Available: http://www.zinwave.com/

[12] [Online]. Available: http://www.innerwireless.com/

[13] [Online]. Available: http://www.powerwave.com/inbuilding.asp

[14] J. J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, vol. COM-24, pp. 447–448, Apr. 1976

[15] W. Luo and A. Ephremides, "Power levels and packet lengths in random multiple access", *IEEE Trans. Inf. Theory*, vol. 48, pp. 46-58, 2002

[16] Z. Hadzi-Velkov and B. Spasenovski, "Capture effect in IEEE 802.11 basic service area under influence of rayleigh fading and near/far effect," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2002*, vol. 1, 2002, pp. 172 – 176

[17] A. Kochut, A. Vasan, A. U. Shankar, and A. Agrawala, "Sniffing out the correct physical layer capture model in 802.11b", in *Proc. 12th IEEE international conference on network protocols (ICNP) 2004*, pp. 252-261, 2004

[18] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function", *IEEE JSAC, 2000*, pp. 535-547

[19] H. Chang, V. Misra, and D. Rubenstein, "A general model and analysis of physical layer capture in 802.11 networks", in *Proc. IEEE INFOCOM 2006*, pp. 1-12, 2006

[20] D. Halperin, T. Anderson, and D. Wetherall, "Taking the sting out of carrier sense: interference cancellation for wireless LANs", in *Proc. 14th ACM international conference on mobile computing and networking*, pp. 339-350, 2008

[21] J. Lee and et. al., "An experimental study on the capture effect in 802.11 a networks", in *Proc. 2nd ACM international workshop on wireless network testbeds, experimental evaluation and characterization*, pp. 19-26, 2007

[22] J. Lee, J. Ryud, S-J Lee, and T. Kwon, "Improved modeling of IEEE 802.11a PHY through fine-grained measurements, *Computer Networks*, 2009, doi: 10.1016/j.comnet.2009.08.003

[23] J. Boer *et al.*, "Wireless LAN with enhanced capture provision (Lucent Technologies, Inc.)", *US Patent 5,987,033*, 1999

[24] N. Santhapuri *et al.*, "Message in message (MIM): A case for reordering transmissions in wireless networks", in *Proc. HotNets 2008*

[25] V. C. M. Leung and A. W. Y. Au, "A wireless local area network employing distributed radio bridges," *Wireless Networks*, vol. 2, no. 2, pp. 97-107, 1996

[26] A. Miu, H. Balakrishnan, and C. E. Koksal, "Multi-radio diversity in wireless networks," *Wireless Networks*, vol. 13, no. 6, pp. 779-798, 2007

[27] K. Whitehouse, A. Woo, F. Jiang, J. Polastre, and D. Culler, "Exploiting the capture effect for collision detection and recovery," in *Proc. Embedded Networked Sensors 2005*, pp. 45-52, 2005

[28] B. Kauffmann, F. Baccelli, A. Chainteau, V. Mhatre, K. Papagiannaki, and C. Diot, "Measurement-Based Self Organization of Interfering 802.11 Wireless Access Networks," in *Proc. IEEE INFOCOM 2007*, Anchorage, Alaska, USA, May 2007

[29] A. Kumar and V. Kumar, "Optimal Association of Stations and APs in an IEEE 802.11 WLAN," in *Proc. National Conference on Communications (NCC) 2005*, Feb. 2005

[30] V. Mhatre, K. Papagiannaki, and F. Baccelli, "Interference Mitigation through Power Control in High Density 802.11 WLANs," in *Proc. IEEE INFOCOM 2007*, Anchorage, Alaska, USA, May 2007

[31] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. Arbaugh, "A client-driven approach for channel management in wireless LANs," in *Proc. IEEE INFOCOM 2006*, Barcelona, Spain, Apr. 2006, pp. 1-12

[32] I. Broustis, K. Papagiannaki, S. V. Krishnamurthy, M. Faloutsos, and V. Mhatre, "MDG: measurement-driven guidelines for 802.11 WLAN design," in *Proc. ACM MobiCom 2007*, Montréal, QC, Canada, pp. 254-265

[33] M. Abusubaih, B. Rathke, and A. Wolisz, "A framework for interference mitigation in multi-BSS 802.11 wireless LANs", in *Proc. IEEE WoWMoM 2009*, Kos Greece, June 2009, pp. 1-11

[34] [Online]. Available: http://www.isi.edu/nsnam/ns/

[35] [Online]. Available: http://www.dei.unipd.it/wdyn/?IDsezione=5091

[36] A. Goldsmith, *Wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005

On Optimization of Wireless Mesh Networks using Genetic Algorithms

Rastin Pries, Dirk Staehle, Barbara Staehle, Phuoc Tran-Gia
University of Würzburg
Institute of Computer Science
Chair of Communication Networks
Würzburg, Germany
Email: {pries, dstaehle, bstaehle, trangia}@informatik.uni-wuerzburg.de

***Abstract*—Next generation fixed wireless networks are most likely organized in a mesh structure. The performance of these mesh networks is mainly influenced by the routing scheme and the channel assignment. In this paper, we focus on the routing and channel assignment in large-scale Wireless Mesh Networks to achieve a max-min fair throughput allocation. As most optimization approaches fail to optimize large wireless mesh network deployments, we investigate the usability of genetic algorithms for this approach. The results show the influence of the genetic operators on the resulting network solution and underline the advantages of a genetic optimization when applied carefully.**

***Keywords*-Wireless Mesh Networks, Planning, Optimization, Routing, Genetic Algorithms**

I. Introduction

The complex structure of Wireless Mesh Networks (WMNs) require a careful network planning and optimization. Our goal is to increase the throughput of the complete WMN while still sharing the resources fairly among the nodes. The planning of WMNs is in contrast to traditional wireless networks much more complex. On the one hand, a WMN consists of a multi-hop structure where not only interference on neighboring paths but also self-interference occurs. On the other hand, each node in the network can be equipped with multiple interfaces operating on different channels. The interference problems are covered by using the concept of collision domains. For the routing and channel allocation, an optimization method is required, which is fast enough to optimize even large WMNs. In our previous papers [1], [2], we evaluated the usability of Genetic Algorithms (GAs) for this optimization approach. GAs are based on the idea of natural evolution by simulating the biological cross of genes. Although GAs are generally not able to find the best solution, they provide near-optimal results in relatively small computation time. In this paper, we extend the evaluation by analyzing the influence of the genetic operators during the evolution and by introducing the concept of local optimization.

Our goal is to increase the throughput of the complete WMN while sharing the resources fairly among the nodes. This is achieved by applying a max-min fair share algorithm presented in [3] and by tuning the genetic parameters. A solution is max-min fair if no rate can be increased without decreasing another rate to a smaller value [4]. A max-min fair share algorithm is used instead of proportional fairness because the main goal is not to optimize the maximum overall throughput on the cost of fairness but to ensure a fair resource distribution between the users.

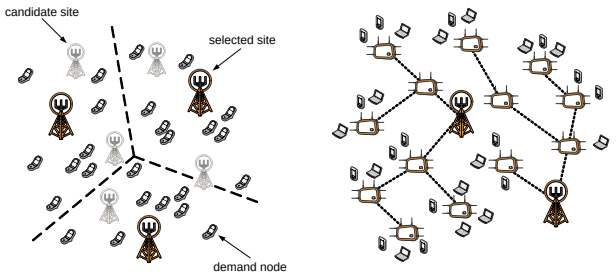
The rest of this paper is structured as follows. In Section II, we first give an introduction to wireless network planning, comparing traditional cellular network planning with wireless mesh network planning. This is followed by a short overview of global optimization techniques, which are applied in the related work section for optimizing WMNs. In Section IV, we describe our genetic algorithms for routing and channel assignment in detail. The performance of different genetic operators is evaluated in Section V and we optimize the genetic algorithm by introducing a local optimization approach in Section VI. Finally, conclusion are drawn in Section VII.

II. Network Planning and Optimization Issues

Network planning and optimization can be done using several techniques. On the one hand, signal quality measurements can be performed, which is very time-consuming and necessitates the access to all areas in which the network should be supported. On the other hand, a demand node concept can be used. This mechanism is often applied to cellular network planning. Furthermore, network planning can be done using an optimization mechanism. Meanwhile, a huge number of optimization techniques have been proposed and we decided to use genetic optimization due to its simplicity and the ability to plan even large networks.

A. Wireless Network Planning

The planning of wireless mesh networks can be applied to a variety of wireless networks, like WiMAX, WLAN, and sensor networks. Although the network technology changes, the planning challenges remain similar. In contrast to traditional cellular network planning, the planning and optimization of WMNs is much more complex. A widely used concept for cellular network planning is the demand node concept introduced by Tutschku [5] and illustrated in Figure 1(a).



(a) Cellular network planning. (b) WMN planning.

Figure 1. Comparison of traditional cellular network planning and wireless mesh network planning.

The algorithm first looks for the demands of cellular services. Therefore, different demographic areas are taken into account. For example, more phone calls occur in urban areas than in rural areas. According to these demographic regions, a different number of demand nodes are set up like shown in Figure 1(a). In addition to the demand nodes, candidate sites for base stations are inserted into the optimization algorithm. As each base station is able to support a fixed amount of users in cellular systems of the second generation, candidate sites are selected for base station placement in such a way that all demand nodes can be served with a certain probability.

In contrast, the planning of WMNs is much more complex. Not only the covered area or number of end users has to be considered, but also the capacity and the interference of the relaying links. The capacity of a link does not only depend on the distance between two mesh points, but also on the interference, which in turn depends on the used channels. Looking again at Figure 1(a), we can see that the channel assignment has to be performed in such a way that neighboring base stations do not use the same channel. In WMNs, such as shown in Figure 1(b), each mesh point can be equipped with multiple interfaces, which can be assigned one channel each.

In addition to the more complex channel assignment in WMNs compared to traditional cellular networks, also the routing has to be considered. In a fixed wireless mesh network where each mesh point is equipped with multiple interfaces, the Modulation and Coding Scheme (MCS), the interference from neighboring nodes, and the number of flows traversing a link have to be taken into account for the routing decision.

B. Global Optimization Techniques

Due to the complexity of routing and channel assignment in WMNs, global optimization techniques are applied. Until now, over 90 different optimization techniques have been proposed, ranging from ant colony optimization to tabu search. We only describe four of them, which are used for the planning and optimization of wireless mesh networks, namely tabu search, branch and bound, simulated annealing, and genetic algorithms.

1) Tabu Search: Tabu search is an extension of the local search technique for solving optimization problems. The algorithm was introduced by Glover in 1986 [6]. It enhances the local search method by using a memory structure. To avoid cycles of the possible solutions found by the algorithm, the solutions are marked as “tabu”. All solutions on the tabu list can not be used for the next iteration step.

The tabu search algorithm starts by using either a random solution or by creating a solution with a heuristic. From this initial solution x, the algorithm iteratively moves to a solution x’ in the neighborhood of x. From all possible solutions in the neighborhood of x, the best one is selected as the new solution if it is not on the tabu list. Afterwards, the tabu list is extended and the algorithm proceeds until a stopping criterion is reached.

2) Branch and Bound: The branch and bound method generally finds the optimal solutions with the disadvantage of being slow. In general, it is a search and comparison of different possibilities based upon partition, sampling, and subsequent upper bounding procedures. The first step, the branch, is used to split a problem into two or more subproblems. The iteration of the branch step creates a search tree. To avoid the calculation of all subtrees, the algorithm uses the bound step. It searches for the first valid solution whose value is the upper bound. All following calculations are canceled if their costs exceed the upper bound. If a new, cheaper solution is found, the upper bound will be set to the value of this new solution. Thus, the branch step increases the search space while the bound step limits it. The algorithm proceeds until either all subtrees have been evaluated or a threshold is met.

3) Simulated Annealing: The goal of simulated annealing is to find a good solution rather than to find the optimal solution like branch and bound. The name of the algorithm comes from metallurgy. Metal is heated up and then cooled down very slowly. The slow cooling allows to form larger crystals, which corresponds to finding something nearer to a global minimum-energy configuration.

When applying simulated annealing for the channel allocation in a WMN, the algorithm starts assigning channels randomly. If a small change in the channel assignment improves the throughput, i.e., lowers the cost or energy, the new solution is accepted and if it does not improve the solution it might be accepted based on a random function. If the change only slightly worsens the solution, it has a better chance to get accepted in contrast to a solution, which heavily decreases the performance. Worse solutions are accepted with a probability given by the Boltzmann factor

$$e^{-\frac{E}{k_B \cdot T}} > R(0,1), \quad (1)$$

where E is the energy, k_B is the Boltzmann constant, T is the temperature, and $R(0,1)$ is a random number in the interval [0,1]. This part is the physical process of annealing. For a given temperature, all solutions are evaluated and then, the temperature is decremented and the entire process repeated until a stable state is achieved or the temperature reaches zero. This means that worse solutions are accepted with a higher probability when the temperature is high. As the algorithm progresses and the temperature decreases, the acceptance criterion gets more and more stringent.

4) Genetic Algorithms: Genetic algorithms are similar to simulated annealing and are also not applied to find the optimal solution but rather good ones. In contrast to the branch and bound method, they are much faster and therefore applicable for the planning and optimization of large wireless mesh networks.

GAs are based on the idea of natural evolution and are used to solve optimization problems by simulating the biological cross of genes. A randomly created population of individuals represents the set of candidate solutions for a specific problem. The genetic algorithm applies a so-called fitness function to each individual to evaluate its quality and to decide whether to keep it in the new population. However, the selection without any other operation will lead to local optima. Therefore, two operators, crossover and mutation, are used to create new individuals. These new individuals are called progenies. Figure 2 shows the influence of crossover and mutation on the fitness landscape of two traits. As mutation is just a swapping of one or two bits, it leads to only small changes in the fitness landscape. The crossover operator instead can lead to complete new solutions as indicated in the figure with the creation of the progeny. Thus, the crossover operator can protect the genetic algorithm from running into local optima, while a mutation is just a small step around a previous solution. Both operators together are used to find a near-optimal solution.

Simulated annealing and genetic algorithms are well suited for the planning of wireless mesh networks. Applying tabu search and the branch and bound algorithm would be too time consuming, especially when considering large WMNs. In the next section, we take a closer look at the work related to WMN planning and optimization where one of the described optimization methods is applied.

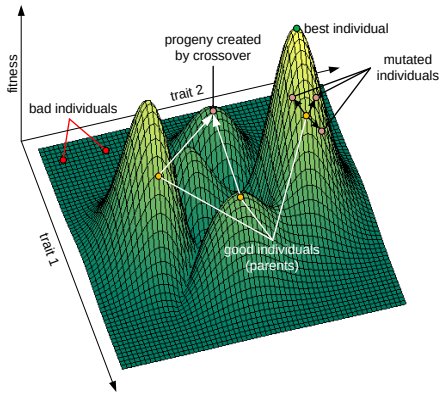


Figure 2. Influences of crossover and mutation in the fitness landscape.

III. Related Work

Wireless mesh networks have attracted the interest of various researchers and Internet providers. Hence, a number of papers have been published on the problem of planning WMNs and estimating their performance. We divide the related work into three parts. The first part shows general WMN planning approaches. In the second part, the work related to channel assignment and routing is presented. Finally, we present papers working with genetic algorithms for planning radio networks.

A. Wireless Mesh Network Planning Using Optimization Techniques

Sen and Raman [7] introduce a variety of design considerations and a solution approach, which breaks down the WMN planning problem into four tractable parts. These subproblems are inter-dependent and are solved by heuristics in a definite, significant order. The evaluations of the presented algorithms show that they are able to generate long-distance WLAN deployments of up to 31 nodes in practical settings.

Other related works [8]–[10] deal with creating a wireless mesh network model, planning its parameters, and evaluating the solutions via linear programming. He et al. [8] propose mechanisms for optimizing the placement of integration points between the wireless and wired network. The developed algorithms provide best coverage by making informed placement decisions based on neighborhood layouts, user demands, and wireless link characteristics. Amaldi et al. [9] propose other planning and optimization models based on linear programming. The aim is to minimize the network installation costs by providing full coverage for wireless mesh clients. Thereby, traffic routing, interference, rate adaptation, and channel assignment are taken into account. Another cost minimizing, topology planning approach is presented by So and Liang [10]. An optimization framework is proposed, which combines a heuristic with Bender's decomposition to calculate the minimum deployment and maintenance cost of a given heterogeneous wireless mesh network. Furthermore, an analytical model is presented to investigate whether a particular relay station placement and channel assignment can satisfy the user demands and interference constraints.

B. Routing and Channel Assignment for Wireless Mesh Networks

One of the first contributions on channel assignment is presented by Raniwala et al. [11], [12]. The channels are assigned according to the expected load evaluated for shortest path and randomized multi-path routing. It is shown

that by using only two network interface cards per mesh point, the throughput increases up to eight times. In contrast to Raniwala et al. [11], [12], Chen et al. [13] do not only consider the expected load for the channel assignment, but also consider the link capacities. Based on the link metrics, called expected-load and expected-capacity, the channel assignment is optimized using simulated annealing.

Further papers based on the paper presented by Raniwala et al. [11] are published by Ramachandran et al. [14] and Subramanian et al. [15]. Both papers take the interference between links into account. The first paper solves the channel assignment using a straightforward approach while the second one uses a tabu search algorithm. Another paper on channel assignment and routing is presented by Alicherry et al. [16]. A linear programming based routing algorithm is shown, which satisfies all necessary constraints for the joint channel assignment, routing, and interference free link scheduling problem. Using the algorithm, the throughput is fairly optimized. The fairness constraint means that for each node the demands are routed in proportion to the aggregate traffic load.

Raniwala and Chiueh [12] and Chen et al. [13] only consider non overlapping, orthogonal channels. Mohsenian Rad and Wong [17], [18] instead also consider partially overlapping channels and propose a congestion-aware channel assignment algorithm. It is shown that the proposed algorithm increases the aggregate throughput by 9.8 % to 11.4 % and reduces the round trip time by 28.7 % to 35.5 % compared to the approach of Raniwala and Chiueh [12].

C. Genetic Algorithms for Radio Network Planning

Genetic algorithms have been used for radio network planning for years [19]–[23]. Calégari et al. [19] apply a genetic algorithm for UMTS base station placement in order to obtain a maximum coverage. It is claimed that the performance of the GA strongly depends on the fitness function. Another paper on UMTS optimization with genetic algorithms was published by Ghosh et al. [20] in 2005. Genetic algorithms are used to minimize the costs and to maximize the link availability of a UMTS network with optical wireless links to the radio network controllers.

Besides Gosh et al. [20], Badia et al. [21] use genetic algorithms for a joint routing and link scheduling for WMNs. The packet delivery ratio is optimized depending on the frame length. It is shown that genetic algorithms solve the studied problems reasonably well, and also scale, whereas exact optimization techniques are unable to find solutions for larger topologies. The performance of the GA is shown for a single-rate, single-channel, single-radio WMN.

Vanhatupa et al. [22], [24] apply a genetic algorithm for the WMN channel assignment. Capacity, AP fairness, and coverage metrics are used with equal significance to optimize the network. The routing is fixed, using either shortest path routing or expected transmission times. An enormous capacity increase is achieved with the channel assignment optimization. Compared to manual tuning, the algorithm is able to create a network plan with 133 % capacity, 98 % coverage, and 93 % costs, while the algorithm needs 15 minutes for the optimization whereas the manual network planning takes hours.

Lee et al. [23] perform an AP assignment for users in smart environments using a genetic algorithm. The AP assignment is optimized in such a way that the load is balanced between the AP and that the bandwidth requirements can be met. The approach is evaluated in a scenario with 16 APs and 70 users. The results show that the load is almost balanced between the APs after 300 generations.

In contrast to the related work, we focus not only on a single-radio or single-rate WMN, but evaluate the performance of a multi-channel, multi-radio, multi-rate WMN using both channel and route assignment. Our genetic algorithm optimizes the throughput while still maintaining a max-min fair throughput allocation between the nodes. In the next section, the complexity of a fair resource allocation in WMNs is described before introducing genetic algorithms and its modifications for the planning and optimization of WMNs.

IV. WMN Planning Using Genetic Algorithms

The objective of this paper is to support the WMN planning process by optimizing the performance of a WMN. With the help of genetic algorithms, near-optimal solutions can be achieved in relatively small computation time. In this section, we show the parameters, which we have to consider and to evaluate in order to achieve a near-optimal WMN solution, meaning that the throughput in the WMN is fairly shared among the mesh points.

A. Problem Formulation

We assume that each mesh point is connected to only one gateway with a fixed routing and we can thus define the mesh network as a directed graph $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of mesh points $n_1, ..., n_V$ and $\mathcal{E} = \mathcal{L}$ is a set of links connecting the mesh points. A subset $GW \subseteq \mathcal{V}$ contains the gateways, which are connected to the Internet. Each mesh point $n_i \in \mathcal{V} \setminus GW$ has a fixed route and gateway to the Internet. The route is denoted as $\mathcal{R}_i$ and consists of a set of links, $\mathcal{R}_i \subseteq \mathcal{L}$. Thus, the mesh points connected to one gateway can be considered as a tree and the complete WMN as a forest.

As we do not have a fully meshed network, a link (i, j) between mesh point i and mesh point j only exists, if a communication between these mesh points is possible within the mesh network. Let $dr_{i,j}$ be the data rate of the link (i, j). The goal is now to optimize the paths from each mesh point $n_i \in \mathcal{V} \setminus GW$ to the gateway so that the throughput in the WMN is fairly shared among the mesh points.

B. Fairness and Capacity in Wireless Mesh Networks

To achieve a fair resource distribution among the mesh points, we use a max-min fair share approach introduced by Bertsekas and Gallager [4]. A solution is max-min fair if no rate can be increased without decreasing another rate to a smaller value. Max-min fairness is achieved by using an algorithm of progressive filling. First, all data rates are set to zero. Then, the data rates of all flows are equally increased until one flow is constrained by the capacity set. This is the bottleneck flow and all other flows have to be faster than this one. Afterwards, the data rates of the remaining flows are increased equally until the next bottleneck is found. This procedure is repeated until all flows are assigned a data rate.

Before assigning the data rates to the flows, the capacity of the network has to be estimated. Therefore, we first have to estimate the link capacities. The capacity of a single link is determined by the pathloss and the Signal to Noise Ratio (SNR). For the pathloss calculation, we use a modified COST 231 Hata [25] pathloss model for carrier frequencies between 2 GHz and 6 GHz. The model is proposed by the IEEE 802.16 working group as the WiMAX urban macrocell model, but is also valid for WLAN mesh networks and is defined as

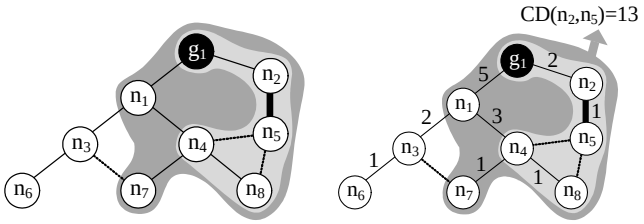
$$PL = 35.2 + 35 \cdot log_{10}(d(n_i, n_j)) + 26 \cdot log_{10}\left(\frac{f}{2}\right). \quad (2)$$

Here, f denotes the operating frequency and d denotes the euclidean distance between mesh points n_i and n_j. The pathloss model is used to calculate the SNR, which is required to determine the maximum achievable throughput. The SNR is calculated as

$$\gamma_{n_i,n_j} = T_x - PL(n_i, n_j, f) - (N_0 + 10 \cdot log_{10}(W)), \quad (3)$$

where T_x is the transmit power, N_0 is the thermal noise spectral density (-174 dBm/Hz), and W is the system bandwidth. Now, the Modulation and Coding Scheme (MCS) is selected with an SNR requirement γ^*_{mcs} that is smaller or equal to the link's SNR γ_{n_i,n_j}. The MCS is chosen in such a way that the frame error rate lies below 1 %. If the SNR requirement for the most robust MCS cannot be met, the two mesh points n_i and n_j are not within communication and interfering range.

Having computed the maximum data rate of each link according to the pathloss, we now have to calculate the capacity of each link taking interference from neighboring mesh points into account. Therefore, we use the concept of Collision Domains (CDs) introduced by Jun and Sichitiu [26]. The collision domain $\mathcal{D}_{i,j}$ of a link (i, j) corresponds to the set of all links (s, t), which can not be used in parallel to link (i, j) because the interference from a transmission on link (s, t) alone is strong enough to disturb a parallel transmission on link (i, j). Figure 3(a) shows the collision domain of link (n_2, n_5). The one-hop collision domain illustrated in light-gray denotes the area for a WLAN-based mesh network without using RTS/CTS. The dark gray area shows the two-hop area where no station can transmit a packet when using RTS/CTS.



(a) One- and two-hop collision domain of link (n_2, n_5).

(b) Link load calculation depending on the carried number of flows.

Figure 3. Collision domain and its link loads.

The nominal load of such a collision domain is the number of transmissions taking place in the collision domain. A transmission $tr_{k,i,j}$ corresponds to the hop from mesh point n_i to mesh point n_j taken by the flow towards mesh point k, i.e., $(i, j) \in \mathcal{R}_k$. The number of transmissions $\lambda_{i,j}$ on link (i, j) corresponds to the number of end-to-end flows crossing it:

$$\lambda_{i,j} = \Big|\{k|(i, j) \in R_k\}\Big|. \quad (4)$$

Figure 3(b) shows the load per link for the same example network as before. Each mesh point on the way to the gateway produces traffic resulting in a traffic load of 5 on the link (n_1, g_1) and a load of 2 on the link (n_2, g_1). Correspondingly, the number of transmissions in collision domain $\mathcal{D}_{i,j}$ is

$$m_{i,j} = \sum_{(s,t)\in\mathcal{D}_{i,j}} \lambda_{s,t}. \quad (5)$$

Thus, the collision domain of link (n_2, n_5) consists of 13 transmissions in total.

In order to fairly supply all mesh points, we share the time resources among all transmissions taking place within the collision domains of the corresponding links. Thereby, we take the rates $dr_{i,j}$ and the number of flows $\lambda_{i,j}$ into account. The throughput $t_{i,j}$ of link i, j is then defined as

$$t_{i,j} = \frac{1}{\sum\limits_{(s,t)\in\mathcal{D}_{i,j}} \frac{\lambda_{s,t}}{dr_{s,t}}}. \quad (6)$$

If we assume that link (n_2, n_5) supports 54 Mbps based on the pathloss and the SNR, the throughput would be 4.15 Mbps due to a collision domain size of 13. However, before setting this throughput to node n_5 we have to follow the principle of max-min fairness.

An algorithm to determine the max-min fair throughput allocation based on the definition of collision domains is given by Aoun and Boutaba [27]. The algorithm iteratively determines the bottleneck collision domain and allocates the

data rates of all flows traversing this domain. If in our example in Figure 3 the link (n_1, g_1) would be the bottleneck, all mesh points traversing the link would be assigned to this throughput, in our case n_3, n_4, n_6, n_7, n_8. As link (n_2, n_5) and link (g_1, n_2) also belong to the collision domain of link (n_1, g_1) but do not transmit over the bottleneck link, the time resources occupied by the bottleneck link are subtracted from the two links.

In the next step of the iteration process, only the remaining collision domains are considered. This way, we calculate the throughput of each flow, which is needed to evaluate the fitness of the WMN. The iteration stops when all flows are assigned. If in our example the next bottleneck collision domain is link (g_1, n_2), the remaining maximum supported rates are assigned to the last two links. Algorithm 1 clarifies the procedure of assigning the rates.

Algorithm 1 Max-min fair resource distribution based on collision domains.

1: $\mathcal{O} = \mathcal{F}$ — all flows are unassigned
2: $\mathcal{L} = \{(i,j) | n_{i,j} > 0\}$ — all active links
3: $p_{i,j} = 1,\ (i,j) \in \mathcal{L}$ — all links have full capacity
4:
5: *Iteration*
6: **for all** links $(i,j) \in \mathcal{L}$
7: $m_{i,j} = \sum_{(s,t) \in \mathcal{D}_{i,j}} \lambda_{s,t}$ — nominal load
8: $t_{i,j} = \frac{1}{\sum_{(s,t) \in \mathcal{D}_{i,j}} \frac{\lambda_{s,t}}{dr_{s,t}}}$ — throughput share per flow
9: **end for**
10: $(u,v) = \arg\min_{(i,j) \in \mathcal{L}} t_{i,j}$ — bottleneck CD
11: $\mathcal{B} = \{k \in \mathcal{O} | \mathcal{R}_k \cap \mathcal{D}_{u,v} \neq \emptyset\}$ — bottleneck flows
12: $b_k = r \cdot t_{u,v}$ for all $k \in \mathcal{B}$ — set bottleneck rates
13: $\mathcal{O} = \mathcal{O} \backslash \mathcal{B}$ — adapt unassigned flows
14: $p_{i,j} = p_{i,j} - \sum_{k \in \mathcal{B}} |\mathcal{R}_k \cap \mathcal{D}_{i,j}| \cdot t_{u,v}$ — adapt free capacity of all CDs
15: $\mathcal{L} = \mathcal{L} \backslash \mathcal{D}_{u,v}$ — adapt active links
16: *Stop criterion:* $\mathcal{O} = \emptyset$

C. Optimization Using Genetic Algorithms

After describing the principle of collision domains and max-min fair throughput allocation, we now explain the workflow of a genetic algorithm in detail. Figure 4 shows the complete procedure of a genetic algorithm for the planning and optimization of WMNs. Firstly, a random population is created with a predefined number of individuals. The fitness of each individual is evaluated using the fitness function and the individuals are ordered according to the fitness value. The best individuals, the elite set, is kept for the new population. Afterwards, the crossover and mutation operator are used to create the remaining number of individuals for the new population. The procedure is repeated until a sufficient solution is achieved. In the following, we explain the steps of our WMN optimization approach in more detail.

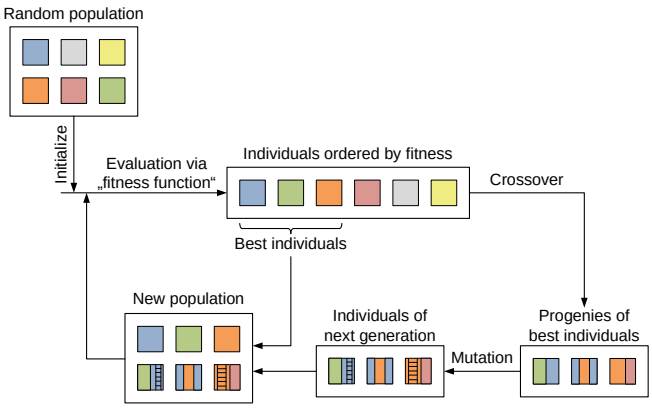
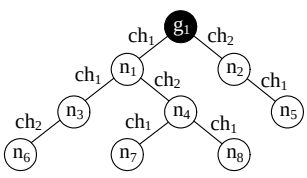


Figure 4. Workflow of a genetic algorithm.

1) Network Encoding: Before going through the steps of the genetic algorithm, the WMN has to be encoded. The encoding must be simple without any redundancy in order not to prolong the runtime of the genetic algorithm. As we assume that each mesh point is connected to only one gateway, the network encoding has to represent a spanning tree with the gateway as root, cf. Figure 5(a). This means that the graph does not contain any cycles and each mesh point has only one route towards the gateway. Such a tree structure can easily be encoded in a list, where the next hop of each mesh point, which the traffic has to take in order to reach the gateway, is stored. This list representation of the example network from Figure 5(a) is shown in Figure 5(b). Considering for example mesh point n_4, the next hop is node n_1 and the next hop of mesh point n_1 is the gateway. Thus, the complete routing of a WMN is handled with a simple list representation.

Besides the routing, we also want to optimize the channel allocation. Although each mesh point can be equipped with several network interface cards, the channel of the link towards the gateway is fixed as shown in Figure 5(a). Thus, the channel allocation can be done in a similar way as the routing. Therefore, the list is extended with one more row, showing the channel of the next hop towards the gateway, cf. Figure 5(b). This simple list represents the tree structure of



(a) Example network.

node	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8
next hop	g_1	g_1	n_1	n_1	n_2	n_3	n_4	n_4
channel	ch_1	ch_2	ch_1	ch_2	ch_1	ch_2	ch_1	ch_1

(b) List representation.

Figure 5. Example network and its list representation.

one gateway and each gateway in the wireless mesh network is encoded in a similar way. The list representation is later used to perform the genetic operations and to evaluate the fitness of the WMN.

2) Evaluation via Fitness Function: The evaluation part of the optimization is the heart of the genetic algorithm. Based on the fitness value, the GA decides, which individuals should be kept in the new population. Hence, it rates the performance of the genes and allows only the best to be replicated.

The fitness of the WMN is estimated using the allocated throughputs of each flow. The fitness function $f(\mathcal{N})$ of the evaluation represents the user satisfaction and the fairness of the resource allocation. Some fitness functions might lead to a complete unfair resource distribution in the WMN. Therefore, we evaluate the performance of several different fitness functions in Section V. Several combinations of the functions $min(\mathcal{R}_\mathcal{N})$, $median(\mathcal{R}_\mathcal{N})$, $mean(\mathcal{R}_\mathcal{N})$, and $var(\mathcal{R}_\mathcal{N})$ are used, which are applied on all routing links of a network solution $\mathcal{N}$. The function $min\,(\mathcal{R}_\mathcal{N})$ calculates for example the minimum throughput of all links used in routing scheme $\mathcal{R}_\mathcal{N}$. We define the following eight different fitness functions:

$$
\begin{aligned}
f_1(\mathcal{N}) &= min(\mathcal{R}_\mathcal{N}) = \text{minimum throughput}(\mathcal{R}_\mathcal{N}) \\
f_2(\mathcal{N}) &= median(\mathcal{R}_\mathcal{N}) = \text{median throughput}(\mathcal{R}_\mathcal{N}) \\
f_3(\mathcal{N}) &= mean(\mathcal{R}_\mathcal{N}) = \text{mean throughput}(\mathcal{R}_\mathcal{N}) \\
f_4(\mathcal{N}) &= min(\mathcal{R}_\mathcal{N}) + \frac{median(\mathcal{R}_\mathcal{N})}{s} \\
f_5(\mathcal{N}) &= mean(\mathcal{R}_\mathcal{N}) - var(\mathcal{R}_\mathcal{N}) \\
f_6(\mathcal{N}) &= min(\mathcal{R}_\mathcal{N}) + \frac{median(\mathcal{R}_\mathcal{N})}{s} + \frac{mean(\mathcal{R}_\mathcal{N})}{|\mathcal{L}|} \\
f_7(\mathcal{N}) &= \sum_{i=0}^{|\tilde{T}|-1} (\left|\tilde{T}\right| - i) \cdot \tilde{T}(i) \\
f_8(\mathcal{N}) &= \sum_{i=0}^{|\tilde{T}|-1} c^{|\tilde{T}|-i} \cdot \tilde{T}(i).
\end{aligned}
$$

The last two functions weight the link throughputs with a factor depending on the corresponding throughput value. Therewith, we aim to achieve a kind of max-min fairness not only with the throughput allocation made by the evaluating algorithm but also with the fitness value from a reasonable fitness function. For this purpose, an ascendingly sorted list $\tilde{T}$ of the throughputs of all routing links in the solution $\mathcal{N}$ is used. Each throughput value from $\tilde{T}$ is weighted with a factor depending on its place in the list, giving more weight to lower positions. This results in a fitness value with which mainly smaller link throughputs are optimized at the expense of higher ones. The parameter c of function $f_8(\mathcal{N})$ is a constant, which we set to 1.5 and s is set to 8 for the experiments in Section V.

3) Selection Principle: After the evaluation of a population, we select a set of solutions, which have the highest fitness of all and keep them in the new generation. This set is called the elite set. In the results section, we vary the size of the elite set in order to see the influence on the solution. As the number of individuals of a population is fixed for all generation steps, the remaining number of individuals are created by crossing and mutating the genes.

The selection of the individuals for applying the genetic operators is thereby based on the fitness and furthermore depends on the number of needed new individuals. Let w be the number of needed new individuals and $s(x)$ be the selection probability for individual x. Then, the number of progenies generated based on individual x are

$$g(x) = \|w \cdot s(x)\|\,. \tag{7}$$

The selection probability $s(x)$ depends on the relation between the fitness of solution x and the sum of all fitness values from the complete population, which means that new individuals are more likely to be created from individuals with a better fitness. This results in

$$s(x) = \frac{f(x)}{\sum_{j=1}^{n} f(j)}. \tag{8}$$

4) Crossover Types: The crossover operator as well as the mutation operator are now applied to the selected number of individuals. For the cross of genes, we use the standard 2-Point Crossover [28] and two other variants, which we especially created for the planning of WMNs, the Cell and the Subtree Crossover.

2-Point Crossover

The 2-Point Crossover is a widely used extension of the 1-Point Crossover. While the 1-Point Crossover changes the list representations of two individuals until a certain point or from a certain point on, the 2-Point Crossover exchanges subsets, which are randomly chosen sublists of the individuals representation, the genotype. Thus, a start and an end point, denoting the range of the sublist, are chosen each time the 2-Point Crossover is applied.

An example of the crossover is shown in Figure 6. The sublists of two individuals should be crossed, namely the routing and channel allocation of mesh points n_2 to n_5. The resulting progenies of the individuals show one characteristic of this reproduction approach. It created solutions, which contain mesh points with no connection to any gateway. This happens due to the unregulated and absolutely arbitrary selection of the gene subset, which is meant to be exchanged.

Looking at the progeny of individual 2, mesh points n_1, n_2, n_6, n_7, n_8 have no connection to any gateway and thus, the crossover results in an unreasonable solution.

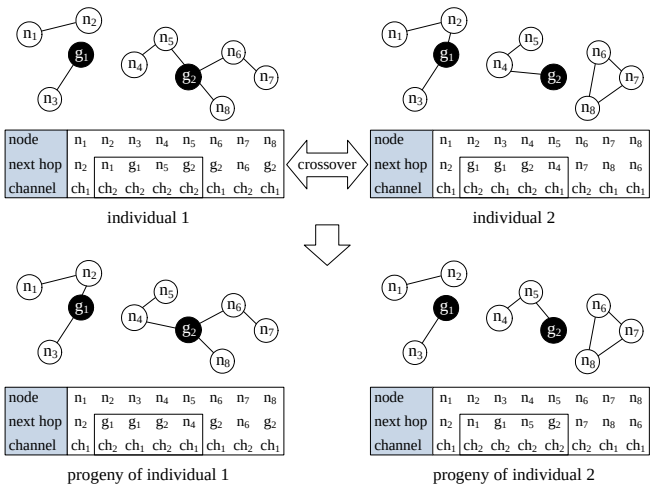


Figure 6. 2-Point Crossover between two individuals.

On the other hand, the 2-Point Crossover has created a reasonable progeny of individual 1.

Since the 2-Point Crossover may lead to unconnected solutions, we have to be careful when evaluating the fitness of the resulting solutions. Thus, we adapt the fitness function to

$$\tilde{f}(\mathcal{N}) = f(\mathcal{N}) - diss(\mathcal{V}), \tag{9}$$

which includes now the $diss(\mathcal{V})$ term denoting the number of nodes with no connection to any gateway. Hence, the throughput contained in $f(\mathcal{N})$ presents the positive costs of the network while $diss(\mathcal{V})$ stands for the penalty costs.

Cell Crossover

In contrast to the 2-Point Crossover, the Cell Crossover does not exchange sublists but complete cells. The crossover operator randomly chooses a gateway and exchanges the entire cell meaning that the routing information as well as the channel allocation is exchanged.

Figure 7 shows an example for the crossover of two solutions. Black nodes denote the network gateways and the light gray areas mark the chosen cell, which is exchanged. In the resulting progenies, the mesh points that have changed their connection are marked dark gray. We can see that not only link connections from mesh points are crossed, but some mesh points are now also connected to other gateways. Mesh points $n_{10}, n_{12}, n_{17}, n_{18}$ are connected to gateway g_2 in the progeny of individual 2 while they were attached to gateway g_1 before. The reason is that the number of mesh points belonging to one cell differ between the individuals. Therefore, we also have to attach unconnected nodes after the Cell Crossover, which can be seen in the progeny of individual 1. In addition to the exchange of routes, the assigned channels are exchanged, which is not shown in the figure for the sake of readability.

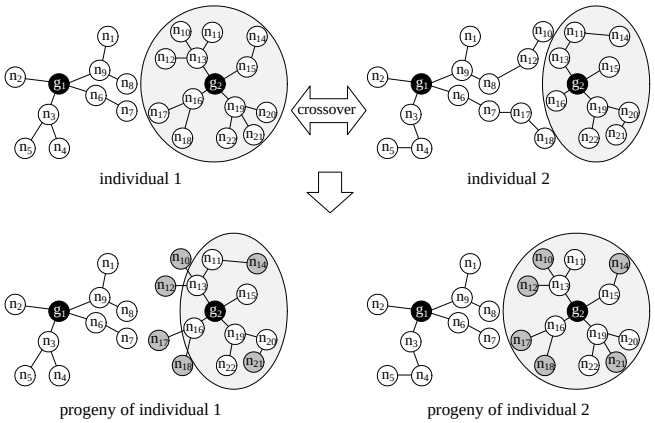


Figure 7. Cell Crossover between two individuals.

Subtree Crossover

The last crossover type is the Subtree Crossover. In contrast to the Cell Crossover, not a complete gateway tree is exchanged but only a subtree. Therefore, the Subtree Crossover chooses mesh points randomly and crosses the entire subtree with the mesh point as root. Similar to the Cell Crossover, the channel allocation is exchanged together with the routing information.

The Subtree Crossover of two subtrees is shown in Figure 8. The chosen mesh points are n_3 and n_{13}. The crossover of subtree n_3 only causes a small change in the tree structure in contrast to the subtree crossover of n_{13}. Here, some nodes of the subtree are connected to different gateways in the two individuals. After the crossover, mesh points n_{10} and n_{12} belong to gateway g_2 in the progeny of individual 2. This reduces the number of long branches of gateway g_1 but there is still potential for further optimization.

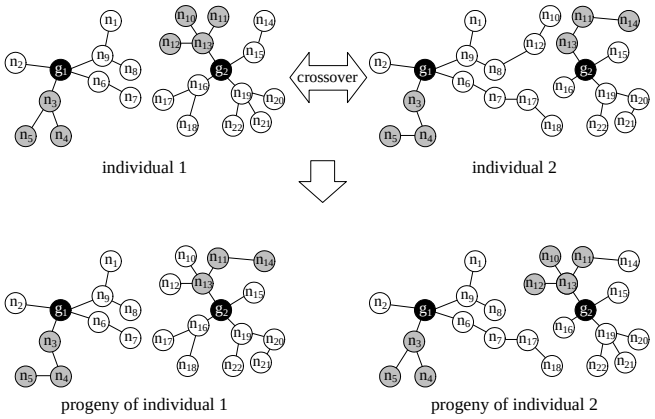


Figure 8. Subtree Crossover between two individuals.

D. Mutation

While the different crossover variants help to avoid running into local optima, the mutation operator increases the

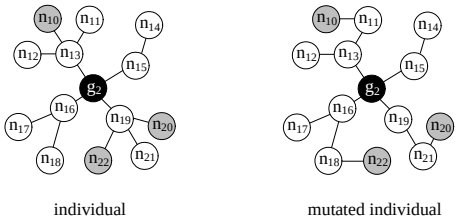


Figure 9. Routing mutation of three mesh points.

performance of WMNs with slightly modifications of the routing structure and channel allocation. For the optimization of WMNs, the number of mutations are chosen based on the scenario size and the mutation of the routing and channel allocation are applied independently from each other.

For the routing scheme, the mutation operator substitutes some randomly chosen positions of the routing code with new information taken from a set of potential neighbors, which would not cause the creation of cycles and would not harm the tree structure of the solution. An example for the mutation of the routing scheme from three nodes is shown in Figure 9. Here, the links towards the gateway of the three gray nodes are mutated. For the channel allocation, the mutation operator randomly chooses a channel from a list of possible channels and substitutes randomly chosen links from the WMN.

According to the workflow diagram shown in Figure 4, the mutation operator is applied after the crossover on the progenies of the crossover. The mutated individuals together with the elite set form then the new population and close the circle of the genetic algorithm.

V. Performance Evaluation

After introducing genetic algorithms in detail and showing our modifications and extensions for wireless mesh networks, we now want to evaluate the performance of the genetic algorithm. The influence of every part of the genetic algorithm's workflow is thereby evaluated separately. First, we take a look at the influence of the fitness function on the resulting solution. Afterwards, the size of the elite set is investigated followed by the population evolution for the three different crossover types. Finally, we show the influence of the two genetic operators crossover and mutation on the resulting network solution.

A. Simulation Settings

For the creation of the results presented in this section, we use the two scenarios introduced in Table I. Although we evaluated a large number of different scenarios, we highlight only the two most different ones here. The first one consists of 2 gateways and 71 mesh points distributed over an area of 2 km x 1.2 km. Thereby, the minimal distance between mesh points is 60 m and between the two gateways it is 700 m. For the sake of readability, we call this scenario G2MP71.

Table I
Simulation Scenarios.

Parameter	Scenario S1	Scenario S2
Topology	G2MP71	G6MP38
Population size	150	
Elite set size	50	
Number of generations	400	
Crossover type	Subtree Crossover Cell Crossover 2-Point Crossover	
Number of crossed subtrees	rand(0,7)	rand(0,5)
Number of mutations	rand(0,20)	rand(0,10)
Fitness function	$f_1(\mathcal{N})$	

The second scenario contains a smaller number of mesh points and a larger number of gateways. We choose this clearly different topology in order to show the influence of the crossover operators depending on the number of mesh points. The 38 mesh points and 6 gateways of the second scenario are allocated in an area of 1.5 km x 1 km. The minimal distance between users is 60 m and between gateways 450 m. We call this scenario G6MP38.

The differences in the settings of the two configurations depend on the used topology of the corresponding scenario. Due to the larger number of mesh points contained in G2MP71, we configure Scenario S1 with more mutations and more exchanged subtrees than Scenario S2. Thereby, we keep the relation between crossover and mutation at a fixed level suitable for the investigation of the genetic operators.

Besides the parameters of the genetic algorithm, the general parameter settings are shown in Table II. These parameters only affect the characteristics of the network connections. The parameters carrier frequency, channel bandwidth, and available channels decide to some extent the performance of the mesh point connections in a network solution but they do not have an impact on the effectiveness of the genetic algorithm. Therefore, we do not consider their impact on the resulting solutions.

Table II
General Parameter Settings.

Parameter	Value
Carrier frequency	3500 MHz
Channel bandwidth	20 MHz
Maximum throughput	67.2 Mbps
Available channels	3500 MHz, 3510 MHz
Antenna power	25 dBm
Pathloss model	WiMAX urban macrocell model

B. Influence of Fitness Function

As the fitness function is the heart of the genetic algorithm, we first take a look on the influence of different fitness functions on the resulting solution. Therefore, eight different fitness functions, described in Section IV, are applied.

Figure 10 shows the throughputs of the mesh points of the best individual after 400 generations of Scenario S1. For the sake of readability, the curves of the eight different fitness functions are shown in two separate subfigures. The x-axis shows the normalized flow IDs, meaning the 71 mesh points sorted by throughput, and the y-axis lists the throughput in Mbps of the flows.

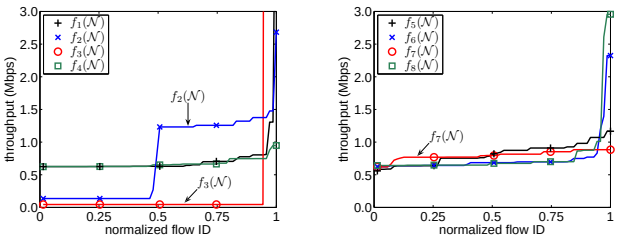
A curve completely parallel to the x-axis would mean a perfect fairness between all flows and a curve whose minimum throughput is above $f_1(\mathcal{N})$ would mean that the solution is max-min fair. This allows to see that the unfairest resource distributions are achieved with the fitness functions $f_2(\mathcal{N})$ and $f_3(\mathcal{N})$.

Optimizing only the median with $f_2(\mathcal{N})$, we do not pay attention to the rest of the throughput allocation. This is why the left part of the $f_2(\mathcal{N})$ curve stays very low. The distribution of $f_2(\mathcal{N})$ also shows that some mesh points have a very high throughput compared to others. This happens accidentally because the fitness function does not control their behavior as it focuses just on the throughput of the median.

Fitness function $f_3(\mathcal{N})$, optimizing only the mean throughput, also results in a very unfair solution. Here, the number of hops towards the gateway are minimized in order to get some nodes with very high throughput, which boost the mean value. In this scenario, four mesh points have a throughput of over 24 Mbps while the throughput of all other flows is about 0.05 Mbps.

All other fitness functions result in a max-min fair resource distribution with a maximized minimal throughput. In the resulting solutions of $f_1(\mathcal{N})$, $f_6(\mathcal{N})$, and $f_8(\mathcal{N})$, some flows have a very high throughput but not at the costs of other flows.

The fairest solution is achieved with fitness function $f_7(\mathcal{N})$ where all flows have a similar throughput of about 0.7 Mbps. The fitness function weights the throughputs of the mesh points. Thereby, smaller throughputs have a stronger influence on the fitness than higher throughputs. This is achieved by multiplying the throughputs with the inverse of the ascendingly sorted flow ID.



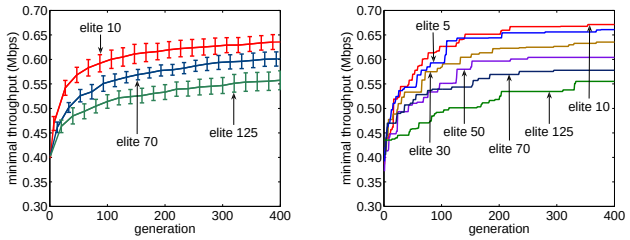
(a) $f_1(\mathcal{N})$ to $f_4(\mathcal{N})$. (b) $f_5(\mathcal{N})$ to $f_8(\mathcal{N})$.

Figure 10. Throughput allocation of the best individual.

C. Elite Set Size

In this section, we examine the impact of the elite set size on the progress of the evolution using Scenario S1 and applying the Subtree Crossover only. Figure 11(a) illustrates the minimal throughput of three different elite set sizes averaged over 15 different initial populations. This time, the x-axis shows the generation number while the y-axis lists the minimal throughputs.

From the figure it can be observed that the best performance is achieved with a small elite set size. On the one hand, a large elite set includes a number of bad individuals, which are kept in the next generation and decrease the minimal throughput. On the other hand, with an elite set size of 125, only 25 new progenies are generated. With this small number of new unexplored genes, the progress of the genetic algorithm slows down, which can be seen on the left side of the figure. Similar solutions compared to an elite set size of 10 might be achieved after several more generations. This means that the larger the elite set size is, the slower is the progress of the genetic algorithm. To prove this statement, we performed the optimization of the same scenario for more different elite set sizes. The results are shown in Figure 11(b).



(a) Three elite set sizes averaged over 15 seeds. (b) Performance of six elite set sizes.

Figure 11. Comparison of different elite set sizes.

The figure reveals almost the same behavior as the previous one. Smaller elite sets cause faster evolution and lead to better solutions. However, a too small elite set size is also bad as the figure shows for an elite set size of 5. With a too small elite set, there might be a discrepancy between the fitness of the elite set and the fitness of new progenies. Thus, the elite set size should be chosen in dependence of the population size.

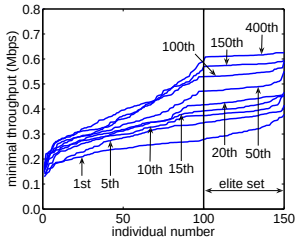
D. Population Evolution

Examining the evolution of the population is an important consideration needed to demonstrate the effectiveness of the genetic algorithm. Observing the evolution of the population with every generation step helps to decide when to terminate the algorithm. When the fitness is not increasing after an additional number of generations, the genetic algorithm can be stopped because either a near-optimal solution is found

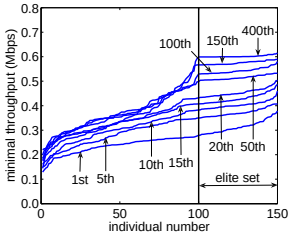
or the genetic algorithm is stuck in a local optimum. As the crossover operator helps to get out of a local optimum, we take a look at the population evolution for all three introduced crossover types.

The results shown in Figure 12 are generated with Scenario S1 from Table I. The x-axis shows the individuals sorted by fitness and the y-axis displays the minimal throughput of each individual. The different curves illustrate the generation progress during the genetic optimization. The elite set size is chosen to be one third of the complete population size.

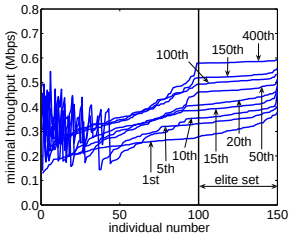
In order to compare all three crossovers, we did not plot the fitness but the minimal throughput on the y-axis. As the penalty costs are included in the fitness function of the 2-Point Crossover, cf. Section IV, the fitness values would be much lower for the 2-Point Crossover. Hence, we consider only the minimal throughputs, which only represent the positive costs. This is also the reason for the strongly varying curves on the left side of Figure 12(c). The individuals have a large minimal throughput but there are a lot of unconnected nodes, which result in a lot of penalty costs and thus in lower fitness.



(a) Subtree Crossover.



(b) Cell Crossover.



(c) 2-Point Crossover.

Figure 12. Generations progress using Subtree, Cell, and 2-Point Crossover.

In all subfigures, we can observe that the higher the generation number is, the smaller is the fitness growth. This slowdown is caused by the similarity of individuals. After several generations, the individuals are quite similar, which means that the crossover does not generate new, unexplored genes. The only possibility to find better solutions is to apply the mutation operator only. Therefore, we introduce the concept of local optimization in Section VI.

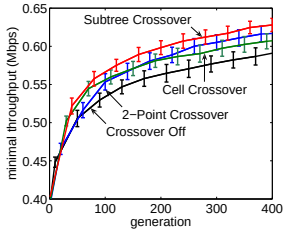
Evaluating the population evolution in other scenarios has shown that it highly depends on the topology structure but a good solution is always found after 400 generations. We tested the performance of Scenario S1 also after 1000 and 1500 generations, but the performance increase was negligible compared to the throughput after 400 generations.

A comparison of the three crossover types shows that the highest minimal throughput after 400 generations is achieved with the Subtree Crossover, followed by the results of the Cell Crossover. The network solution with the worst performance is achieved when applying the 2-Point Crossover. In the next subsection, we want to see if this is an exception or if the Subtree Crossover always leads to the best solutions.

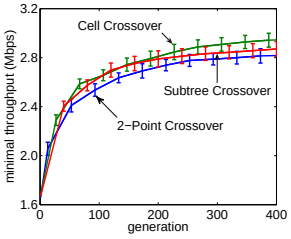
E. Effectiveness of Crossover

In order to show the effectiveness of the crossover type, we compare the performance of the three crossover operators depending on the number of mesh points and gateways in the network. Furthermore, we want to find out if there is an interaction between the efficiency of the crossover types depending on the topology.

The results for both scenarios from Table I are presented in Figure 13. Figure 13(a) shows the evolution of the best individual during 400 generations with different crossover types and for not using the crossover operator at all for Scenario S1. It illustrates the average results of 20 seeds while applying a 95% confidence interval.



(a) Scenario S1.



(b) Scenario S2.

Figure 13. Effectiveness of the crossover operator.

This scenario includes a high number of mesh points, which are distributed in the coverage areas of only two gateways. This results in deep tree structures with long ways over multiple hops towards the corresponding gateway. Such network structures seem to be crucial for the effectiveness of the crossover types. We can observe that the Subtree Crossover leads to a better solution than the other two crossover types. The better performance of the subtree approach is the result of the exchange of small connectivity components, which causes reasonable gene variations without disturbing the tree structure. The other two crossover types show a lower performance whereby the unregulated 2-Point Crossover even outperforms the intelligent Cell Crossover approach. This results from the small number of gateways, which causes the cross of only one cell per new progeny and quickly leads to similar individuals.

The results from Scenario S2 are shown in Figure 13(b). In contrast to the previous scenario, the higher number of available gateways leads to a better efficiency of the Cell Crossover. Moreover, the small number of nodes belonging to one gateway allows a larger variety of individuals. This is due to the fact that small changes in the routing structure cause higher changes in the network performance than in Scenario S1. However, the Cell and Subtree Crossover, which exchange only connectivity components have a better performance than the 2-Point Crossover.

The comparison of the crossover types shows that the crossover operator should be selected based on the considered topology to achieve the best solutions. In the next subsection, we take a look at the influence of the mutation operator on the evolution of the population.

F. Effectiveness of Mutation

The mutation operator causes small changes in the fitness landscape and normally does not help to get out of local optima. However, in the last subsection we have seen that applying only the mutation operator almost increases the performance of the wireless mesh network to the same level as compared to a scenario where both, crossover and mutation are applied. To investigate the influence of the mutation operator, Scenario S1 is considered. Both mutation operations, the routing and the channel mutations, are applied on the progenies of the crossed individuals. The number of routing and channel mutations on each individual are chosen randomly in the interval [0,20]. Figure 14 shows the minimal throughputs during the progress of the genetic algorithm for all three crossover types.

Surprisingly, the performance of the genetic algorithm without mutation is generally low and the genetic algorithm runs into a local optimum after a few generations. For the Cell Crossover, the reason is simple because only 2 gateways are placed in the scenario. The minimal throughputs for the other two crossover variants are higher compared to the Cell Crossover but still way below the throughputs achieved when mutation is used together with crossover. This is because after a few generations, the created individuals are quite similar and thus, the genetic algorithm gets stuck in a local optimum. In contrast, when activating the mutation operator, the fitness of the solution grows even after 400 generations and there is still potential for further evolution.

This shows how crucial the mutation operator is for the evolution of the genetic algorithm. Without using the mutation operator, similar individuals are created by the three different crossovers. The best performance is here seen for the Subtree Crossover as the Subtree Crossover has the largest possibilities to create new genes. The mutation operator instead ensures the creation of new unexplored genes with slight changes in the routing scheme and channel allocation, which fosters the evolution.

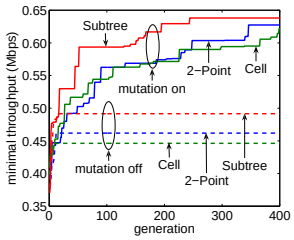


Figure 14. Mutation ON/OFF in combination with three crossover types tested on Scenario S1.

VI. Optimization of the WMN Planning Approach

In the last section, we have seen the influence of the genetic operators on the performance of the resulting wireless mesh network. In this section, we take a look at the influence of the genetic operators in dependence of the GA progress and introduce a local optimization technique to quickly improve the performance of the wireless mesh network.

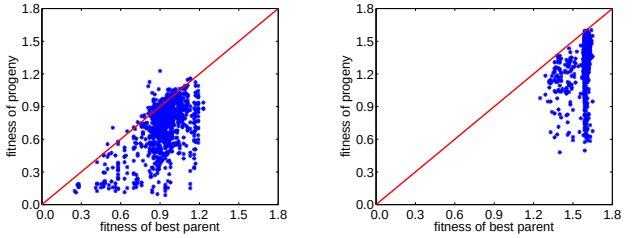
A. Influence of the Crossover on the GA Progress

As crossover operations are very time consuming, we want to see if the crossover types lead to better network solutions during all generations. Therefore, we compare the fitness of the best parent with the fitness of the resulting progeny for early generations as well as for late generations. The genetic optimization runs for 500 generations and the results in Figure 15 show the fitness of the Cell Crossover and Subtree Crossover of 2000 samples.

Looking at Figure 15(a), we can see that about 10 % to 20 % of all crossover operations lead to better progenies. Although this amount seems to be very low, we have to take a look at the exact improvements. One early Cell Crossover increases the fitness from 0.9 to 1.2. This might be a step out of two local optima in the fitness landscape. However, performing a Cell Crossover in the late stages of the genetic algorithm always leads to worse progenies. The reason is simple as a Cell Crossover of two near-optimal solutions are likely to create unreasonable progenies.

When applying the Subtree Crossover, the results are a little bit different as shown in Figure 15(c) and Figure 15(d). Although the percentage of better progenies is similar to the Cell Crossover, the improvements are lower. The reason is that the Subtree Crossover performs only small variations by exchanging subtrees, whereas the Cell Crossover changes two complete cells. However, these small changes also have a bad influence when performing them at the end of the generation process and only one or two progenies are better than their parents.

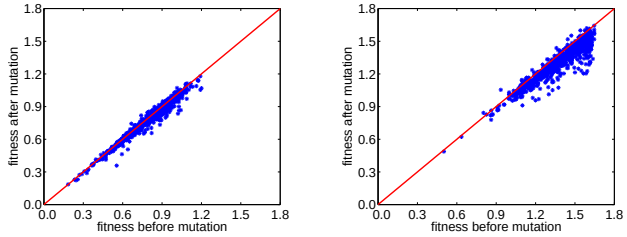
Thus, the amount of crossovers can be reduced with increasing number of generations. Before doing this, we take a look at the influence of the mutation operator in dependence of the number of generations.



(a) Early generation Cell Crossover. (b) Late generation Cell Crossover.

(c) Early generation Subtree Crossover. (d) Late generation Subtree Crossover.

Figure 15. Influence of the crossover on the fitness of the resulting progenies.



(a) Early generation routing mutation. (b) Late generation routing mutation.

(c) Early generation channel mutation. (d) Late generation channel mutation.

Figure 16. Influence of the mutation operator on the GA progress.

B. Influence of the Mutation Operator Depending on the GA Progress

The mutation operator conducts only small modifications of the individuals and it is thus expected that the fitness only slightly changes after the mutation is performed. Furthermore, we want to evaluate if, in contrast to the crossover, the mutation also leads to better results when applied on late generation steps. The results for the two mutation operators, routing and channel allocation, are shown in Figure 16. The plots are generated based on 2000 samples taken at the beginning and at the end of a 500 generation run.

As expected, the change in the fitness value is only small after the mutation is applied. However, the number of improved individuals is larger for both mutation operators compared to the crossover operations. The channel mutation even yields better results in 50 % of all mutations. Although the performance of both mutation operators decreases with an increasing number of generations, still better individuals are achieved in 5 % to 10 % of all mutations, cf. Figure 16(b) and Figure 16(d).

Thus, the mutation operator should be applied during the complete generation process. However, when reducing the number of crossover operations with an increasing number of generations, the number of performed mutations are also decreased. In order to keep the number of mutations, the following mechanism is applied. Firstly, the elite set size is increased with each generation, which means that increasingly more individuals are kept for the following population. This reduces the number of crossover and mutation operations. Secondly, in order to apply the mutation operator during the complete generation process, both mutation operations are performed with each individual of the elite set. If the fitness after the mutation is higher than the fitness before the mutation, the new individual is taken for the next population instead of the old one. If the fitness is worse, the new individual is discarded.

In Figure 17, we compare the fitness values of the ten best individuals in a scenario with an enlargement of the elite set size with increasing generation number and without an enlargement. The values are averaged over 10 simulation runs with 500 generations. Except for the worst of all 10 individuals, the enlargement of the elite set has a positive influence on the fitness. On average, the fitness is increased by 8 %.

Summarizing, a reduction of the number of crossover operations achieved by a stepwise enlargement of the elite set size has a positive effect on the fitness value. In addition, the runtime of the genetic algorithm is reduced due to the smaller number of complex calculations of the crossover operations.

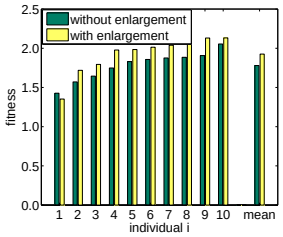


Figure 17. Influence of the enlargement of the elite set.

C. Local Optimization and Population Size

As we have seen in the previous figures, applying the crossover operator on late generations almost always results in worse individuals. However, the mutation operator might improve the individuals because it only slightly changes the individuals. To take advantage of this, we introduce the concept of local optimization. After the normal genetic algorithm finishes, we take the five best individuals of the last generation, copy them three times, and perform several mutations with them. Similar to the previous improvement, the resulting individual is only kept if its fitness value is higher compared to the fitness value before the mutation, else it is discarded. This can be repeated more than a thousand times because the computation time for mutating 15 individuals is negligible.

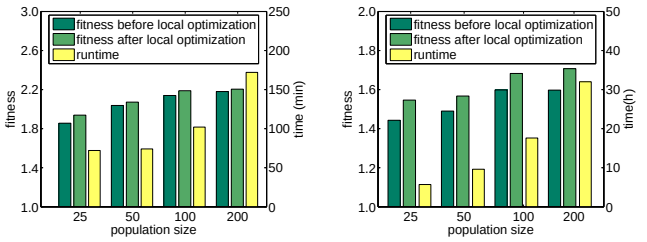
In order to investigate the effect of the local optimization, we take a look at the influence of the population size. The larger the population size, the more new individuals are created per generation resulting in a larger number of good individuals. This means that a large population size has the potential to get to the optimal solution but requires more computation time. In order to find a good population size, we need to look at the fitness of the best individual for a variety of population sizes and compare it to the runtime.

To see the influence of the population size as well as the local optimization, a genetic algorithm run with 500 generations is performed with an additional local optimization of 2500 generations. We investigate the influence on two different scenarios, with different average numbers of mesh points per gateway, and increase the population size from 25 to 200. The results are shown in Figure 18.

The fitness values are averaged values of the best individual over ten runs of the genetic algorithm. The runtime shows the minimal total runtime. In Figure 18(a) the local optimization only slightly increases the fitness of the best individual. However, in a scenario with a larger number of mesh points, the local optimization increases the fitness between 5 % and 7 % depending on the population size, cf. Figure 18(b). The reason is that such a scenario offers more possiblities to assign the routes and channels, which are evaluated in the local optimization process.

Taking a look at the population size, we want to point out that the performance increase is only visible up to a population size of 100. When increasing the population size to 200, a run takes twice as long as a run with a population size of 100, while the fitness increases only by 1.5 % at most. Thus, a population size of 100 is a good compromise between the runtime of the genetic algorithm and the fitness of the resulting individuals.

Summarizing, we want to point out that a local optimization of the best individuals is a good means to get to better solutions without significantly prolonging the runtime of the genetic algorithm. A similar result might be achieved after an additional 500 or 1000 generations but this would take much more time. Also the performance increase by enhancing the population size is negligible and almost doubles the runtime.



(a) Scenario with an average of 18.6 MPs per mesh gateway.

(b) Scenario with an average of 23.6 MPs per mesh gateway.

Figure 18. Relationship between population size, fitness, and runtime.

VII. Conclusion

In this paper, we investigated the usability of genetic algorithms for optimizing wireless mesh networks. Thereby, we showed that the performance of the genetic algorithm depends on the applied fitness function. The fitness function is used to evaluate the resulting network solution. We investigated eight different fitness functions optimizing for example the minimum, mean, and maximum throughput. The results show that the fitness function should be chosen with care because some functions lead to an unfair share of resources. Using a fitness value built on weighted throughputs of all network flows results in the best solutions. In addition to choosing a good fitness function, we illustrated that it is also important to choose the elite set size according to the population size. A small population with a large elite set size often results in a local optimum. The elite set size also has an impact on the required number of generations to get to a good solution. We showed that with an elite set size of one third of the population size, a near-optimal solution is achieved after 400 generations.

Besides the fitness function and the size of the elite set, the genetic operators crossover and mutation have to be carefully applied. We adapted the operators to the requirements of wireless mesh networks and introduced two new crossover variants called Cell and the Subtree Crossover. The evaluation of the influence of these operators revealed that the WMN-specific Cell and Subtree Crossover lead to better solutions compared to the well-known 2-Point Crossover. However, they have to be applied according to the network topology. The Subtree Crossover shows the best performance in scenarios with a large number of mesh points per gateway whereas the Cell Crossover leads to the best solutions in scenarios with a small number of mesh points per gateway.

During the progress of the genetic algorithm, the contribution of the crossover operator to find the optimal solution decreases. After several generation steps, almost no better solutions are achieved by applying the crossover operator.

Here, only mutation leads to a better fitness of the solution. We have shown that a reasonable network optimization is only possible by using mutation. The influence of the mutation operator in combination with all crossover types was tested and it was proven that in all cases it strongly fosters the evolution. Even in late generation steps, the fitness of the resulting solution improved.

In order to benefit from the crossover operator to get out of local optima at the beginning of the evolution process and to still get to better solutions at the end of the genetic optimization, we introduced the concept of an elite set increase and a local optimization. With every generation of the genetic algorithm, the elite set is increased, which decreases the number of crossover and mutation operations. In order to still mutate the individuals, the mutation operator is applied to the elite set and if a better solution is found, it is taken to the next generation. The local optimization is done after the normal generations procedure finishes. Thereby, several mutations of the five best individuals are performed and the resulting individuals are only kept in the new generation of the local optimization if the fitness value is higher compared to the fitness value before the mutation. Using these concepts, the performance of the WMN can be significantly increased with a minimal computational overhead.

Thus, we showed that genetic algorithms are well-suited for the optimization of wireless mesh networks. While other optimization techniques like linear programming fail to optimize large WMNs, genetic algorithms solve the complex structure of WMNs in relatively small computation time. However, the parameters of the genetic algorithm have to be carefully chosen and adapted to the applied topology.

References

[1] R. Pries, D. Staehle, M. Stoykova, B. Staehle, and P. Tran-Gia, "Wireless Mesh Network Planning and Optimization through Genetic Algorithms," in *The Second International Conference on Advances in Mesh Networks (MESH)*, Athens/Glyfada, Greece, June 2009.

[2] ——, "A Genetic Approach for Wireless Mesh Network Planning and Optimization," in *First International Workshop on Planning and Optimization of Wireless Communication Networks (PlanNet2009) in conjuction with the 5th International Wireless Communications and Mobile Computing Conference (IWCMC)*, Leipzig, Germany, 6 2009.

[3] D. Staehle, B. Staehle, and R. Pries, "Max-Min Fair Throughput in Multi-Gateway Multi-Rate Mesh Networks," in *IEEE VTC Spring 10*, Taipei, Taiwan, May 2010.

[4] D. P. Bertsekas and R. G. Gallager, *Data networks*. Prentice-Hall, 1987.

[5] K. Tutschku, "Demand-Based Radio Network Planning of Cellular Communication Systems," in *IEEE Infocom 1998*, San Francisco, CA, USA, March 1998.

[6] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Comput. Oper. Res.*, vol. 13, no. 5, pp. 533–549, 1986.

[7] S. Sen and B. Raman, "Long distance wireless mesh network planning: problem formulation and solution," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, New York, NY, USA, 2007, pp. 893–902.

[8] B. He, B. Xie, and D. P. Agrawal, "Optimizing deployment of Internet gateway in Wireless Mesh Networks," *Computer Communications*, vol. 31, no. 7, pp. 1259–1275, 2008.

[9] E. Amaldi, A. Capone, M. Cesana, I. Filippini, and F. Malucelli, "Optimization models and methods for planning Wireless Mesh Networks," *Computer Networks*, vol. 52, no. 11, pp. 2159–2171, August 2008.

[10] A. So and B. Liang, "Minimum Cost Configuration of Relay and Channel Infrastructure in Heterogeneous Wireless Mesh Networks," in *Networking*, Atlanta, GA, USA, May 2007, pp. 275–286.

[11] A. Raniwala, K. Gopalan, and T. Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 8, no. 2, pp. 50–65, April 2004.

[12] A. Raniwala and T. Chiueh, "Architecture and Algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *IEEE Infocom 2005*, Miami, FL, USA, March 2005, pp. 2223–2234.

[13] Y.-Y. Chen, S.-C. Liu, and C. Chen, "Channel Assignment and Routing for Multi-Channel Wireless Mesh Networks Using Simulated Annealing," in *IEEE Globecom 2006*, San Francisco, CA, USA, November/December 2006.

[14] K. N. Ramachandran, E. M. Belding, K. C. Almeroth, and M. M. Buddhikot, "Interference-Aware Channel Assignment in Multi-Radio Wireless Mesh Networks," in *IEEE Infocom 2006*, Barcelona, Spain, April 2006.

[15] A. P. Subramanian, H. Gupta, S. R. Das, and J. Cao, "Minimum Interference Channel Assignment in Multiradio Wireless Mesh Networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 12, pp. 1459–1473, December 2008.

[16] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *MobiCom '05: Proceedings of the 11th annual international conference on Mobile computing and networking*, Cologne, Germany, 2005, pp. 58–72.

[17] A. H. Mohsenian Rad and V. W. S. Wong, "Joint Channel Allocation, Interface Assignment and MAC Design for Multi-Channel Wireless Mesh Networks," in *IEEE Infocom 2007*, Anchorage, AK, USA, May 2007, pp. 1469–1477.

[18] ——, "Congestion-aware channel assignment for multi-channel wireless mesh networks," *Computer Networks*, vol. 53, no. 14, pp. 2502–2516, September 2009.

[19] P. Calégari, F. Guidec, P. Kuonen, and D. Wagner, "Genetic Approach to Radio Network Optimization for Mobile Systems," in *IEEE VTC Spring 97*, Phoenix, AZ, USA, May 1997.

[20] S. Ghosh, P. Ghosh, K. Basu, and S. K. Das, "GaMa: An Evolutionary Algorithmic Approach for the Design of Mesh-Based Radio Access Networks," in *LCN '05: Proceedings of the The IEEE Conference on Local Computer Networks 30th Anniversary*, Washington, DC, USA, November 2005, pp. 374–381.

[21] L. Badia, A. Botta, and L. Lenzini, "A genetic approach to joint routing and link scheduling for wireless mesh networks," *Elsevier Ad Hoc Networks Journal*, vol. Special issue on Bio-Inspired Computing, p. 11, April 2008.

[22] T. Vanhatupa, M. Hännikäinen, and T. D. Hämäläinen, "Performance Model for IEEE 802.11s Wireless Mesh Network Deployment Design," *Journal of Parallel and Distributed Computing*, vol. 68, no. 3, pp. 291–305, March 2008.

[23] J.-H. Lee, B.-J. Han, H.-J. Lim, Y.-D. Kim, N. Saxena, and T.-M. Chung, "Optimizing Access Point Allocation Using Genetic Algorithmic Approach for Smart Home Environments," *The Computer Journal*, 2008.

[24] T. Vanhatupa, M. Hännikäinen, and T. D. Hämäläinen, "Genetic Algorithm to Optimize Node Placement and Configuration for WLAN Planning," in *4th International Symposium on Wireless Communication Systems, 2007. ISWCS 2007*, Trondheim, Norway, October 2007, pp. 612–616.

[25] E. Damosso and L. M. Correia, *Digital Mobile Radio Towards Future Generation Systems, COST 231 Final Report*. European Commission, 1999.

[26] J. Jun and M. L. Sichitiu, "The Nominal Capacity of Wireless Mesh Networks," *IEEE Communications Magazine*, vol. 10, no. 5, pp. 8–14, October 2003.

[27] B. Aoun and R. Boutaba, "Max-Min Fair Capacity of Wireless Mesh Networks," in *IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, Vancouver, BC, Canada, October 2006, pp. 21–30.

[28] J. H. Holland, *Adaptation in natural and artificial systems*. Cambridge, MA, USA: University of Michigan Press, 1975.

A multiple channel selection and coordination MAC Scheme

Mthulisi Velempini
Department of Electrical Engineering
University of Cape Town
Cape Town, South Africa
mvelempini@crg.ee.uct.ac.za

Mqhele. E. Dlodlo
Department of Electrical Engineering
University of Cape Town
Cape Town, South Africa
mqhele.dlodlo@uct.ac.za

***Abstract* - The realization that single channel MAC protocols do not offer adequate end-to-end throughput has prompted researchers to explore more scalable approaches such as multi-channel MAC protocols. Multi-channel MAC protocols implementing a dedicated control channel offer promising solutions. However, it has been suggested that the use of a single control channel may lead to saturation problems. The saturation problem needs to be investigated. The paper proposes a cyclic scheduling algorithm, which schedules data transmission in phases. The scheme reduces the signalling overhead of the control channel and improves its capacity by reducing the effects of the channel switching delay and the idleness of the control channel. The scheme is connection oriented and implements the services of a network support systems, which provides the network with the required intelligence for data channels reservation. The scheme takes advantage of the integration of mesh routers and mesh clients in an overlaid wireless mesh networks (WMN). Mesh routers are also deployed in the ad hoc network of mesh clients to act as a network support backbone. The mesh routers forming the network support backbone are assumed to be within the communication range. We first analyze the control bottleneck problem of the proposed scheme as a channel selection and coordination problem requiring an effective channel scheduling technique. The scheduling techniques should be designed to minimize the effects of channel switching penalty on the control channel. The techniques should also increase the scheduling capacity of control channel. A single dedicated channel with at least two data channels and one transceiver system was considered in the analyses. The capacity of a single control channel is investigated as the number of data channels is increased from two to fourteen. Channel saturation is observed on data channels. Analytical results show that a single dedicated control channel causes no bottlenecks. Its capacity is affected by the saturation of data channels. The proposed scheme was also evaluated through NS 2 simulations. The numerical results show that the scheme is effective in reducing the signalling overhead.**

Keywords-Channel bottleneck, Channel coordination, Channel saturation, Channel selection, Connectivity, Multiple Channel MAC

I. INTRODUCTION

The paper analyzes multi-channel MAC protocols, which implement a dedicated control channel. The motivation is to establish the effects of channel saturation on the performance of the network and the causes of the saturation problem. The choice of multichannel schemes, which employ a dedicated control channel as a single signalling channel was motivated by the ability of these schemes in ensuring total network connectivity. The schemes under this category do not segment a network into logical segments.

A scheme, which implements one transceiver, a dedicated control channel, and at least two data channels is proposed. The proposed scheme facilitates network connectivity and is designed to use network resources efficiently. The saturation levels of channels are first investigated including the capacity of the control channel. The study seeks to establish through analytical means whether a dedicated control channel is a bottleneck in multi-channel systems. It also investigates how channel switching penalty can be employed effectively in improving the capacity of a control channel. The proposed scheme also attempts to reduce the idle periods of the control channel with a view of increasing its scheduling capacity.

The proposed a multi-channel scheme schedules flows on data channels in cycles. The scheme incorporates channel switching delay. Thereafter, through analytical means, the paper studies the effect of the saturation problem on both control and data channels. Key to the analysis is the capacity of the control channel as the number of data channels is increased steadily from two to fourteen. The paper tries to establish the soundness of this approach through analytical and simulations.

The rest of the paper is organized as follows: the need for a new multi-channel MAC scheme is discussed in Section II and related works are highlighted in Section III. The model is presented in Section IV and Section V presents the analytical results. The effectiveness of the proposed scheme is evaluated in Section VI. The simulation model and the numerical results are presented and discussed in Section VII. Section VIII then concludes the paper.

II. MOTIVATION

Multi-channel MAC schemes have a potential of increasing the capacity of current wireless access technologies. A number of multi-channel MAC protocols show that this increase in capacity

is possible. However, channel selection and coordination needs to be addressed if the full potential of multi-channel schemes is to be realized. The multi-channel MAC protocols, which implement a dedicated control channel architecture, are promising as they facilitate total network connectivity.

Furthermore, a single transceiver system employing two data channels and one dedicated control channel has been investigated elsewhere. The saturation of the control channel did not impact severely on the performance. It was established that the control channel can support a reasonable number of data channels. However, channel switching penalty has not been adequately considered in channel coordination and in multi-channel selection. Its effect has not been considered in the context of data channels degradation and how it increases the scheduling capacity of a dedicated control channel. When data transmission is scheduled in cycles, it is possible to minimize the effect of channel switching delay and to improve the capacity of the control channel.

III. RELATED WORK

In this Section, multi-channel MAC protocols, which employ either a temporary or a dedicated common control channel, are reviewed. The evaluation focuses on the efficiency of MAC protocols in reducing the signalling overhead of the control channel and how they improve the scheduling capacity of the control channel.

In [1] a temporary signalling channel called a default channel is implemented. Nodes reserve data channels during the ATIM window through the default channel. The reservation of the data channels is done through a data structure called the Preferable Channel List (PCL). Nodes exchange the ATIM/ATIM-RES/ATIM-ACK packets during the ATIM window and then the RTS/CTS packets during the data window to reserve one of the data channels. The signalling duration is long and its overhead cost is too high. All the signalling packets have been increased in size. These challenges impact negatively on the efficiency of the control channel and its scheduling capacity is reduced. During the ATIM window, only the default channel is used while all the data channels lie idle. The bandwidth of the data channels is therefore, wasted during the ATIM window.

In [2] a temporary common signalling channel is implemented during the control window. One data channel is selected by a number of pairs thereby increasing the probability of data collisions. The sizes of all the control packets have been increased, which degrades the performance of the control channel and reduce its capacity to support many data channels. There is also an additional signalling packet called the reserve (RES) packet, which increases the signalling payload of the scheme. The protocol is similar to [1] and it suffers from similar constraints. The bandwidth of data channels is also wasted during the control window.

A multi-channel MAC scheme employing a dedicated control channel is proposed in [3]. The reservation of data channels is done through the control channel with an aid of the local channel tables. If a pair fails to reserve a data channel, they are given a second chance to reserve a free data channel. Assuming that all pairs succeed in their second attempts, a significant control channel overhead will be incurred. The nodes also defer for a DIFS and for the multi-channel switching durations, which erodes the capacity of the control channel. The reservation of data channels is done when the current transmissions have ended. Unfortunately, the control channel lies idle for fairly long durations, which can be reduced to improve its capacity. The bandwidth of the control channel is wasted when data frames are being transmitted on the data channels. On the other hand, the bandwidth of data channels is wasted as they are queued at the control channel waiting for their next transmissions.

The feasibility and functionality of the scheme is challenged in newly deployed networks and by the joining terminals. Joining nodes have to first set all the channels unavailable until they have updated their local channel tables. With inadequate network information, nodes will defer indefinitely their next transmissions. Nodes returning to the control channel are also equally affected as they have to set all the channels unavailable except the ones they have just visited. Nodes have to defer their transmission until they have acquired adequate information about the status of the network. This is not possible with a newly deployed network where all nodes have no knowledge of the network. In such a situation nodes will defer indefinitely their transmission resulting in a blocked network state. There will be no transmission that will take place, until nodes acquire adequate network information.

CTS packets reserves a data channel, unfortunately they fail to calm hidden nodes, which are in the sender neighbourhood. The probability of collisions and destroyed packets will increase forcing the scheme to retransmit a substantial amount of packets, which will further degrade the capacity of the control channel. Furthermore, nodes have limited processing power and storage capability and cannot process efficiently the local node tables and store the processed information.

The Dynamic Channel Assignment (DCA) [4] uses two transceivers, the control channel and the data channels transceivers. The use of two transceivers is expensive in terms of hardware costs. It also comes with increased design complexity. There is also a signal linkage problem where signals from the adjacent transceivers interfere with each other.

Each node keeps two data structures called the channel usage list (CUL) and the free channel list (FUL). Their use is similar to the three schemes above. The introduction of the RES packet also causes further delays. The control channel and the control channel transceiver will be idle when all data channel transceivers are busy on the data channels. The capacity of the control channel is therefore degraded. The bandwidth of the data channels is equally wasted when they lie idle waiting for the control channel to schedule data transmissions.

The scheme proposed in [5] also exploits the common control channel approach. Nodes listen on the common channel to

synchronize their hopping sequences. Long signalling durations are experienced by nodes when they negotiate and share hoping sequences. Furthermore, nodes send RTR packets without sensing channels, which may results in RTR collisions. Frequent hopping involves a sizeable amount of channel switching delays, which may significantly degrade the performance of the proposed scheme.

In [6] a dedicated signalling channel is employed. The dedicated channel is also used as a data channel after the contention period. The protocol is divided into contention reservation interval (CRI) and the contention free interval (CFI). Nodes contend for network resources and data channels during the CRI and they all defer their transmissions until the beginning of the CFI. The deferment wastes resources and degrades the capacity of the signalling channel. The protocol requires global synchronization a challenge in wireless networks.

In [7] a system employing busy signals is proposed. A single channel is divided into a control and a data channel. Busy signals are sent on the control channel while data frames are being transmitted on the data channels. The scheme assumes that a node can send and listen at the same time. It also assumes that a node can transmit on two channels simultaneously. The major shortcoming of this scheme is the wastage of bandwidth of the control channel – the busy signal channel.

Nodes in [8] [9] randomly select independent home channels to listen on when they are idle. This approach segments a network and does not facilitate network connectivity for effective communication. The idea of data structures is also employed.

To address the problem of synchronization a guard time is implemented unfortunately the guard time degrades the performance of the protocol. On the other hand every packet sent should include a 32 bit current time and seed. The time stamp increases the payload of the protocol.

A new node, which has just joined a network, has to first wait for ten seconds before it establishes and follows its own hoping sequence. This waiting time increases the signaling overhead cost. Courtesy HELLO packets are also sent to newly discovered neighbours. The HELLO messages do degrade the performance of the protocol and may fail to reach all the nodes on different channels.

A scheme implementing a separate control channel and N traffic channels is proposed in [10] and in [11]. A CTS selects the clearest channel, unfortunately a CTS based data channel selection scheme fails to calm hidden nodes at the sender's neighbourhood as a result there will be numerous retransmissions, which will degrade the control channel. The size of the control packets has been increased. The enlarged packets will further degrade the capacity of the control channel.

The protocol assumes that nodes can sense all the channels and receive on all the channels simultaneously. Nodes can sense only one channel at a time and thereafter can switch to the next channel incurring very high overhead costs in both channel switching and channel sensing.

A multi-channel MAC protocol called dynamic channel assignment with power control (DCA-PC) is proposed [12]. However, it uses the same channel access and reservation mechanisms discussed in [4]. The proposed protocol therefore suffers from the same challenges.

The paper in [13] proposed a Distributed Queue Dual Channel (DQDC) scheme, which seeks to increase the utilization of the data channels and to increase the achieved throughput. The scheme implements the notion of a control channel however; the DQDC introduces a four way packet handshake negotiation scheme. The following packets are exchanged before a data channel is reserved: Mesh Transmission Opportunity Request (MTXOP REQ), Mesh Transmission Opportunity Response (MTXOP RSP), Mesh Transmission Opportunity Acknowledgment (MTXOP ACK) and Agreement Indicator (AID). A three way handshake is also possible when the receiver accepts the MTXOP REQ.

When the channel has been reserved, neighbouring nodes are notified through the AID, which is sent by the node receiving the MTXOP ACK. The AID is a broadcast message, which fails to reach nodes, which are currently transmitting on data channels and those, which are hidden from the AID broadcasting node. Returning nodes assume that the data channels are busy until they receive the reservation messages or after the expiry of a set threshold. This delays the next transmissions of the returning nodes and forces the control channel to be idle for longer time frames. Furthermore the signaling overhead is significant and it degrades the performance of the control channel. Nodes have to update their DQ each time they receive an AID or MTXOP ACK message. The DQ processing requires a node with unlimited processing power and storage capacity.

The signalling delay should be reduced further. As noted above, most multi-channel protocols suffer from high control channel overhead and wastage of bandwidth of both the control and data channels. These should be reduced to improve the scheduling power of the control channel and to increase its capacity. The idle durations of the control channel should also be reduced for improved control channel performance. The processing and storage constraints of mobile nodes should also be considered. A number of other multi-channel MAC protocols are discussed in [14] to [22]. The scheme proposed in [26] seeks to address these and other challenges of multi-channel MAC protocols, which employ the services of a control channel. A common control channel is presented as a driver that can increase the capacity of multi-channel MAC protocols.

IV. THE MODEL

A Cyclic Scheduling algorithm (CSA), which is equipped with network support infrastructure, is proposed. The network infrastructure is designed to provide terminals with adequate network information, which is required for the reservation of data channels through the control channel. The main objective is to

reduce the signalling overhead of the control channel and increase its scheduling capacity. The network support infrastructure is made up of a network of mesh routers overlaid in the ad hoc network of mesh clients. The mesh routers forming a network support infrastructure are called the NST nodes. Every NST node is within the communication range of the next NST node. The mesh routers are powerful and intelligent. Their communication ranges are wider and do cover as many mesh clients as possible.

The proposed scheme requires the services of a single transceiver and it divides the channels into one dedicated control channel and *n* data channels. The number of data channels should be at least two. The control packets, the Request To Send (RTS) and the Clear To Send (CTS) will be transmitted on the control channels. The data frames and the Acknowledgement (ACK) packets will the transmitted on the data channels.

Data transmission on data channels is scheduled in cycles when data channels are about to be free to improve the scheduling capacity of the control channel. The sending nodes will start contending for the control channel when the first channel is about to be free. A communicating pair reserves the next available data channel through the control channel. However the reservation and the handshake of control packets will be done before the data channel to be reserved becomes idle. Data channel reservation starts when the remaining transmission time of the current transmission on the data channel is exactly equal to the duration of the control channel handshake. The data channel will only become free as the new pair is switching onto the reserved data channel. The switching will be completed when the data channel is free. To ensure that the timing is perfect, an inter-cycle duration, which is a function of the number of data channels, is designed.

The switching penalty is considered in the design of our scheme. The nodes switch to the reserved data channel after the exchange of RTS/CTS. They do not defer for the summation of DIFS and the channel switching duration. When the acknowledgement has been sent, the nodes will switch back onto the control channel. Given this double effect of switching delay on data channels, a control channel is unlikely to result in a bottleneck. The scheduling capacity of the control channel is likely to improve as nodes take longer to return to it. The control channel will have more capacity to service the sender-receiver pairs, which are currently in the queue. The data channels may instead create a bottleneck in the system. The scheme takes advantage of switching penalty to improve scheduling capacity of a control channel. Carrier sensing and channel contention are limited to the control channel to improve the capacity of the protocol. Data channels are reserved thorough the network support infrastructure.

The inter-cycle duration will separate data transmission cycles and indicate when control channel reservation should start. Given the inter-cycle duration, the next pair will initiate communication on the control channel, exchange control packages and then switch onto the data channel while it is still busy sending data. The inter-cycle duration is the duration that separate two consecutive cycles. Its value varies with the number of data channels. It is an inter-cycle hold off duration whose value is determined by (3). The busy data channel will be expected to be free just before the next pair arrives on the data channel. The implementation of the network support infrastructure ensures that there are no retransmissions on the data channels. Retransmissions are possible on the control channel which implements the IEEE 802.11 channel reservation mechanisms.

In a given cycle the first pair to reserve the control channel will automatically reserve the first data channel; the second pair will reserve the next data channel while the last pair in a cycle by default will reserve the last data channel. There will be no contention required for the data channels however; the nodes have to contend for the access of the control channel. The cyclical scheduling algorithm will tie access to data channels to phases in a cycle to reduce the reservation duration and signalling overheard. Access to data channels will be linked to the reservation of the control channel. The access to data channels will be in phases within a cycle.

The pair, which wins the control channel in phase one will access the first data channel and in the N^{th} phase the N^{th} data channel will be reserved by the N^{th} pair. The cyclic scheduling algorithm will be memory based (network support) keeping track of all the activities of the data channels, cycles and phases within cycles. These details will be stored on the network support infrastructure of multitasking mesh routers in a hybrid Mesh network.

To explain further the concept behind the cyclic scheduling algorithm Figure 1 is employed. The diagram though not in scale is based on the length of control packets, channel switching duration of 224μs and the length of data packets as stipulated in the standard. The durations however, were changed to make them more manageable and easy to work with. For example short durations were scaled up while longer ones were scaled down. This manipulation of the durations only reduces the duration of the inter-cycle hold off time. However, the idea the paper conveys can still be appreciated.

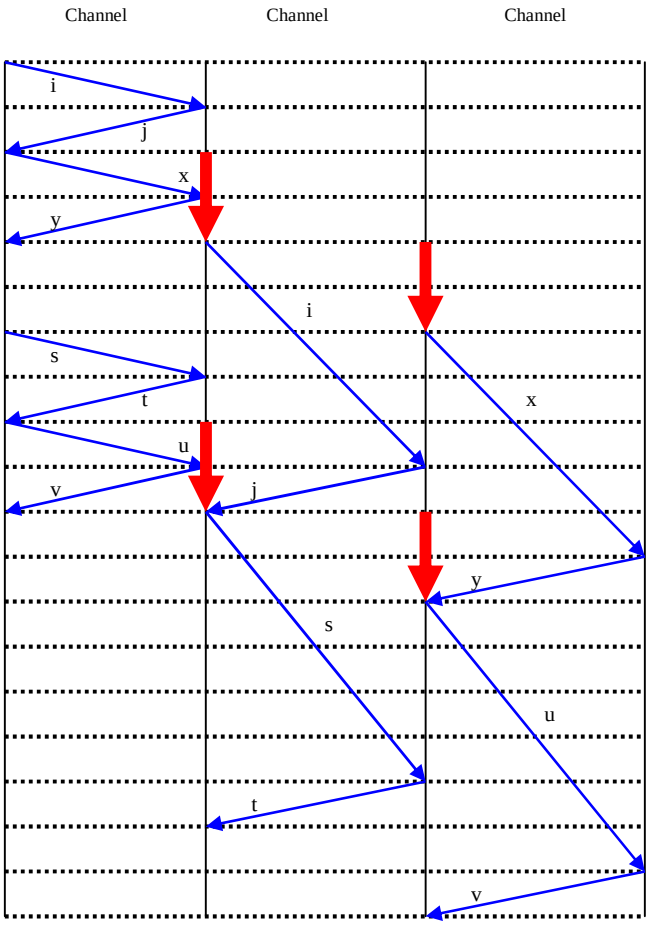


Figure 1. Packet switching analytical diagram.

Channel 0 denotes the control channel, while channels 1 and 2 are the two data channels. The red arrow depicts the channel switching delay. Lastly, the blue arrows represent data and control packets durations.

In Figure 1 the two communicating pairs (ij) and (xy) will automatically reserve data channels one and two respectively during the first cycle. After the first cycle terminals will back off for an inter-cycle duration to accommodate the switching delay and to properly mark the beginning of the next cycle. For a three channel system, the inter-cycle duration will be equal to the total transmission duration of a data packet plus two switching durations minus the summation of three control packets handshakes. The inter-cycle is computed using equation (3). During the inter-cycle duration, the control channel lies idle until the onset of the next cycle. The inter-cycle duration will reduce as more data channels are implemented, which improves the capacity of the control channel.

In the next cycle, the next pairs (st) and (uv) will reserve the two data channels while the previous two pairs are still communicating on the same two data channels. As the next pairs are switching to data channels, the first pairs will complete their transmission. The first pairs will then switch back to the control channel to wait for the next cycle if they still have data to send. The first pair in each cycle must reserve the first data channel. This rule ensures a collision free data exchange on data channels.

The design of the hold off duration and the channel coordination scheme is fundamental to the success of this scheme. In Figure 1 only channel switching delay to a data channel is depicted. The reverse channel switching delay is not shown in the diagram.

In Figure 2, RTS, CTS, DATA and ACK packets are presented in a form of a block diagram. It can be seen that this approach waste bandwidth in the data channels during the first cycle. This is caused by the switching delay and the use of multiple channels. We call this bandwidth wastage; the multi-channel scheduling cost (MSC). This degradation is not avoidable, but can be limited to the first cycle like in our case. The MSC has been eliminated in the subsequent cycles by our scheme. In the existing protocols, this cost is periodic and repetitive. The switching cost and the multi-channel scheduling cost are not evident in the subsequent phases as our cyclic scheduling protocol takes advantage of switching penalty coupled with the inter-cycle duration to schedule concurrent transmissions in a proactive manner. The multi-channel scheduling cost is first identified in [26] and is common with protocols implementing a common control channel.

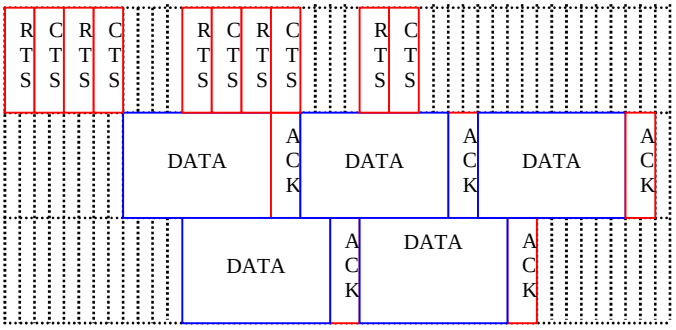


Figure 2. Packet scheduling block diagram.

To calculate the amount of wasted bandwidth caused by the multi-channel scheduling cost in our scheme, the following *for loop* can be implemented:

```
for (i =0; i <= n; i++)
    x += i;
Wasted bandwidth = x * handshake duration;
```

This cost is limited to the first cycle in our scheme and its value depends on the number of data channels *N*. As the number of data channels is increased, the multi-channel scheduling cost will increase as well. In the earlier protocols discussed in the

related section, the effect of the multi-channel scheduling cost is repetitive.

To evaluate the proposed scheme an analytical model was designed to model channel occupancy and investigate the capacities of control and data channels. The model examines the saturation of data channels and the capacity of the control channel to schedule successful data transmission onto the available data channels. To test the efficiency and the scalability of the control channel, the number of data channels was increased from two to fourteen.

To investigate the capacity of the channels the following equations based on our proposed idea were employed. The equations capture the essence of our proposed scheme; they emulate the allotment of bandwidth. The variables used in the equations are explained in Table 1. The equations are based on the Institute of Electrical and Electronics Engineers 802.11 Carrier Sensing Multiple Access with Collision Avoidance - IEEE 802.11 CSMA/CA mechanism. They were adjusted to suit and meet the specifications of our protocol.

$$DC = B_{dc} - D_l + 2 * sw * DC_n - msc \quad (1)$$

$$CC = B_{cc} - hd * DC_n - Intcyc \quad (2)$$

$$Intcyc = D_l + 2 * sw - DC_n * hd - sw \quad (3)$$

TABLE I. LIST OF VARIABLE USED IN THE EQUATIONS.

Variable		Variable Meaning
DC		Data Channel
B_{dc}		Data Channel Bandwidth
D_l		Data packet length
Sw		Channel switching delay
DC_n		Number of Data Channels
Msc		Multi-Channel scheduling cost
CC		Control Channel
B_{cc}		Control Channel Bandwidth
Hd		Control Channel handshake duration
Intcyc		Inter-cycle duration

Given the above equations, the capacities of both the control and the data channels were computed, allotment to nodes done and channel saturation investigated. All channels were considered to be having the same bandwidth of 1Mbs. Both data and basic rates were set to 1Mbs. The number of nodes was varied between 30 and 210 depending on the number of data channels. A system with fourteen data channels had the largest topology. In the analysis, however, an average of 30 nodes and two data channels was considered in each case. In some cases the total number of nodes was considered. The data frames were assumed to be 1000 bytes. Other parameters such as the control packet sizes were set to standard lengths specified in the IEEE 802.11 specification.

We now describe the functionality of the network support and its significance in reducing the signalling overhead. We also show how it facilitates communication in a newly deployed network. The network support is also designed to provide joining and returning nodes with information, which allows them to initiate their next transmission immediately instead of deferring them to a later stage. The network support takes advantage of the composition of the WMN and its different nodes, which have different capabilities. Of interest is a hybrid WMN, which has in addition to a backhaul of mesh routers and an Ad hoc network of mesh clients, it has a backbone of fully connected mesh routers within the ad hoc network of mesh clients.

Each node, which is part of the network support, maintains a data structure called a Network Status Table (NST). These nodes are referred to as the NST nodes. The NST nodes will store information about the availability of data channels, list of data channels, which are currently in use and when they will become available. The data channel will be said to be available when the remaining transmission time is equal to the amount of time required for the next pair to reserve it. This remaining time is determined by the inter-cycle duration, which was discussed earlier. The inter-cycle duration is stored in the NST as the duration of the data transmission duration of a given data channel. The network support nodes will also maintain a sequence of data channels to ensure that data transmission is scheduled in a round robin bases. The information maintained in the status tables is made available to any node, which probes the NST node.

When a node wishes to send data and does not have a complete understanding of the network status, it first probes the nearest NST node. Upon receiving this information it will be able to exchange the control packets (RTS/CTS) on the control channel and reserve a data channel, which will be available next. The reservation is done before it becomes idle after the inter-cycle duration.

The network support system is of paramount importance for a network, which has just been deployed. In such a scenario nodes would not have a complete picture of the network status. Instead of waiting indefinitely in attempt to gather information from overheard control packets, nodes will simple probe the nearest NST node. All the nodes current on the control channel and within the coverage of the NST node will receive the probe response. The responses will be used by a number of nodes to update their own local tables. The local tables are expected to be small and limited in size. This will help nodes with limited processing power to store and update their local tables effectively.

The NST information will also be helpful to joining and returning nodes. A node, which has just been registered, is considered to be a joining node. A joining node could be a node, which has moved from one NST node zone into the zone of the

next NST node. Nodes can also join a network from other adjacent networks when they have been handed over to the next network. On the other hand, a node is said to be returning if it was busy transmitting or receiving on a data channel. Upon the completion of the transmission, it switches back onto the control channel. This node would have missed the details of data channels, which were reserved during its visit to one of the data channels. When the returning node wants to initiate communication immediate upon its return, they must first send a probe to the nearest NST node otherwise it will not communicate due to lack of sufficient network information. The network support system therefore, prevents the deferment of transmission and reduces bandwidth wastages due to false blocking of nodes with insufficient information.

V. ANALYTICAL RESULTS

In this section, the channel saturation problem is investigated analytical. Both control and data channels are investigated. The number of data channels was increased from two to fourteen to investigate how increased capacity and congestion affects the capacity of the control channel. All the channels were assumed to have a bandwidth of 1Mbs. The channels were assumed to be orthogonal for testing purposes. The three equations discussed in the previous section were used in the analysis of the cyclic scheduling algorithm.

It was noted that data channels transmit long packets as compared to the control channel. Data channels are also degraded by channel switching delays. These two observations do affect negatively the capacity of the data channels. On the other hand short packets provide the control channel with more capacity to drive many data channels.

In Figure 3 a system with one control channel and two data channels was considered. The general topology had a total of thirty nodes. It was noted that the capacity of the two data channels caused a bottleneck in the system. Their combined capacity could support up to fourteen nodes. This translates to seven nodes per a data channel. On the other hand the control channel had enough capacity for the thirty nodes.

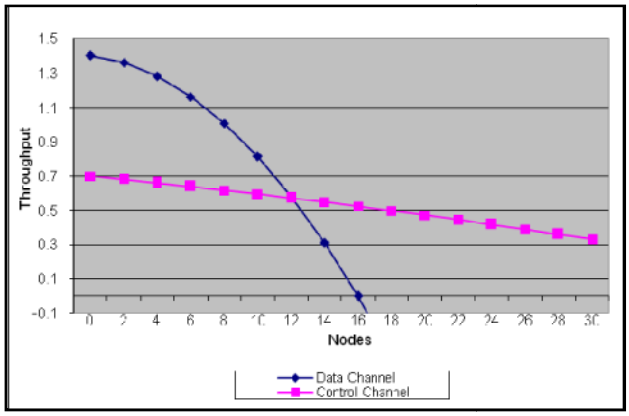


Figure 3. Performance of the control channel when two data channels are considered

The inter-cycle has the longest duration in a two data channels system. Its duration reduces with the every increase in data channels. The inter-cycle duration degrades the capacity of the control channel. Despite this, the capacity of the control channel was still adequate. It can be seen that the control channel was underutilized when there were very few data channels. The number of data channels should be steadily increased to improve the utilization of the control channel.

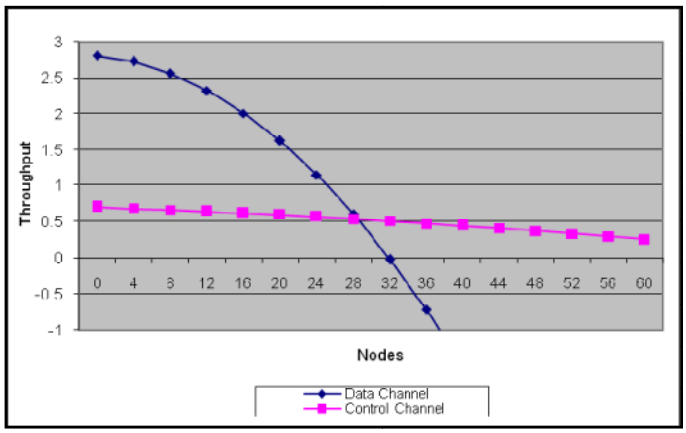


Figure 4. Saturation levels in a system with four data channels.

When the number of data channels was increased to four in Figure4, it was noted that the capacity of the control channel was still underutilized. On the other hand the number of nodes was increased to sixty. The data channels saturated first and their performance was unchanged. However, there was small change in the performance of the control channel. This proves that the number of data channels do affect the performance of the control channel though it this case in is insignificant.

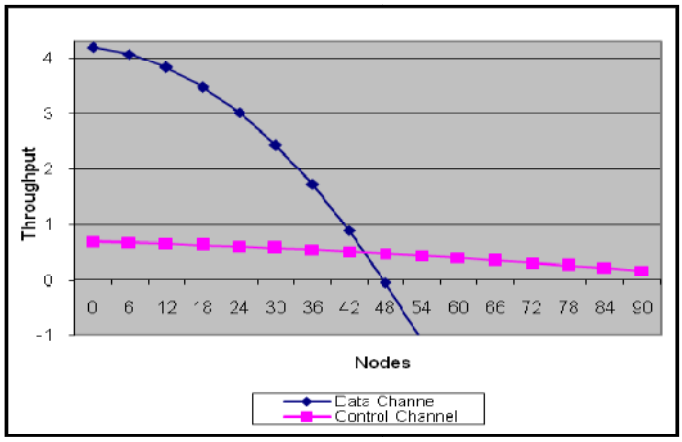


Figure 5. The saturation levels of channels in a network with six data channels.

The number of data channels was further increased to six while the nodes were increased to ninety. The increases were designed to evaluate the effects of the network size and the number of data channels on the performance and capacity of the

control channel. As can be seen in Figure5 the capacity of the control channel continues to degrade gracefully, while the performance of the data channels is unchanged. The performance of the data channels does not improve as the number of data channels. It is only affected by the saturation of the control channel is increased. However, it is important to note that the control channel at this stage is still has enough capacity for the six data channels and ninety nodes.

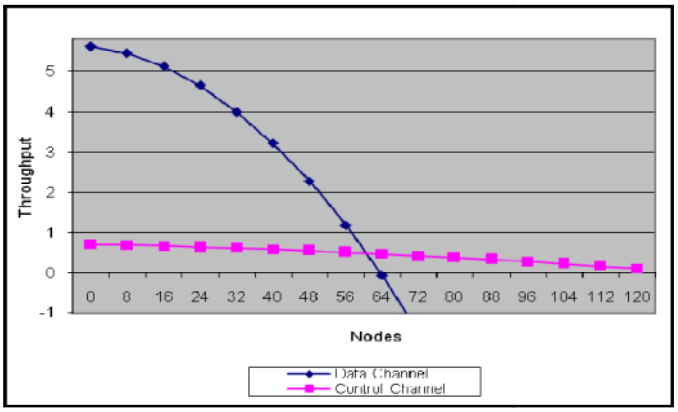


Figure 6. The saturation point of the control channel

In Figure6 the capacity of the control channel continues to degrade, thought it had enough capacity for the 120 nodes, its capacity was now limited. It was fast approaching zero. Figure 6 had eight data channels in total. The control channel still had enough capacity for the eight data channels.

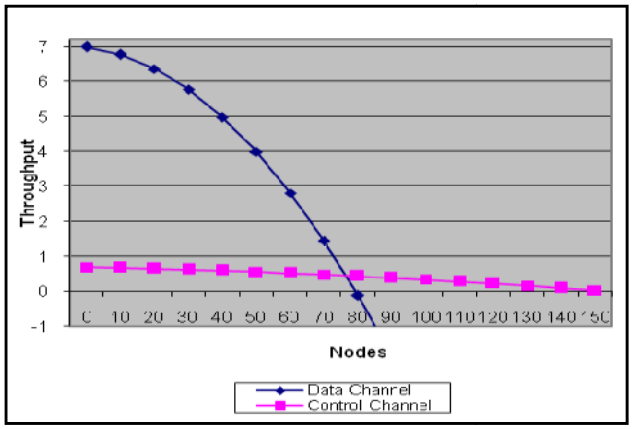


Figure 7. Performance of the control channel when ten data channels are considered.

In Figure 7 the number of data channels was increased steadily to ten. The control channel could drive all the data channels its capacity was not enough for the 150 nodes. The control channel did not have adequate capacity for all the available nodes. Interestingly when the data channels were fewer, the control channel was underutilized. It began degrading as the number of data channels was increased. To optimize the performance of the scheme, the number of the data channels should be increased to a level, which does not underutilize the capacity of the control channel. On the same token, the control channel should run at a level, which does not degrade the performance of the data channels. Although, the control channel began saturating, it did not cause a bottleneck in the system. The bottleneck was caused by the data channels whose capacity was just enough for the first seven nodes in each data channel. There was no degradation on the data channels, which was caused by the saturating control channel.

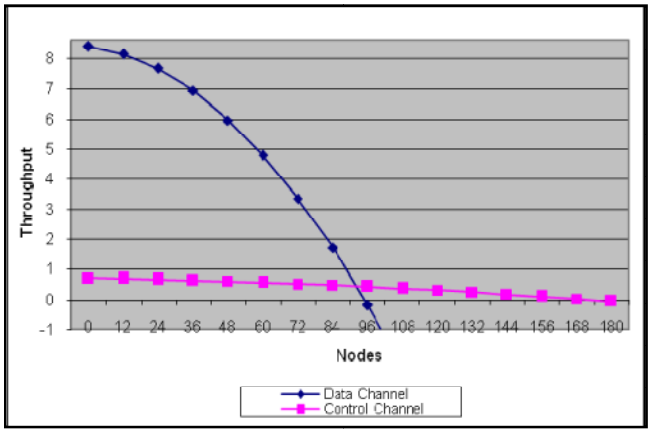


Figure 8. Performance of the control channel when twelve data channels are considered.

The saturation of the control channel becomes more apparent in a system with twelve data channels. In a general topology with 180 nodes, the capacity of the control channel is limited to only 168 nodes in Figure 8. The impact of the saturation of the control channel becomes severe as more data channels are added. The saturation of the control channel at this stage does not degrade the performance the system. The performance of the system would have long been degraded by data channels, which cause a system bottleneck after the 7th node on each data channel. To improve the system performance, higher data rates should be considered for the data channels.

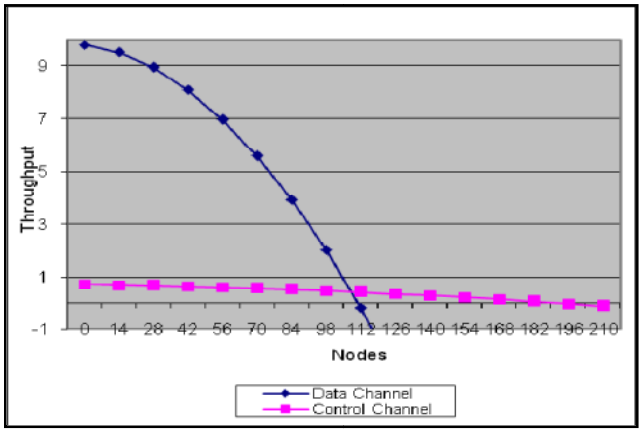


Figure 9. Performance of the control channel when fourteen data channels are considered.

In Figure 9 similar observations were made. In this case a total of fourteen data channels were considered. The general topology had 210 nodes. The combined capacity of data channels was only enough for ninety-eight nodes. The control channel could support 182 nodes instead of 210. The control channel began saturating after the 182nd node mark. This was the largest general topology, which was investigated. The network provided the shortest inter-cycle duration in our experimentation. The inter-cycle reduces with every increase in the data channels as more capacity is required by the control channel to schedule more transmission and drive more data channels. The control channel becomes busier servicing an increasing load of data channels. Furthermore, the multi-channel scheduling cost had its biggest impact on the network with fourteen data channels.

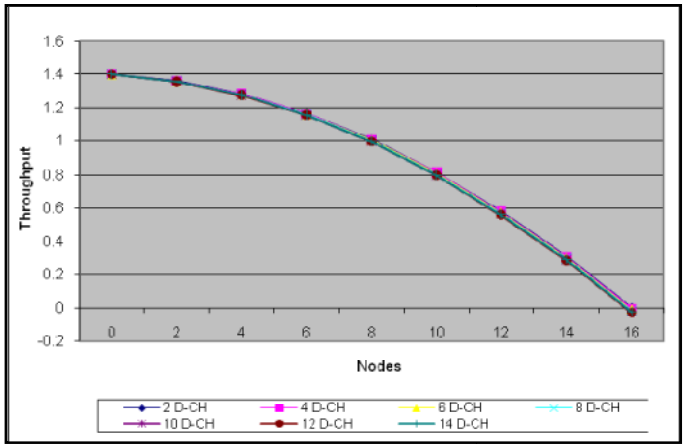


Figure 10. Analysis of the saturation levels of data channels.

A snap shot of the seven different data channels systems ranging from the two data channels to the fourteen data channels system shows a similar pattern in performances of the networks. The performances of these seven networks are depicted in Figure 10. All the data channels in each case saturated after the seventh node. Therefore the average performance of data channels does not increase as the number of the data channels is increased. However the overall increase in capacity can be observed as data channels are increased, though the performance remains the same. It can be concluded that an increase in data rates will also improve system capacity while the performance will remain unchanged. The performance is therefore not expected to improve though network capacity may show an increase in the end to end throughput.

The analysis of the control channel for the seven different cases reveals an interesting development in Figure11. The performance and the capacity of the control channel is affected by the number of data channels. Its performance reduces with each and every increase in data channels. Following this observation, we can conclude that the number of data channels do degrade the performance of a control channel. It should be noted that despite the degradation of the control channel, it still has enough capacity to drive as many as fourteen data channels. This statement is valid given the fact that data channels saturate after every seventh node. However, the fourteen data channels are not orthogonal; they were assumed to be overlapping for experimentation purposes. The idea was to investigate the capacity of the control channel as the network size increases and more data channels are added.

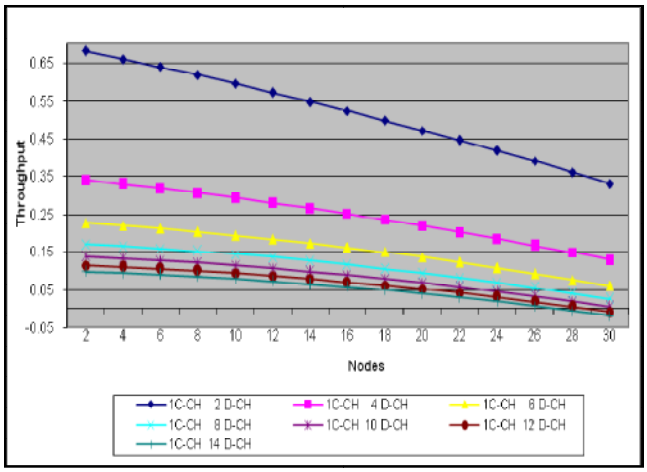


Figure 11. Analysis of the capacity of control channel as data channels are increased from two fourteen.

VI. ANALYZING THE CYCLIC SCHEDULING ALGORITHM

Multichannel MAC protocols implementing a single control channel can be modeled as a queuing network. The network has three distinct service points. The service points are depicted in Figure12. These are the nodes marked Ns1 to Nsn, the control channel identified as Cc1 and then the data channels marked as Dc1 to Dcn. In the queuing model it can be seen that the control channel can slow down the speed of the network as a single service station fed by multiple servers and in turn sending its output to multiple servers. The capacity of the control channel therefore needs to be improved substantially. The control channel server should have reasonable capacity to offer well balanced service so that it can provide an efficient and a very fast link to multiple servers at its input and output ends.

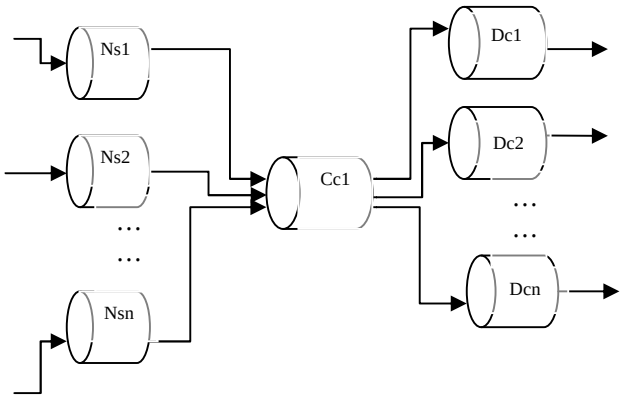


Figure 12. A Multichannel queuing network with a single control channel server

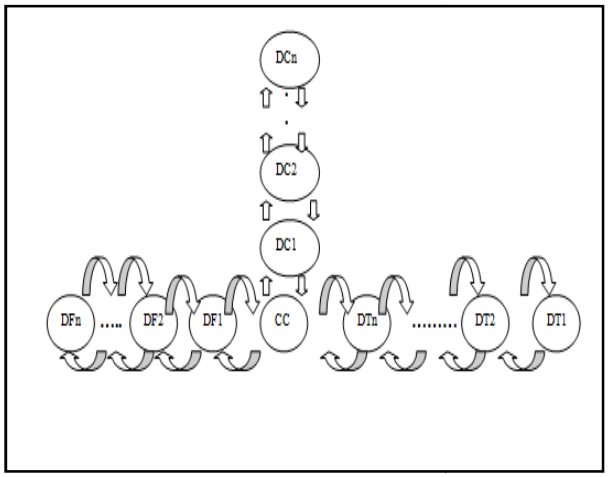


Figure 13. A multiple senerio in multichannel systems with a single signalling channel

Figure 13 shows how the implementation of the common control channel as a single signaling channel results in the formation of multiple queues. Data packets are first queued when the nodes wait for the control channel to be free and contend for it. In our analysis, we consider these packets as data flows queued at the control channel and are denoted as DF1 to DFn. When the data flow is de-queued from the control channel it is then queued in the data channel queue. Thereafter the data flows are served by the data channels. These two models clearly show the significance of the control channel in ensuring better performance of multichannel networks, which implement the notion of a signaling channel.

In the following three figures we evaluate four multichannel MAC protocols, which employ the idea of a control channel as a single signaling channel. We assume a Markovian packet arrival with an exponential inter arrival times. The arrival rate was assumed to vary between 0.1 and 2.9. The control channel service rate is based on the amount of time the control channel will service a pair, which wants to reserve one of the available data channels. The service rates were based on the payload of the control channel of the schemes primarily to show the effects of signaling on the capacity of the control channel. For the AMCP we considered the worst case for control channel utilization where all communicating pairs have to re-initiate data channel reservation after failed first attempts. This is according to the protocol, which allows nodes to attempt to reserve a data channel, which is available at both ends after the first attempt, which was unsuccessful. However, for both waiting and response times we assumed that all the initial attempts would be successful.

The following are of interest in our analysis of the schemes: system utilization, the average time in the queue and system. All the three were limited to the capacity of the control channels of the individual protocol, which were evaluated. The Little's theorem was used in calculating the values of the above parameters. In Figure 14 we evaluate the utilization of the control channel in the four protocols, the MMAC, LCM, AMCP and the CSA. The MMAC had the worst utilization factor followed by the LCM. The AMCP and the CSA offer the best performance. However, the CSA is slightly better. The low utilization factor shows that a given scheme has more capacity to handle increased volumes.

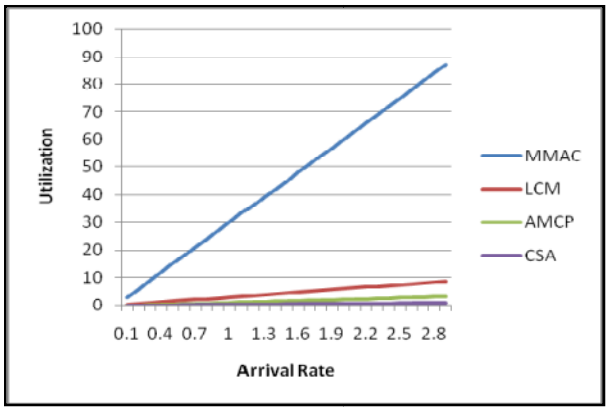


Figure 14. The utilization of the control channel.

The MMAC is not stable for the inter arrival rates that were considered in this analysis in both Figures. 15 and 16 because of its very low service rate. The results of MMAC are therefore not shown in these two figures.

In Figure 15 the CSA offered the best performance. The packets of the CSA were subjected to the smallest amount of delay in the queue. The turnaround of packets in the queue was the fastest as compared to the AMCP and the LCM. The AMCP was the second best. When the protocol offers the least delay in the queue it shows that the service rate of the control channel is good and does not cause significant degradation of the performance of the protocol. The LCM had the largest amount of delay and therefore the control channel is likely to degrade significantly the performance of the protocol. The results for the MMAC are not shown due to its instability within the inter-arrival range considered for this analysis.

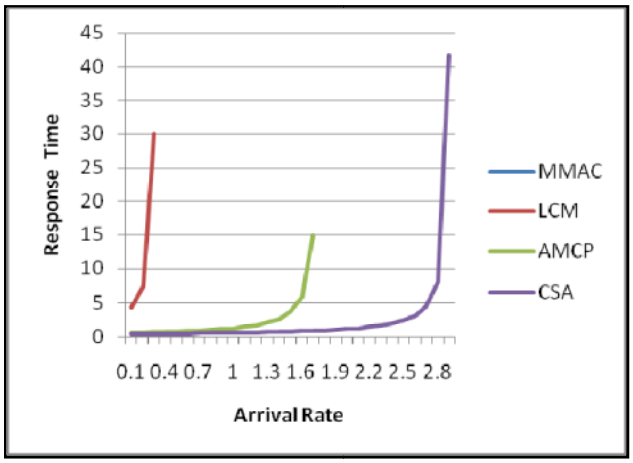


Figure 15. Analysis of the system response time in the queue.

In Figure 16 waiting time was considered. The waiting time is the summation of the response time and the processing time. In this case, it is the amount of delay a packet is likely to be subjected to before it is transmitted on the data channel, where both the queuing and the service times are considered. There was no significant difference between the queue and the system average times results. A similar trend was observed in the two graphs. However the delays were slightly higher in Figure. 16 due to the addition of the service times to the queue delay.

The CSA is therefore effective in reducing the signaling delay of the control channel. The reduction of the signaling delay means that the capacity of the control channel has been improved and its scheduling capacity increased.

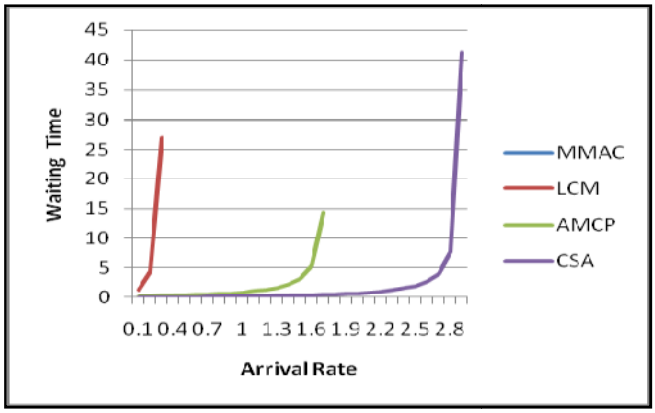


Figure 16. The analysis of the system waiting time in the system

VII. THE SIMULATION MODEL

In this section, we evaluate the performance of our Cyclic Scheduling Algorithm and compare it to the Asynchronous Multi-channel Coordination Protocol (AMCP). The channel switching delay has been included in the AMCP platform for better comparison with our approach. Total throughput achieved was employed as a metric for evaluation purposes. The analysis sought to find out, which of the two schemes achieves better throughput.

The AMCP was evaluated against after multi-channel schemes and was found to be superior in [3]. For this reason, the proposed scheme was only compared with the AMCP.

Default Network Simulator 2 (NS 2) parameters were used. We considered IEEE 802.11 MAC standard in our simulation. The channels were assumed to be orthogonal and of the same bandwidth. The bandwidth of each channel was assumed to be 2Mbs each.

A total of five channels were employed in the simulation with four data channels and one control channel. The number of the channels was fixed throughout of the simulation. However, different network sizes were considered. There were four different network sizes, which were considered. We assumed that all the networks had general topologies.

The switching delay was set to 224µs and two switching delays were considered. The first channel switching delay was incurred when terminals switched from the control channel to the reserved data channel to transmit data frames and ACK packets. The second one is when the nodes switch back onto the control channel after finishing their transmissions on the data channel.

The NO Ad Hoc (NOAH) routing agent was implemented in all the four networks. For each of the four network sizes, at least twenty simulation runs were considered with each simulation run, running for three hundred simulation time. Different network sizes were either expressed in terms of the number of terminals or the number of data flows. The number of terminals was always double the number of data flows. A given data flow links two distinct terminals, a sender and a receiver.

The data packets were assumed to be of type CBR and were all set to 1000 bytes. Both data and control packets were sent between a sender and a receiver. The network was assumed to be single hop network, hence packets were not relayed. The RTS and CTS packets were sent on the control channel, which was set aside as a signaling channel, while the DATA and ACK packets were sent on the data channels. The timers were reconfigured and reset according to the inter-cycle design. The rest of the parameters were unchanged and were set to values specified in the IEEE 802.11 standard.

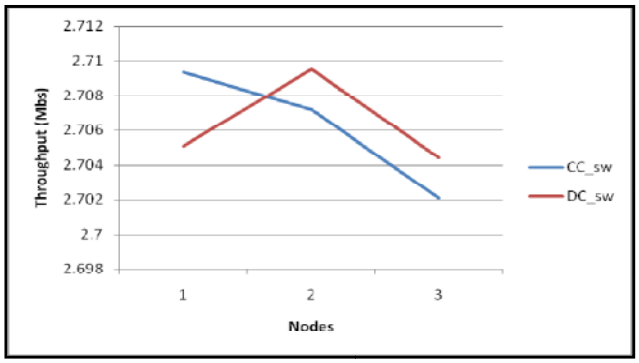


Figure 17. The performance of a network with three data flows subjected under the two channel switching schemes

The smallest network size, which was investigated, had six nodes, three transmitters and three receivers. It can be seen in Figure 17 that the first node in the reference model (AMCP) achieved higher throughput than the CSA. The proposed approach performed better in the second and third nodes. In general, the proposed approach was superior to the reference model.

It should also be noted that in Figure 17, the number of data channels was more than the number of data flows. This resulted

in the highest achieved throughput as compared to the subsequent results where data flows were either equal to or more than the number of data channels. Therefore, the highest achieved throughput in Figure 17 was possible due to less interference experienced in a three data flow network.

The network size was changed in Figure.18; the number of nodes was increased to eight with four transmitters and four receivers. The number of data flows was equal to the number of data channels. There was one to one pairing of data channels to data flows.

It can be noted that there was a slight decrease in achieved throughput in the two schemes in Figure 18. The decrease in achieved throughput was caused by the non availability of a free data channel as was the case in the previous figure. The nodes therefore, did not benefit from the extra data channel, which resulted in a decrease in the achieved throughput. The decrease in achieved throughput was caused by the increase in the amount of interference.

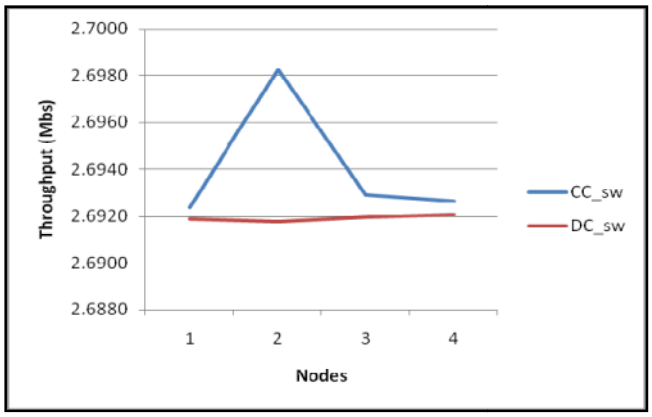


Figure 18. Comparison of two switching approaches with four data flows.

In Figure 18, the reference model was superior to the proposed model in all the data flows. It achieved a higher throughput in all the cases. The reason for the poor performance of the proposed scheme can be attributed to the design of the inter-cycle duration and the observations made on the performance of the inter-cycle duration in the analysis section. It was observed that the inter-cycle duration degrades the performance of the control channel when few data channels are implemented. Its performance improves with the increase of data channels. The duration of the inter-cycle is longer in the smallest possible network and shorter in the largest possible network that can be supported by a single control channel.

The increase in the amount of interference couple with the design aspect of the inter-cycle duration degraded the performance of the proposed protocol in a network with four data flows.

The network size was further increased in Figure.19. The number of nodes was increased to ten with five transmitters and five receivers. The number of data flows was more than the number of data channels. At any given time, there would be one interfering data flow. The interfering data flow caused a severe degradation to the two protocols.

The proposed approach performed better in the last four nodes. It was outperformed in the first node. This shows that the proposed protocol offers better performance as the size of the network is increased. On the other hand, the inter-cycle duration improves the performance of the proposed protocol as more data channels are added. The results in Figure 19 validate this assertion on the performance of the inter-cycle duration.

In Figure 20 the number of nodes was increased to thirty with fifteen data flows. This was the largest network scenario, which was considered in this evaluation. However the achieved throughput was more than the one, which was achieved in Figure. 19. It was almost the same as the throughput, which was achieved in Figure 18. This shows that the proposed scheme is scalable and that the performance of the inter-cycle improves as more data channels are added.

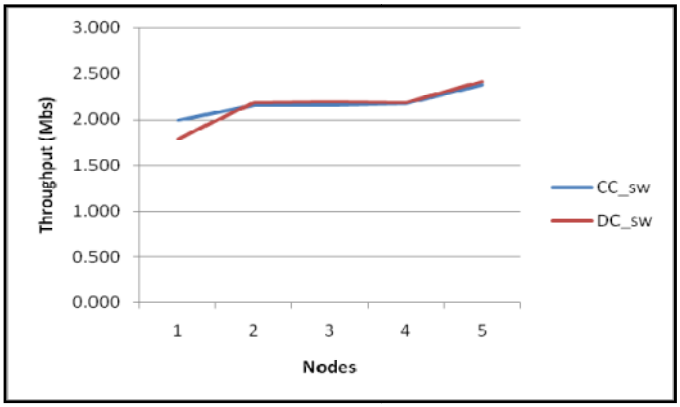


Figure 19. The performance of a system with five data flows implementing both channel switching delay schemes.

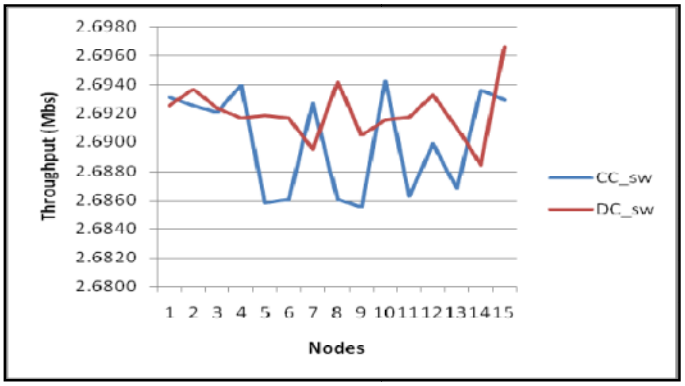


Figure 20. Throughput achieved by the fifteen data flows implementing the two channel switching delay approaches

Figure 20 creates a notion of highly congested and backlogged network. The proposed scheme was evaluated in the largest possible network. As can be seen in the figure, the performance of the proposed scheme was very good. Its performance did improve remarkable in a large network. The performance gains can be attributed to the improved performance of the inter-cycle duration in large networks. Secondly, the achieved throughput remained the same as in Figure. 18 largely due to spatial reuse. These results also show that an interfering data flow causes a significant degradation as compared to the size of the network.

The network in Figure 20 was three times larger than the one in Figure 19. However, its achieved throughput is better despite its size, which would have otherwise degraded its performance. This confirms the argument that the CSA is more scalable and that its performance does improve with network size.

VIII. CONCLUSION

The paper proposed a cyclic scheduling algorithm, which incorporates channel switching delay in channel scheduling and coordination. The scheme has been analyzed through analytical means and through numerical simulations. The analysis shows that the control channel's capacity does not degrade the system. The data channels on average saturate after the seventh node degrading the performance of the system. To improve the capacity of the data channels higher data rates can be considered. They should not be increased to levels that degrade the performance of the control channel.

The numerical results show that the proposed scheme reduces the signaling delay of the control channel. The capacity of the control channel and its scheduling capacity was show to improve with the addition of more data channel. The implementation of the inter-cycle duration is therefore very effective in large networks. The inter-cycle ensures that a data channel is reserved before the current transmission is completed, and that the multi-channel scheduling cost is not repetitive but limited to the first cycle.

The new multi-channel interference problem referred to as the multi-channel scheduling cost in this project, is associated with multi-channel MAC protocols implementing a single control channel as a signaling channel. The analysis shows that the multi-channel scheduling cost can be limited to the first cycle when the cyclic scheduling algorithm is implemented. In the subsequent cycles it can be eliminated by varying the size of the inter-cycle duration and allowing data channels to be reserved when the ongoing transmissions are about to end.

It is envisaged that this protocol provides a platform through, which interference challenges such as the missing receiver problem; the hidden terminal problem and the exposed terminal problem can be addressed. These interference problems can be reduced and cannot be eliminated.

REFERENCES

[1] Jungmin So and Nitin vaidya (2004). Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using A Single Transceiver. MobiHoc 2004, Roppongi, Japan, ACM.

[2] Ritesh Maheshwari, Himanshu Gupta, and Samir R. Das. Multichannel MAC protocols for wireless networks. SECON 2006, Volume 2, Issue , 28-28 Sept. 2006

[3] Jingpu Shi, Theodoros Salonidis, and Edward W. Knightly (2006). Starvation Mitigation Through Multi-Channel Coordination in CSMA Multi-hop Wireless Networks, MobiHoc'06, May 22–25, 2006, Florence, Italy.

[4] S.-L. Wu, C.-Y. Lin, Y.-C. Tseng, and J.-P. Sheu (2000). A New Multi-Channel MAC Protocol with On-Demand Channel Assignment for Multi-Hop Mobile Ad Hoc Networks, in Int'l Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN), 2000.

[5] A. Tzamaloukas and J.J. Garcia-Luna-Aceves (2001). A Receiver-Initiated Collision-Avoidance Protocol for Multi-Channel Networks," in *Proc. of IEEEINFOCOM*, 2001. www.cse.ucsc.edu/ccrg/.../jamal.infocom01.pdf,

[6] Jenhui Chen and Shiann-Tsong Sheu (2004). Distributed multichannel MAC protocol for IEEE 802.11 ad hoc wireless LANs. Computer Communications 28 (2005).

[7] J. Deng and Z. Haas (1998). Dual Busy Tone Multiple Access (DBTMA): A New Medium Access Control for Packet Radio Networks, in *Proc. of IEEE ICUPC*, Florence, Italy, 1998J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[8] Hoi-Sheung Wilson So and Jean Walrand (2006). Design of a Multi-Channel Medium Access Control Protocol for Ad-Hoc Wireless Networks**.** University of California at Berkeley

[9] Hoi-Sheung Wilson So, Jean Walrand, and Jeonghoon Mo. McMAC: A Parallel Rendezvous Multi-Channel MAC Protocol, www.walrandpc.eecs.berkeley.edu/Papers/**McMAC**.pdf

[10] N. Jain and S. Das (2001). A Multichannel CSMA MAC Protocol with Receiver-Based Channel Selection for Multihop Wireless Networks, in *Proc. of the 9th Int.Conf. on Computer Communications and Networks (IC3N), October 2001.*

[11] Jain Nitin, Samir R. Das, and Asis Nasipuri. A multichannel CSMA MAC protocol with Receiver-Based Channel Selection for Multihop Wireless Networks.

[12] Shih-Lin Wu, Yu-Chee Tseng, Chih-Yu Lin, and Jang-Ping Sheu (2002). A multi-channel MAC protocols with power control for multi-hop mobile ad hoc networks. The computer journal, Vol. 45, No. 1, 2002.

[13] Ali Khayatzadeh Mahani, Majid Naderi, Claudio Casetti, and Carla F. Chiasserini (2007). ENHANCING CHANNEL UTILIZATION IN MESH NETWORKS. IEEE MILCOM, Orlando, FL, 2007, 29-31 October 2007. www.ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4454826

[14] A. Nasipuri, J. Zhuang, and S. R. Das (1999). A Multichannel CSMA MAC Protocol for Multihop Wireless Networks, in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, September 1999.

[15] A. Nasipuri and S. R. Das (2000). Multichannel CSMA with Signal Power-based Channel Selection for Multihop Wireless Networks, in *Proc. of IEEE Vehicular Technology Conference (VTC)*, September 2000.

[16] W. Hung, K. Law, and A. Leon-Garcia (2002). A Dynamic Multi-Channel MAC for Ad Hoc LAN," in *Proc. Of 21st Biennial Symposium on Communications*, April 2002.

[17] Arup Acharya, Archan Misra, and Sorav Bansal (2003). MACA-P : A MAC for Concurrent Transmissions in Multi-hop Wireless Networks. Proc. Of PerCom, March 2003

[18] Sudthida Wiwatthanasaranrom and Anan Phonphoem. Multichannel MAC Protocol for Ad-Hoc Wireless Networks. Department of Computer Engineering, Kasetsart University.

[19]] Jeonghoon Mo, Hoi-Sheung Wilson So, and Jean Walrand (2005). Comparison of Multi-Channel MAC Protocols. MSWiM 2005, 10 – 13 October, Montreal, Quebec, Canada.

[20] Pradeep Kyasanur Jungmin So, chandrakanth Chereddi, and Nitin H. Vaidya. Multi-Channel Mesh Networks: Challenge and Protocols. University of Illinois

[21] Myunghwan Seo and Joongsoo Ma (2007). Flexible Multi-channel Coordination MAC for Multi-hop Ad hoc Network. Technical Report, January 2007.

[22] Chandrakanth Chereddi, Pradeep Kyasanur, and Nitin H. Vaidya (2006). Design and Implementation of a Multi-Channel Multi-Interface Network. REALMAN'06, May 26, 2006, Florence, Italy.

[23] EunSun Jung and Nitin H. Vaidya (2002). A Power Control MAC Protocol for Ad Hoc Networks, *MOBICOM'02*, September 23–28, 2002, Atlanta, Georgia, USA.

[24] The editors of IEEE 802.11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Speci_cation, 1997.

[25] Danilo Valerio, Fabio Ricciato, and Paul Fuxjaeger (2008). On the feasibility of IEEE 802.11 multi-channel multi-hop mesh networks, Computer Communications 31 (2008), www.elsevier.com/locate/comcom

[26] Mthulisi Velempini and Mqhele. E. Dlodlo (2009). A multiple channel selection and coordination MAC Scheme. MESH 2009, June 18 – 23 2009, Glyfada/Anthens, Greece.

Service area deployment of IEEE 802.16j wireless relay networks: service area coverage, energy consumption, and resource utilization efficiency

Shoichi Takemori
Graduate School of Information Science and Technology
Osaka University
1–5, Yamadaoka, Suita, Osaka, Japan
s-takemr@ist.osaka-u.ac.jp

Go Hasegawa, Yoshiaki Taniguchi, Hirotaka Nakano
Cybermedia Center
Osaka University
1–32, Machikaneyama-cho, Toyonaka, Osaka, Japan
{hasegawa, y-tanigu, nakano}@cmc.osaka-u.ac.jp

***Abstract*—In wireless relay networks based on IEEE 802.16j, each relay node has its own service area that provides wireless Internet access service to the client terminal. The performance of such networks is heavily affected by how each relay node determines its service area size. In order to determine the service area size for each relay node, it is important to use the location information of other neighboring nodes and their service area sizes. However, in general, such information is completely unknown or only partially known. In the present paper, we introduce three methods to determine the service area size, each of which assumes a different level of the knowledge regarding neighboring nodes. We conduct extensive simulation experiments to evaluate the performance of these three methods in terms of coverage ratio, service area overlap characteristics, energy consumption, and utilization efficiency of wireless network resources. We confirm the trade-off relationships between the knowledge level and performance for these three methods.**

***Keywords*-WiMAX; IEEE 802.16j; relay networks; service area; energy consumption;**

I. INTRODUCTION

IEEE 802.16j relay networks [2] (hereinafter referred to as relay networks), which are often referred to as wireless mesh networks, have received significant attention as extensible, cost-effective means to provide a wide-area wireless broadband access environment. In relay networks, each relay node connects to other relay nodes through wireless links so that the overall topology becomes a tree-like structure, as shown in Figure 1. Each relay node is connected to the Internet through a gateway node that has a wired connection to the Internet. The relay node provides wireless Internet access service to client terminals within its service area. Generally, the wireless channel used for communication between relay nodes and client terminals is different from that used among relay nodes. Communication quality for client terminals is heavily affected by the service area construction. In an area where multiple service areas overlap, the communication quality for the client terminals degrades due to radio interference. IEEE 802.16 uses the TDMA-based transmission mechanism, which assigns time slots to each communication link between relay nodes and client terminals. This means that the links in the overlap area would be assigned different time slots in order to avoid transmission collapse. Therefore, when the overlap area size becomes large or the number of overlapping areas increases, the total number of time slots required for all communication links increases, which decreases the network throughput. Furthermore, since IEEE 802.16 uses the contention-based mechanism for client terminals to join the network [3, 4], service area overlaps would increase access collisions.

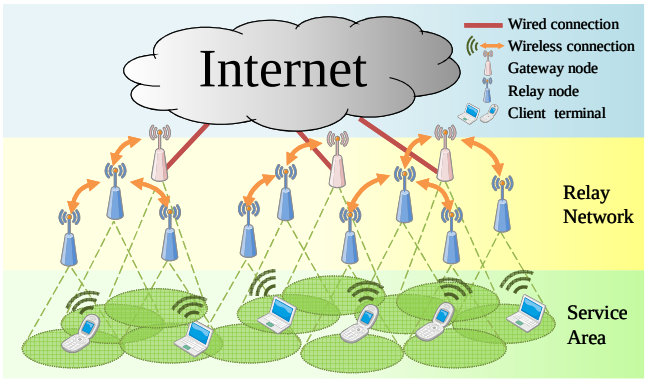


Figure 1: Wireless relay network and client terminals

On the other hand, we should increase the total coverage ratio to provide wide-area and uniform service to client terminals. One possible way to do this is to increase the service area size for each relay node, but this increases the service area overlaps. Furthermore, this also increases the energy consumption of the relay nodes. In other words, there are complex trade-off relationships among network performance, energy consumption, coverage ratio, and service area overlaps.

In order to determine the service area size for each relay node, the information on the location of other neighboring nodes is quite important. However, such information is un-

[1]The present manuscript is an extended version of [1], which was presented at UBICOMM 2009, October 2009.

known, especially when considering the random installation of relay nodes. In other cases, such information is partially known through the topology construction procedure. In the extreme case, we can obtain precise location information of relay nodes when we install the nodes in a well-organized manner. Therefore, various methods are needed to determine the service area size according to the knowledge level of the location information of neighboring relay nodes.

In [1], we proposed two methods, and compared their communication performance in terms of the size of the total service area, the size of the single-covered area, and the maximum number of overlapped service areas through preliminary simulation experiments. We assume that the service area is a circle. These methods determine the radius of this circle (hereinafter referred to as the service radius). The first method, referred to as the identical radius method, is for the situations, in which there is no information about other relay nodes. Therefore, all of the relay nodes use an identical service radius. The second method, referred to as the NND (Nearest Neighbor Distance) method, is used in situations, in which some degree of topology information can be obtained from the topology construction procedure. Using this information, each node estimates the distance to the nearest neighboring node and sets the service radius.

In this paper, we propose the third method. The third method, referred to as the Voronoi method, is used in situations, in which we can obtain precise location information of other neighboring relay nodes. In this method, each relay node sets its service radius to the distance from the relay node to the furthest point in its Voronoi area [5].

These three methods assume a different level of knowledge. The goal for the future is to develop a method whose performance is highly competitive with Voronoi method, using limited information such as topology construction information. For this purpose, in this paper, we conduct extensive simulation experiments to evaluate the performance of three methods in terms of coverage ratio, service area overlap characteristics, energy consumption, and the utilization efficiency of wireless network resources. We confirm the trade-off relationships between the knowledge level and the performance of the three methods. The simulation results reveal that the method that uses more information performs better in terms of all of the above metrics. In particular, the Voronoi method has the best performance among the three methods, whereas the location information cannot be easily obtained. We also confirm that the performance of the NND method is significantly improved by the inclusion of a small amount of additional information, which is readily available.

The remainder of the present paper is organized as follows. In Section II, we discuss research related to the coverage problem. In Section III, we describe the model of the relay networks and the topology construction method. In Section IV, we introduce three methods to determine the service area size. We present the simulation results in Section V. Finally, in Section VI, we conclude the paper and describe areas for future research.

II. Related work

As mentioned in Section I, a relay node must determine its service radius so that overlapping areas becomes small and the total service area becomes large. This problem can be classified as a disc coverage problem, which has been researched extensively with respect to sensor networks [6, 7], relay networks [8], and image processing [9]. However, these studies cannot be applied to the problem in the present paper because the assumptions are quite different. For example, in [6, 7], individual sensors cannot change their coverage ranges. In [9], the authors discussed a method to choose a set of discs from various size discs so that the total area is covered by the smallest possible number of discs and the size of discs is fixed. In [8], the authors focused on the coverage problem in relay networks and investigated the adequate number of relay nodes for client terminals to establish a path to the gateway nodes. They also investigated the process of changing the coverage ratio, which is the ratio of the number of client terminals with an active link to the total number of client terminals. However, the previous authors assumed that they could install additional relay nodes so that the client terminals could establish a path to the gateway node.

III. IEEE 802.16J relay network

In this section, we describe the network model and the radio interference model used in the present paper. We also introduce the topology construction method.

A. Network model

Figure 2 shows the network model used in the present paper. In the network model G, we assume that a set of relay nodes $V = \{v^{(0)}, v^{(1)}, ..., v^{(N-1)}\}$ and a set of client terminals $W = \{w^{(0)}, w^{(1)}, ..., w^{(Q-1)}\}$ are randomly located in the field and that the relay node at the center of the field is the gateway node $(v^{(0)})$.

The relay nodes and client terminals construct a tree topology based on the topology construction procedure described in Subsection III-C. In topology construction, each relay node $v^{(i)}$ sets its transmission power in a non-stepwise manner. Let $d^{(i)}$ denote the transmission distance of each relay node. Here, $M^{(i)}$ represents the number of physical links of node $v^{(i)}$ that connect to neighboring relay nodes. Note that a physical link is defined as a link that is established between two relay nodes located within each other's transmission distance, and an active link is defined as a physical link that constructs a tree topology and is used to transmit data to the gateway node.

We consider two types of wireless communications: wireless communications among relay nodes and wireless communications between relay nodes and client terminals. These two types of communications do not interfere with each

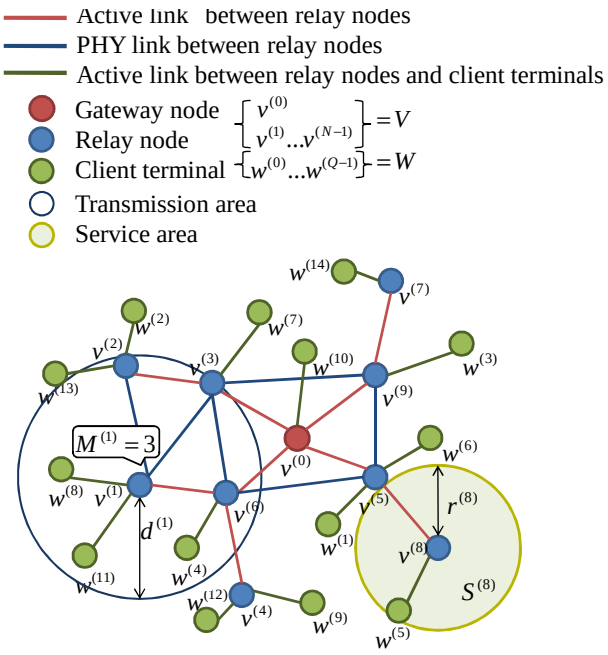


Figure 2: Network model

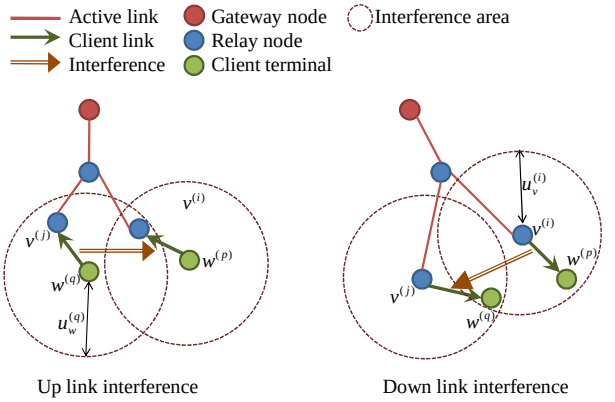


Figure 3: Directed interference model

other because they use different wireless channels [10]. We assume that the service area of relay node $v^{(i)}$ is a circle $S^{(i)}$, the radius of which is $r^{(i)}$, and the service radius can be chosen in a non-stepwise manner. We also assume that client terminals always set a constant transmission distance regardless of the distance to the connecting relay node. After the topology construction introduced in Subsection III-C, each relay node determines its service area size in order to provide wireless Internet access service for client terminals. Each client terminal connects to the nearest relay node having a service radius area that covers the client terminal.

B. Directed interference model

We next introduce the radio interference model [11] used in the present paper, as is depicted in Figure 3. We define a directed communication graph $I = (X, E)$, where $X = V \cup W$ and E is the set of directed communication links $l_{v^{(i)},w^{(p)}}$ that defines an edge directed from relay node $v^{(i)}$ to client terminal $w^{(p)}$. Note that the communication links between relay nodes are not included in E. The model defines the interference relationship between two directed transmission links $l_{v^{(i)},w^{(p)}}$ and $l_{v^{(j)},w^{(q)}}$ based on the distance among four vertices $v^{(i)}$, $v^{(j)}$, $w^{(p)}$, and $w^{(q)}$. In the communication between relay nodes and client terminals, we assume that $v^{(i)}$ and $w^{(p)}$ have interference ranges given by circles of radii are $u_v^{(i)}$ and $u_w^{(p)}$, respectively. In IEEE 802.16j, upward links and downward links become active in different sub-frames. Therefore, we have to consider the following two types of interferences. For upward interference, $l_{v^{(j)},w^{(q)}}$ interferes with $l_{v^{(i)},w^{(p)}}$ when and only when $||v^{(i)} - w^{(q)}|| < u_w^{(q)}$. For downward interference, $l_{v^{(i)},w^{(p)}}$ interferes with $l_{v^{(j)},w^{(q)}}$ when and only when $||v^{(i)} - w^{(q)}|| < u_v^{(i)}$. Here, $||v^{(i)} - w^{(q)}||$ is the distance between $v^{(i)}$ and $w^{(q)}$. Typically, $u^{(i)} > d^{(i)}$, and the ratio of the interference range to the communication range for node $v^{(i)}$, denoted as $\gamma^{(i)} = \frac{u^{(i)}}{d^{(i)}}$, takes ranges from 2 to 4 in practice [11]. When the two links interfere, they are not used at the same time slot in the IEEE 802.16 frame.

C. Topology construction procedure

IEEE 802.16j does not define the details of the method used to construct the network topology [12]. The topology construction procedure we introduce here is targeted at situations, in which relay nodes are deployed randomly. The typical situation includes network construction at a disaster site, where it is difficult to deploy relay nodes in a well-organized manner. We also assume that relay nodes are deployed to the network incrementally. In the following, we explain in detail the algorithm for a newly joining relay node to connect to an existing network topology.

A newly joining relay node, denoted as $v^{(i)}$, waits for replies from other existing relay nodes while increasing its radio transmission power [3]. When a found relay node does not have a path to a gateway node, $v^{(i)}$ establishes a physical link to the relay node and continues increasing its transmission power. When the found relay node has a path to the gateway node, $v^{(i)}$ establishes an active link to the relay node and stops increasing its transmission power. When $v^{(i)}$ cannot find any relay nodes even when the transmission power reaches the maximum, $v^{(i)}$ maintains the maximum transmission power to detect the joining relay nodes which is successfully linked to the gateway node. If the relay node finds multiple relay nodes that have a path to the gateway node at the same time, the relay node sets an active link to the relay node that has the smallest hop count to the gateway node. When multiple relay nodes have an identical

Algorithm 1 Algorithm for complete coverage of the field using the identical radius method

Input: topology T_n
Output: length of service radius r_{idt}

```
Declare variable tempDist
Declare variable coverageRatio
tempDist = 0.6
coverageRatio = 100%
while coverageRatio = 100% do
    tempDist = tempDist − 0.005
    Calculate the coverageRatio using all of the service radii r^(i) = tempDist
    if coverageRatio < 100% then
        BREAK
    end if
end while
Return r_idt = tempDist + 0.005
```

hop count to the gateway node, $v^{(i)}$ chooses the relay node that is nearest $v^{(i)}$. When a relay node $v^{(j)}$ that does not have an active link finds a path to the gateway node as a result of the entry of $v^{(i)}$, $v^{(i)}$ and $v^{(j)}$ set active links as the shortest path to the gateway node.

Note that this topology construction procedure may generate isolated relay nodes, which do not connect any other relay nodes and which cannot provide a service area, depending on its transmission distance, deployment order of relay nodes, and location. The simulation experiments shown in Section V include such a case.

IV. Methods to Determine the Service Area Size

A. Identical radius method

As the simplest method, we first discuss the identical radius method. This method is used in situations, in which there is no information about other relay nodes. In this method, all relay nodes use an identical service radius. This method determines the service radius as the minimum value so that the coverage ratio becomes 100%.

Each relay node determines its service radius $r^{(i)}$ according to the following equation, where r_{idt} is obtained by algorithm 1:

$$r^{(i)} = r_{idt}. \quad (1)$$

Figure 4 shows an example of coverage with the identical radius method.

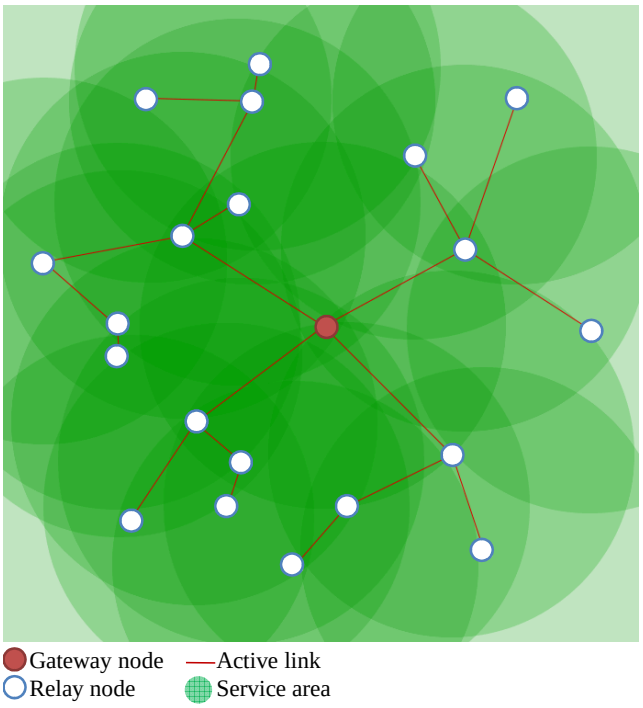


Figure 4: Example of coverage by the identical radius method

B. Nearest neighbor distance method

Nearest neighbor distance (NND) method is used in situations, in which some degree of topology information can be obtained by the topology construction. The NND method estimates the nearest neighbor distance, which is the distance to the nearest neighboring relay node, based on the node density, using this information. This estimation result is used to determine the service radius.

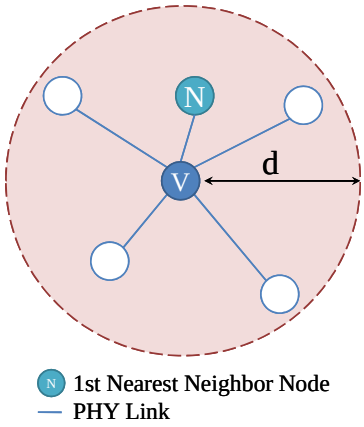


Figure 5: Estimation of the nearest neighbor distance

1) Estimation of the nearest neighbor distance: Figure 5 shows the situation, in which relay node v has M (= five) neighboring nodes in its transmission distance, which means that relay node v has M physical links. Since we assume the random location of all relay nodes, M neighboring relay nodes are randomly located in the circle for which the center is located at node v and the radius is d. We then derive the nearest neighbor distance using parameters M and d.

First, we consider a lemma for the estimation.

Lemma IV-B.1. *When the node density is constant and*

there are M relay nodes in a circle C_α of radius α, the average distance $\overline{R}$ between these nodes and the center node becomes:

$$\overline{R} = \frac{2\alpha}{3}.$$

Proof: Let $P_{(t|0\leq t\leq \alpha)}$ be the node density function on the circumference of a circle of radius t in C_α. Since nodes are randomly located, $P_{(t)}$ is proportional to t. Therefore, with invariable c, $P_{(t)}$ becomes as follows:

$$P_{(t)} = c \cdot t. \tag{2}$$

Here, since the number of relay nodes in C_α is M, we have

$$\int_0^\alpha P(t)dt = M. \tag{3}$$

Using Eq. (2) and (3), c satisfies

$$c = \frac{2M}{\alpha^2}. \tag{4}$$

Substituting Eq. (4) into Eq. (2), we obtain the following:

$$P(t) = \frac{2M}{\alpha^2}t. \tag{5}$$

Therefore, $\overline{R}$ becomes as follows:

$$\begin{aligned} \overline{R} &= \frac{1}{M} \cdot \int_0^\alpha t \cdot P(t)dt \\ &= \frac{2\alpha}{3}. \end{aligned} \tag{6}$$

■

Using lemma IV-B.1, we have the following theorem.

Theorem IV-B.1. *If the number of PHY links M and the transmission distance d of a relay node are known, the average distance to the nearest neighbor node $\overline{R_1}$ becomes as follows:*

$$\overline{R_1} = \frac{2d}{3\sqrt{M}}.$$

Proof: Considering that the fraction between 1 and the number of relay nodes in the circle C_d is equal to the fraction between the area of C_{R_1} and C_d, we have

$$\frac{M}{\pi d^2} = \frac{1}{\pi (R_1)^2}. \tag{7}$$

Thus, R_1 satisfies

$$R_1 = d\sqrt{\frac{1}{M}}. \tag{8}$$

Using lemma IV-B.1, the average distance to the nearest neighbor node $\overline{R_1}$ becomes as follows:

$$\overline{R_1} = \frac{2d}{3\sqrt{M}}. \tag{9}$$

■

2) Accuracy of estimation: We conducted simulations to evaluate the performance of the estimation algorithm. In this simulation, 100 relay nodes are randomly deployed in a 1×1 field according to the joining process described in Subsection III-C. We then obtain the estimated and actual values of the nearest neighbor distance for all relay nodes. The simulation was conducted 100 times for different node locations. Figure 6 show the relation between the actual and estimated values of the nearest neighbor distance. The two straight lines represent the least-squares approximation and the $y = x$ relationship between the actual and estimated values. When the relay node estimates its nearest neighbor distance precisely, the dots in the graph should be plotted on the $y = x$ line. However, as shown in Figure 6, this algorithm has a large estimation error. Moreover, this algorithm overestimates the nearest neighbor distance when it is less than 0.05 and underestimates the nearest neighbor distance when it is larger than 0.05 on average. This error stems from the topology construction procedure. As mentioned in Subsection III-C, a relay node stops increasing its transmission power when the node finds a neighboring node that has a path to the gateway node. This means that, in most cases, some neighboring nodes exist on the circumference of a circle. On the other hand, this algorithm assumes that the neighboring nodes exist randomly in the circle.

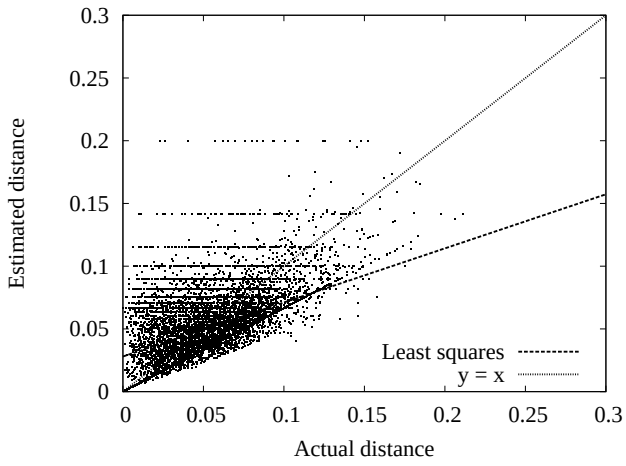


Figure 6: Estimation accuracy

3) Algorithm for complete coverage of the field using the NND method: In the NND method, each relay node sets its service radius based on the estimation result of the nearest neighbor distance (Eq. 9), where d and M can be obtained through the topology construction procedure. Specifically, each relay node determines its service radius $r^{(i)}$ according to the following equation:

$$r^{(i)} = \overline{R_1^{(i)}} \times k_{nnd}, \tag{10}$$

where $\overline{R_1^{(i)}}$ is the nearest neighbor distance of node $v^{(i)}$ defined in Eq. 9 and k_{nnd} is the parameter for determining the overall degree of the service radius (referred to in

Algorithm 2 Algorithm for complete coverage of the field using the NND method

Input: Topology T_n
Output: Value of service ratio k_{nnd}

Declare variable $tempRatio$
Declare variable $coverageRatio$
$tempRatio = 5$
$coverageRatio = 100\%$
while $coverageRatio = 100\%$ **do**
  $tempRatio = tempRatio - 0.1$
  Calculate the $coverageRatio$ using all of the service radii $r^{(i)} = \overline{R_1^{(i)}} \times tempRatio$
  if $coverageRatio < 100\%$ **then**
    BREAK
  end if
end while
Return $k_{nnd} = tempRatio + 0.1$

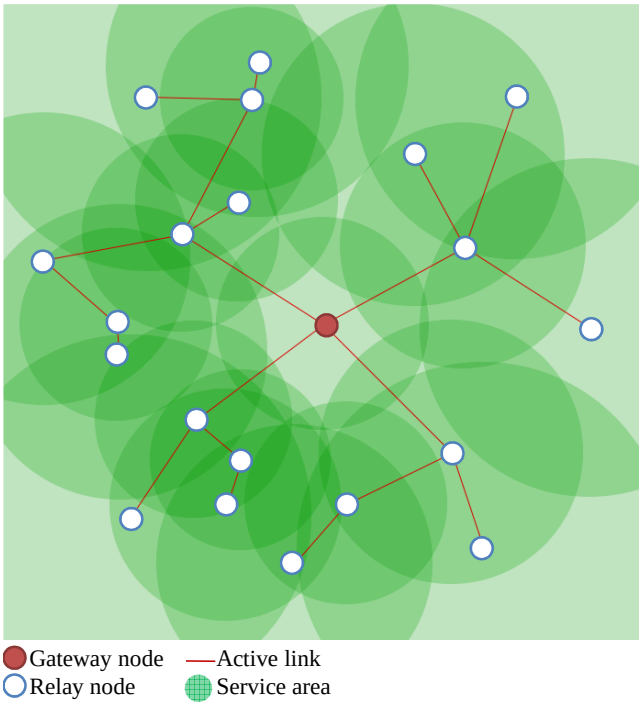


Figure 7: Example of coverage by the NND method

algorithm 2 as the service ratio). Using algorithm 2, we set k_{nnd} so that the coverage ratio becomes 100%. Figure 7 shows an example of coverage with the NND method.

C. Voronoi method

As mentioned in Subsection III-C, we assume that each client terminal connects to the nearest neighboring node. This means that each relay node accepts connections from the client terminals in its Voronoi area [5]. Based on this

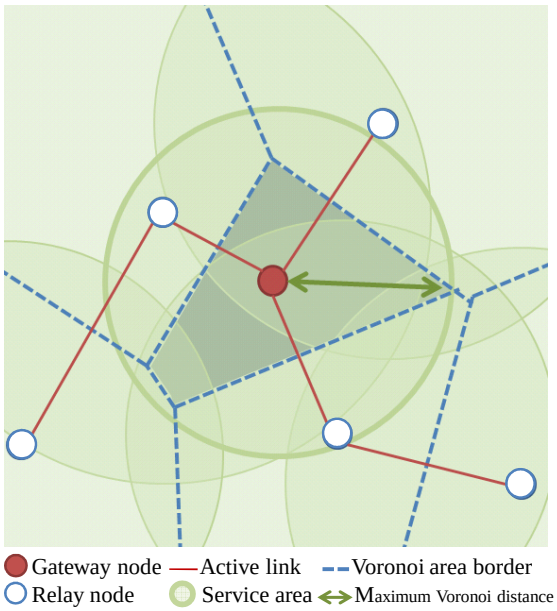


Figure 8: Voronoi method

observation, in the Voronoi method, each relay node set its service radius to cover its Voronoi area. Figure 8 shows an image of the coverage obtained by using the Voronoi method. We assume in this method that we can obtain precise location information of all neighboring relay nodes and calculate the Voronoi area. The Voronoi method sets the service distance to the farthest point in the Voronoi area (referred to in Figure 8 as the maximum Voronoi distance). Note that this method can provide 100% of the coverage ratio because the coverage ratio is based on the Voronoi area. Figure 9 shows an example of coverage obtained by using the Voronoi method.

V. Performance Evaluation

We show the evaluation results for the three methods proposed in Section IV by conducting simulation experiments.

A. Simulation settings and performance metric

In the simulation, a gateway node is located at the center of a 1×1 field, and 99 or 49 relay nodes and 500 client terminals are randomly located. As described in Subsection III-C, the relay nodes construct the network topology based on their location and transmission distance and determine the service area size using each method. Each client terminal connects to the nearest neighboring relay node. The maximum transmission distance of relay nodes and client terminals to construct network topology for communication between relay nodes is set to 0.3, and γ is set to 2. Each plot in the following graphs is the average of $1,000$ simulation experiments.

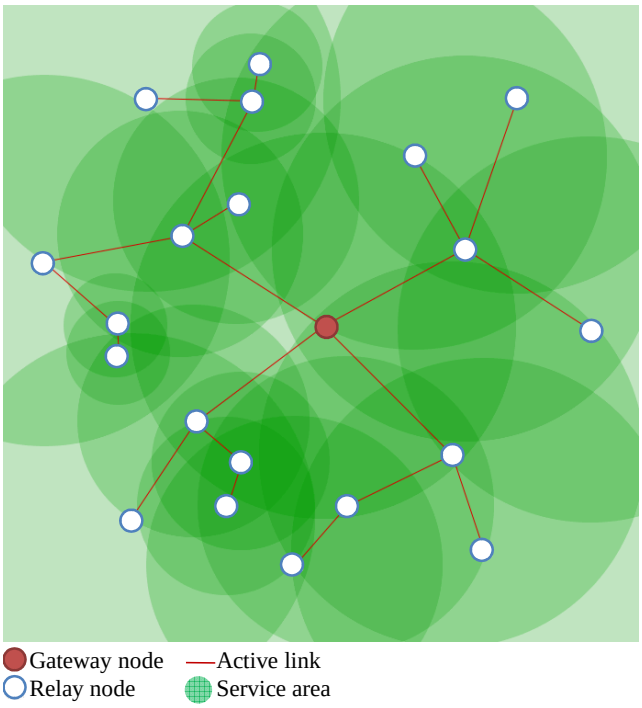


Figure 9: Example of coverage by the Voronoi method

We evaluate the performances of the three methods from following three viewpoints:

- Overlapping service area: We define the *overlap number* of a certain point in the field as the number of relay nodes having a service area that includes the point. We then look at the size distribution of the area in the field summarized by the overlap number. We also use the maximum overlap number, which is defined as the maximum value of the overlap number in the field.
- Energy consumption: We define the energy consumption $E^{(i)}$ of the relay node $v^{(i)}$ as follows [13, 14]:

$$E^{(i)} = {r^{(i)}}^2. \quad (11)$$

  Then, we calculate the energy consumption of the entire relay network as the sum of the energy consumptions of the relay nodes in the network.
- Wireless resource efficiency: In TDMA-based wireless multi-hop relay networks, in order to prevent radio interference, different time slots are assigned to the links that interfere with each other. The frame length is defined as the sum of the number of time slots that are assigned to all of the network links. When the frame length becomes smaller, we can say that we achieve better efficiency of wireless network resources. Therefore, we adopt the frame length as the metric for wireless resource efficiency. Note that we evaluate the frame length for downlink transmissions since the methods in Section IV affects the performance of downlink communication between gateway nodes and relay nodes. By sharing time slots among multiple wireless links that do not interfere with each other, the frame length can be reduced. Note that the method to determine the frame length is beyond the scope of the present paper. However, in the following, we briefly explain the algorithm presented in [11] used to determine the frame length.
  The problem of finding the time slot assignment to the minimize frame length is known as the NP-hard problem [15]. A heuristic algorithm to obtain the interference-free time slot assignment with a small frame length, based on the greedy algorithm is presented in [11], where this resource allocation problem was considered as a point coloring problem and the number of colors was considered as the frame length. This algorithm is divided into the following two major steps.

  Step 1 Determine the order of links for the coloring based on a conflict graph

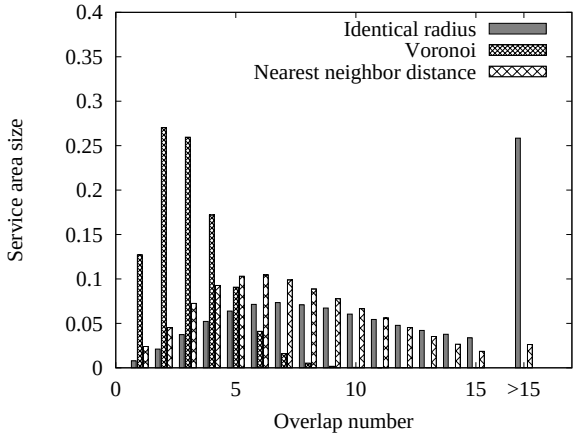
  Step 2 Color each link using the greedy algorithm by the order determined in Step 1

  See [11] for the detailed algorithm. In general, the frame length becomes smaller when the degree of radio interference in the relay network is small. Therefore, the frame length is one possible metric for evaluating service area deployment algorithms.

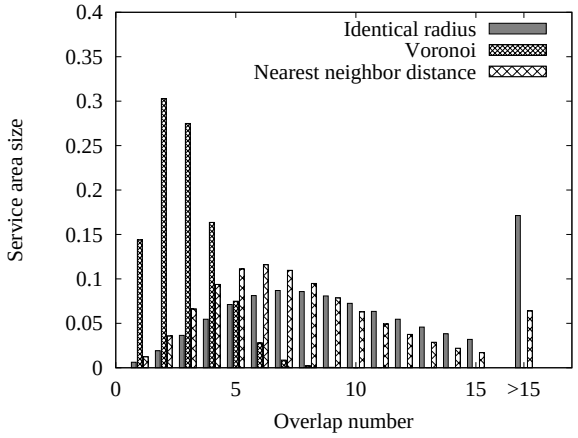
Note that we do not evaluate the coverage ratio because we set the service radius to the minimum value so that the coverage ratio reaches 100%.

B. Overlapping service area

Figures 10(a) and 10(b) show the area size distributions for each overlap number when we use 50 and 100 relay nodes, respectively. In this figure, the area size with an overlap number of 1 represents the single-covered area size, and k of the overlap number indicates that the area is covered by the service areas from k relay nodes. We summarize the area size where the overlap number is larger than 15 at the rightmost bins in the graph. Based on these figures, we observe that when we compare the NND method and identical radius method, the NND method slightly outperforms the identical radius method because the NND method has smaller values of the overlap number than the identical radius method. On the other hand, the results obtained by using the Voronoi method are much better, where roughly 80% of the field has an overlap number of less than five. This indicates the effectiveness of the Voronoi method. Figure 11 shows the CDF of the maximum overlap number for $1,000$ simulation experiments. In Figure 11, we can observe that, as the number of deployed relay nodes increases, the maximum overlap number also increases when we apply the

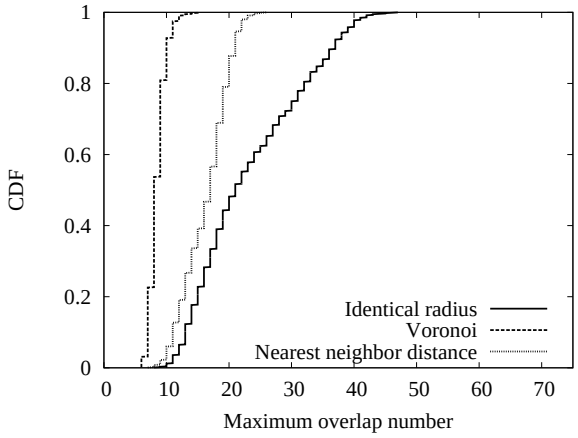


(a) 50 relay nodes

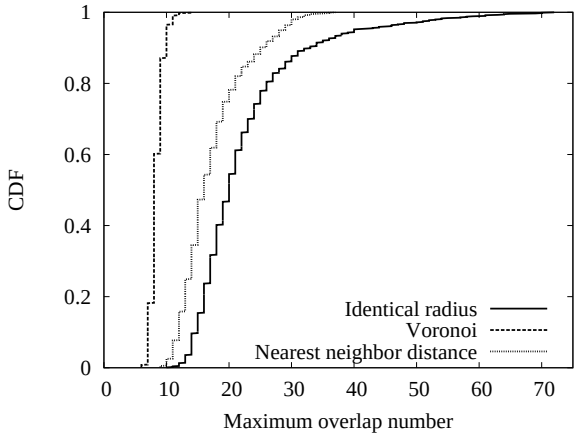


(b) 100 relay nodes

Figure 10: Size distribution of the overlapping area



(a) 50 relay nodes



(b) 100 relay nodes

Figure 11: Distribution of the maximum overlap number

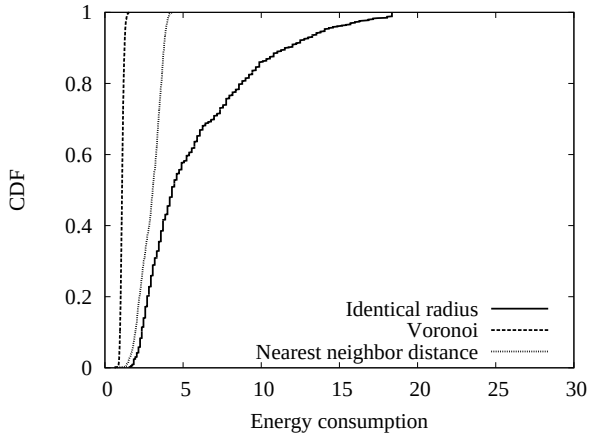
NND method and the identical radius method. On the other hand, when we apply the Voronoi method, the maximum overlap number is not affected by the number of relay nodes.

C. Energy consumption

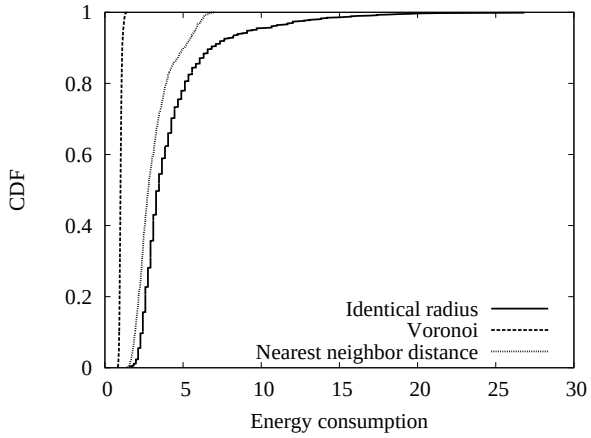
Figure 12 shows the CDFs of the energy consumption of 1,000 simulation experiments for networks with 50 and 100 relay nodes, respectively. Figure 12 resembles Figure 11, which means that the service area coverage efficiency largely affects the energy consumption efficiency. In addition, for the Voronoi method, there is a little difference between the results obtained for 50 relay nodes and the results obtained for 100 relay nodes. This indicates the effectiveness of the Voronoi method, which requires precise location information of all of the relay nodes in the network.

D. Wireless resource efficiency

Figure 13 shows the CDF of the frame length of $1,000$ simulation experiments. We can observe that the frame length of the Voronoi method becomes 54% in the case of 50 relay nodes and 59% in the case of 100 relay nodes, as compared to the identical radius method. For the NND method, the frame length in the case of 50 relay nodes is 82% and that in the case of 100 relay nodes is 119%, respectively, as compared to the identical radius method. This means that, in the NND method, the frame length is affected by the node density. However, as shown in Figure 13, the variance of the identical radius method is larger than that of the other methods. This is because the identical radius method cannot accommodate the biased node density.

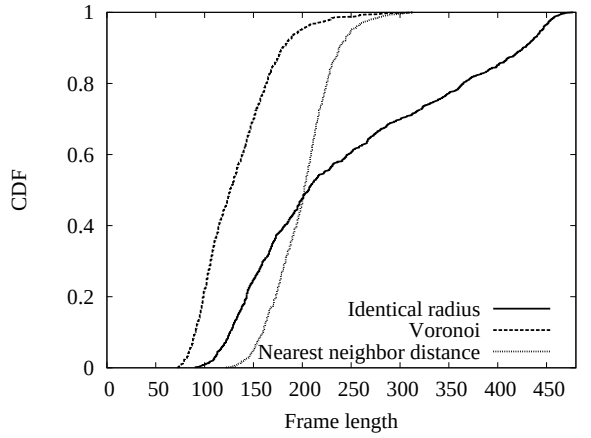


(a) 50 relay nodes

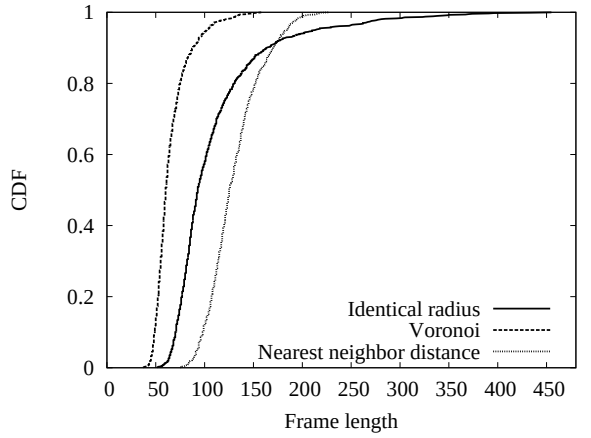


(b) 100 relay nodes

Figure 12: Distribution of network energy consumption



(a) 50 relay nodes



(b) 100 relay nodes

Figure 13: Frame length distribution

Therefore, it is necessary to set a large service radius in order to obtain a coverage ratio of 100%, which causes a large degree of radio interference and a large frame length.

VI. Conclusion and Future Work

In the present paper, we introduced three methods to determine the service area size for IEEE 802.16j relay nodes, referred to as the identical radius method, the NND method, and the Voronoi method, each of which assumes a different level of knowledge regarding neighboring nodes. The identical radius method is the simplest method, in which all relay nodes use an identical service radius. The NND method is used for situations, in which some degree of topology information of the relay network can be obtained. In this method, each relay node estimates the nearest neighbor distance and sets its service radius based on the estimation results. The Voronoi method is used in situations, in which we can obtain precise location information of other neighboring relay nodes. In this method, each relay node sets its service radius to cover its Voronoi area. Through performance evaluation in terms of service area overlap characteristics, energy consumption, and utilization efficiency of wireless network resources, we confirmed the trade-off relationships between the knowledge level and performances of the three methods considered herein. Specifically, the Voronoi method performs significantly better than the other methods.

In future studies, we intend to develop a method in which each relay node can obtain the shape of its Voronoi area using less information, whose performance is highly

competitive with Voronoi method. Furthermore, we intend to evaluate the effect of the service area size on connection establishment procedure based on the contention-based request mechanism.

ACKNOWLEDGMENTS

The authors would like to thank Ms. Akiko Miyagawa for her invaluable support.

REFERENCES

[1] S. Takemori, G. Hasegawa, Y. Taniguchi, and H. Nakano, "Improving coverage area quality using physical topology information in IEEE 802.16 mesh networks," in *Proceedings of UBICOMM 2009*, pp. 163–168, Oct. 2009.

[2] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks*, vol. 47, pp. 445–487, Mar. 2005.

[3] G. Nair, J. Chou, T. Madejski, K. Perycz, D. Putzolu, and J. Sydir, "IEEE 802.16 medium access control and service provisioning," *Intel Technology Journal*, vol. 8, Aug. 2004.

[4] J. Delicado, F. M. Delicado, and L. Orozco-Barbosa, "Study of the IEEE 802.16 contention-based request mechanism," *Telecommunication Systems*, vol. 38, pp. 19–27, June 2008.

[5] F. Aurenhammer, "Voronoi diagrams—a survey of a fundamental geometric data structure," *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, 1991.

[6] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, "Coverage problems in wireless ad-hoc sensor networks," in *Proceedings of INFOCOM 2001*, vol. 3, pp. 1380–1387, Apr. 2001.

[7] C.-F. Huang and Y.-C. Tseng, "The coverage problem in a wireless sensor network," in *Proceedings of WSNA 2003*, pp. 115–121, Sept. 2003.

[8] S. Allen, S. Hurley, V. Subodh, and R. Whitaker, "Assessing coverage in wireless mesh networks," in *Proceedings of MeshNets 2005*, July 2005.

[9] T. Asano, P. Brass, and S. Sasahara, "Disc covering problem with application to digital halftoning," in *Proceedings of ICCSA 2004*, pp. 11–21, Apr. 2004.

[10] D. Ghosh, A. Gupta, and P. Mohapatra, "Adaptive scheduling of prioritized traffic in IEEE 802.16j wireless networks," in *Proceedings of WiMob 2009*, pp. 307–313, Dec. 2009.

[11] W. Wang, Y. Wang, X.-Y. Li, W.-Z. Song, and O. Frieder, "Efficient interference-aware TDMA link scheduling for static wireless networks," in *Proceedings of MobiCom 2006*, pp. 262–273, Sept. 2006.

[12] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16j relay-based wireless access networks: An overview," *Wireless Communications*, vol. 15, pp. 56–63, Oct. 2008.

[13] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of HICSS 2000*, p. 8020, Jan. 2000.

[14] R. Bhatia and M. Kodialam, "On power efficient communication over multi-hop wireless networks: Joint routing, scheduling and power control," in *Proceedings of INFOCOM 2004*, vol. 2, pp. 1457–1466, Mar. 2004.

[15] B. N. Clark, C. J. Colbourn, and D. S. Johnson, "Unit disk graphs," *Discrete Mathematics*, vol. 86, pp. 165–177, Dec. 1990.

MIMO Capacity of Wireless Mesh Networks

Sebastian Max, Bernhard Walke
Communication Networks (ComNets) Research Group
Faculty 6, RWTH Aachen University, 52074 Aachen, Germany
Email:{smxlwalke}@comnets.rwth-aachen.de

***Abstract*—A Wireless Mesh Network (WMN) serves to extend the wireless coverage of an Internet gateway by means of Mesh Stations (MSTAs) that transparently forward data between Stations (STAs) and the gateway. This concept reduces deployment costs by exchanging the multiple gateways, required to cover a larger area with wireless Internet access, by a wireless backbone. Unfortunately, this also reduces capacity, owing to multiple transmissions of the same data packet on its multi-hop route. Hence, different mechanisms to increase the capacity of WMNs are investigated.**

Multiple Input/Multiple Output (MIMO) is a technique that is able to increase the capacity of a single link in the same bandwidth and transmission power: Both the transmitter and the receiver is configured with multiple antennas. If multiple streams are transmitted in a rich scattering environment, these streams can be separated and decoded by the receiver successfully.

However, it is unclear how this single-link capacity increase translates into a system capacity increase in a WMN. In this paper, we will combine a realistic MIMO model with a capacity calculation framework to show the combined effect of the two technologies. The results show that although not the full link capacity increase of MIMO can be exploited, especially WMNs benefit from the MIMO gain.

***Keywords*–Wireless Networks, Capacity, Multiple-Input/Multiple-Output (MIMO), Mesh**

I. Introduction

In the last years, two parallel research areas have produced fundamental innovations for wireless data networks: First, the exploitation of multipath propagation by multiple transmit and receive antennas to increase the link capacity using the same bandwidth. Second, the upcoming of Mobile Ad-Hoc Networks (MANETs) where data is forwarded by intermediate nodes on dynamic, self-configured paths to extend the range of a single wireless link.

In both areas the innovations have successfully found their way into standards and products: Multi-antenna technology, also known as Multiple Input/Multiple Output (MIMO), is for example a crucial part of the latest amendment of IEEE 802.11, "n" [2], to reach the maximum gross throughput of 600 Mb/s (using, among other techniques, 4 transmit and receive antennas). And while MANETs are not deployed themselves, the results from the research of wireless path selection protocols are now standardised and implemented in Wireless Mesh Networks (WMNs), e. g., in [3]: In contrast to MANETs, data forwarding is restricted to special Mesh Stationss (MSTAs), implementing the *mesh facility*. This facility enables forwarding of frames between MSTAs so that for example the limited radio coverage of a Internet-connected MSTA (named *mesh gate* according to [3]) is extended without new wires. Throughout the paper, it is assumed that MSTAs also provide the Access Point (AP) facility for association and management of mobile Stations (STAs). From their viewpoint, the coverage extension via relaying is completely transparent.

The two research areas are parallel two each other because they are applied to different layers of the OSI/ISO protocol stack: MIMO is a technology applied mostly by the Physical Layer (PHY), plus some intelligence in the Medium Access Control (MAC) required for the advanced Rate Adaptation (RA) that now incorporates, next to the selection of the Modulation- and Coding Scheme (MCS), the number of streams to be transmitted. In contrast, the ability to forward data transparently for the application over multiple hops is at the heart of the Network Layer (NL), with probably some improvements in the MAC to measure the wireless link quality or to schedule multi-hop transmissions. Hence, due to the characteristics of the OSI/ISO protocol stack model, it is straightforward to combine the advances of the PHY to those of the NL; as a matter of fact, the two technologies should be transparent to each other.

While this statement is true from a qualitative perspective, its quantitative implications are unknown: In theory, MIMO provides a link capacity improvement which scales linearly with the number of transmit/receive antennas. Of course, this capacity increase would be advantageous especially for the wireless backbone between the MSTAs where the aggregated data of the mobile STAs is transported. However, it is not clear how much of the link capacity increase of MIMO remains to the system capacity of the WMN.

The scope of this paper is to estimate the improvements of MIMO in a WMN. It is structured as follows: After reviewing the related work on capacity estimation of wireless networks in Section II, Section III details the applied system model. At first, this system model is

applied to a single Basic Service Set (BSS), consisting of multiple STAs and a single AP, in Section IV to estimate the upper bound BSS capacity for a WMN. Then, Section V introduces the capacity calculation method for WMNs. This calculation method is applied in Section VI to evaluate the effect of different MIMO configurations. Finally, the paper concludes with Section VII.

II. Related Work

Capacity calculation of wireless communication networks is a popular research topic. Two major trends have evolved:

1) To determine the capacity bound of a random network with certain properties, where the capacity is considered to be a random variable and asymptotic properties are calculated, and
2) To compute the capacity of a given, arbitrary network using graph-theory based algorithms.

Besides these trends, work has been published that considers a given network for calculating its capacity from the shortest possible schedule by means of Linear Programming.

1) Analytical Bounds: In their seminal paper Gupta and Kumar [4] explored the limitations of multi-hop radio networks with random source-destination traffic relationships by computing the achievable throughput for a random network obtained under optimal conditions to be $\Theta\left(W/\sqrt{n}\right)$, where n is the number of nodes and W the radio bandwidth.

Gupta and Kumar conclude that efforts should be targeted to small networks, where nodes communicate with near neighbours only.

Several researchers have considered to extend this basic model, e. g., by incorporating different network structures [5] or mobility of nodes [6]. Due to the same approach chosen, these papers have in common that they derive asymptotic scaling laws to describe the capacity bounds for the considered random network. Application of these results to any real WMN instance with a given topology appears not to be possible.

2) Graph-based capacity calculation: The mentioned disadvantage is avoided when concentrating on the calculation of capacity bounds in a given network instance. In [7] and [8] this is done by translating the properties of the wireless medium (shadowing, interference, receive probability) into two graphs: The connectivity graph $G = (V_G, E_G)$ and the conflict graph $C = (V_C = E_G, E_C)$. While in G each vertex represents a node and an edge represents a link between two nodes, C represents links that cannot transmit simultaneously. The capacity of the network is computed then using methods from graph-theory.

Since computation of the capacity bounds of a given wireless network is NP-complete even under simple assumptions [9], approximation algorithms must be used.

For real-world wireless networks where link adaptation is state of the art, the problem is even aggravated: A node may choose among alternate MCSs to be used for a transmission. If a high-rate, but interference susceptible MCS is chosen, the link should be operated under low interference only, in contrast to a more robust, low-rate MCS that may function well under high interference. Hence, it is impossible to generate the conflict graph C without assigning to each link a single MCS in advance. Therefore, most papers on capacity calculation restrict the link model to one MCS only, thereby ignoring an important characteristic of current wireless standards.

3) Linear Programming-based capacity calculation: Algorithms published for calculating the system capacity of a given network taking different MCSs into account all use a model similar to the one introduced in [10]: Alternate assignments of MCSs are compared by computing the set of achievable data rate combinations between all source-destination pairs in the network. *Basic Rates* are introduced as a key element, describing a set of links active at a given time. The challenge is to find the schedule of all feasible basic rates that minimises the schedule's duration. A capacity bound can be derived from the duration of the shortest schedule and the amount of carried traffic. Consistent with [9], the number of basic rates and thus the algorithm runtime complexity grows exponentially, rendering it useless for networks with more than 30 nodes.

Reference [11] proves the computation of the shortest schedule to be NP-complete and extends the work of [10] by a column-based approach to solve the Mixed-Integer Programming (MIP)-formulation of the optimisation problem. Although this method makes use of modern branch-and-price methods to solve the MIP, large networks with 40 and more nodes cannot be solved exactly. Instead it is proposed to stop the branching process using a heuristic.

4) Previous Publications by the Authors: The concept of linear programming-based capacity calculation is picked up by the authors in [12], where the heuristics Selective Growth (SG) and Early Cut (EC) to control the number of network states are introduced and applied to WMNs first. A more detailed analysis and the additional heuristic Selective Growth/Delete (SG/Del) is provided by [13]. These extensions to the linear programming method are crucial to compute the capacity of large-scale scenarios with 100 and more nodes.

The improved calculation methods have been applied by the authors to calculate the capacity of WMNs under different conditions: [14] considers hybrid wireless/wired mesh networks, [15] uses Ultra Wideband

(UWB) as transmission technology and [1] shows the effect of transmit power control on the capacity. Throughout the publications, the capacity calculation method has proven to be a versatile tool to estimate the effect of a PHY technology to the system capacity.

III. System Model

The system model is especially concerned with the characteristics of a wireless network, i. e., the the wireless channel and the performance capability of the PHY to transmit information using the wireless channel. In compliance with the topic of the paper. According to the topic of the paper, special treadment is given to model MIMO transmissions.

A. Wireless Channel Model

The wireless channel determines the received signal strength of a transmission from node N_i to N_j, positioned at p_i and p_j, respectively. Typically, a wireless channel model is of the form

$$\mathrm{P}\left(N_i, N_j\right)[\mathrm{dBm}] = \mathrm{P}_i + g_i + g_j \quad (1)$$
$$- \mathrm{pl}(p_i, p_j) - \mathrm{s}(p_i, p_j) \quad (2)$$

where

- P_i is the transmission power of node N_i;
- g_i and g_j are optional transmit and receive antenna gains;
- $\mathrm{pl}(p_i, p_j)$ is the pathloss function that models the attenuation of the radio wave due to the distance between N_i and N_j;
- $\mathrm{s}(p_i, p_j)$ is a shadowing fading component having log-normal distribution.

The system performance highly depends on the characteristics of this model, i. e., the parameterisation of $\mathrm{pl}(p_i, p_j)$ and $\mathrm{s}(p_i, p_j)$. Therefore, it is crucial that the selection is based on extensive real-world measurements campaigns. Furthermore, the usage of a standardised model allows direct comparison with results from the literature that use the same assumptions.

Based on these considerations and the typical application scenario of a WMN, we select the Urban Micro (UMi) channel model described in [16], which is designed to evaluate radio interface technologies in the IMT-Advanced process. This model provides pathloss functions for Line Of Sight (LOS), Non Line Of Sight (NLOS) and Outdoor-to-Indoor (OtoI) links as well as a description of a random process with correlated log-normal distribution for the shadowing fading.

B. Physical Layer Model

The Physical Layer (PHY) model decides under which conditions a packet transmission is successful, i. e., the packet is decoded error-free at the receiver.

In our model, the success probability depends on two factors:

1) How much noise from the background and other active transmissions interferes with the signal and
2) which MCS is selected at the transmitter.

Throughout the paper we will assume IEEE 802.11n-2009 [2], as the physical layer standard. Adaptation of the methodology to other wireless transmission technologies based on Orthogonal Frequency Division Multiplexing (OFDM) is possible by adapting the calculations to different MCS; however, this is not in the scope of this paper.

As in [2], the MCS does not only comprise the modulation and the channel coding, but in addition the number of spatial streams n_{ss}. Hence, the MCS comprises all information that determines the number of data bits per OFDM symbol.

1) SINR: Signal degradation at the receiver is caused by two factors: First, the thermal- and receiver noise; second, interference from other active transmissions.

The power of thermal noise (dBm) is given by $N_{th} = -174 + 10\log_{10}(\Delta_f)$, where Δ_f is the bandwidth in Hz ; the receiver noise N_{rx} is an additional degradation caused by components in the RF signal chain and assumed to be 5 dB.

Let now denote N_j the current receiver, trying to decode a signal from N_j, starting at time t_0 and ending at t_1. Furthermore, let $\mathbb{I} = \{k : k \neq i, k \neq j\}$ be the set of active transmitters of all other overlapping transmissions, having start time $t_{k,0}$ and end time $t_{k,1}$.

Then the interference $I_{i,j,\mathbb{I}}$ at N_j for the transmission from N_i is computed in mW as

$$I_{i\to j,\mathbb{I}} = \sum_{k\in\mathbb{I}_{i\to j}} \frac{\min\left(t_1, t_{k,1}\right) - \max\left(t_0, t_{k,0}\right)}{t_1 - t_0} \cdot \mathrm{P}\left(N_k, N_j\right). \quad (3)$$

This calculation averages each interfering signal over the transmission time; hence, the effect of strong but short interference peaks are underestimated. This simplification is tolerable when using the capacity calculation methods described below, as interference will be aligned optimally.

The final quality of the received signal is measured by the Signal to Interference plus Noise Ratio (SINR):

$$\mathrm{SINR}_{i\to j} = \frac{\mathrm{P}\left(N_i, N_j\right)[\mathrm{mW}]}{I_{i\to j,\mathbb{I}}[\mathrm{mW}] + \left(N_{th} + N_{rx}\right)[\mathrm{mW}]}. \quad (4)$$

2) Packet Error Rate: The resulting Packet Error Rate (PER) of a received frame, which corresponds to the probability of a data burst with faulty Cyclic Redundancy Check (CRC), depends on three parameters: the MCS which is selected by the transmitter, the frame size, and the SINR measured at the receiver.

Based on the SINR and the modulation scheme, the pre-decoder Bit Error Rate (BER) can be derived analytically for the modulation schemes defined in IEEE 802.11 as shown in [17].

IEEE 802.11 specifies two channel coding schemes, namely Binary Convolutional Code (BCC) and Low-Density Parity-check Code (LDPC). In this paper, we restrict ourselves to the BCC scheme. Hence, the results from [18] can be applied, allowing for estimating an upper bound for the PER dependent on the SINR and the packet length.

3) Link Throughput: For an error-free link, the link gross throughput using MCS m is given by the number of data bits per symbol, n^m_{DBPS}, divided by the duration of one symbol:

$$T^m = n^m_{DBPS}/t_{symbol} \tag{5}$$

C. Modelling MIMO Links

Coarse classification distinguishes two types of MIMO techniques (both part of IEEE 801.11n-2009) based on the propagation channel properties, i. e., on the structure of the spatial correlation matrix at the antenna array. In the case of high correlation of the transmitted signal beamforming can be applied, whereas in the case of low correlation diversity and multiplexing approaches apply [19]. The focus of this work is MIMO methods in the later sense, namely Spatial Multiplexing (MUX) and Spatial Diversity (DIV) schemes.

In MUX schemes, $n_{ss} > 1$ streams are transmitted simultaneously, each one using one dedicated antenna of the transmitter. In a rich scattering environment the signal of the combined streams takes different paths with none or low correlation. Hence, different signals arrive at the multiple receive antennas which can be processed to gain the different streams. Obviously, the number of data streams is limited by the number of transmit antennas, n_{tx}. Furthermore, the receiver must contain at least as many receive antennas, n_{rx}, as streams. Consequently, a MUX scheme increases the data rate at most by $\min(n_{tx}, n_{rx})$.

DIV schemes, in contrast, exploit the diversity of the multiple receptions of the same signal: The receiver with multiple antennas has multiple copies of the transmitted signal, each distorted by a different channel function. Thus, appropriate signal processing algorithms can increase the SINR of the signal by combining the different streams.

In the schemes combining MUX and DIV, more than one transmit antenna is active, but the receiver, as in DIV schemes, has more antennas than the number of spatial streams transmitted. To describe a link with n_{tx} transmit and n_{rx} receive antennas, the common notion "$n_{tx} \times n_{rx}$" will be used. If not mentioned otherwise, we will assume that either $n_{ss} = n_{tx}$ or that the transmitter deactivates $n_{tx} - n_{ss}$ antennas to transmit $n_{ss} < n_{tx}$ streams.

A detailed introduction to the history, benefits and problems of MIMO systems can be found in [20].

1) Signal Model: A $n_{tx} \times n_{rx}$ MIMO system is represented by Equation 6 where it is assumed that the total transmit power is equally divided over the n_{tx} transmit antennas:

$$\mathbf{y} = \sqrt{\frac{E_S}{n_{tx}}}\mathbf{H}\mathbf{s} + \mathbf{n}; \tag{6}$$

$\mathbf{s} \in \mathbb{C}^{n_{tx} \times 1}$ is the transmitted signal vector whose jth component represents the signal transmitted by the jth antenna. Similarly, the received signal and received noise are represented by $n_{rx} \times 1$ vectors, $\mathbf{y}$ and $\mathbf{n}$, respectively, where y_j and n_i represent the signal and noise received at the ith antenna. E_S denotes the average signal energy during the transmission. Finally, $\mathbf{H} \in \mathbb{C}^{n_{rx} \times n_{tx}}$ is the matrix representing the $n_{rx} \cdot n_{tx}$ channels between the n_{tx} transmit and n_{rx} receive antennas.

If $n_{ss} = n_{tx} = n_{rx}$ and $\mathbf{H}$ has full rank, i. e., $\mathbf{H}^{-1}$ exists, the Zero-Forcing (ZF)-receiver can extract $\mathbf{s}$ as follows:

$$\hat{\mathbf{s}} = \left(\sqrt{\frac{E_S}{n_{tx}}} \cdot \mathbf{H}\right)^{-1} \mathbf{y}. \tag{7}$$

This equation can be generalised for $n_{tx} \neq n_{rx}$ by using the Moore-Penrose pseudo-inverse matrix $\mathbf{H}^\dagger = \mathbf{H}^* / (\mathbf{H}^*\mathbf{H})$ instead of $\mathbf{H}^{-1}$, where $\mathbf{H}^*$ is the conjugate transpose of $\mathbf{H}$.

Under ideal circumstances, one may increase the data rate of the system by merely adding transmit and receiver antennas. Under realistic conditions, there is non-neglectable correlation between the transmit and receive antennas: In the extreme case, the channel $\mathbf{H}$ is equal to $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, which resembles a completely correlated channel. In this case, the matrix is singular and cannot be inverted by the receiver; hence, the reception fails, independently of the SINR.

In practice, the MIMO channel does not fall completely in either of the theoretical cases described. The antenna correlation and the matrix rank are influenced by many different parameters, as the antenna spacing, antenna height, the presence and position of local and remote scatterer, the degree of LOS and more.

Using a widely accepted channel model, the MIMO channel with correlated antennas can be described by the matrix product

$$\mathbf{H} = {\mathbf{R_{rx}}}^{1/2} \mathbf{H_0} {\mathbf{R_{tx}}}^{1/2}, \tag{8}$$

where $\mathbf{H_0}$ represents the i. i. d. block fading complex Gaussian channel according to [21] and $\mathbf{R_{rx}}$ and $\mathbf{R_{tx}}$ are the long-term stable normalised receive and transmit correlation matrices.

Link Type	Angle Spread (rad)	
	APs	STAs
LOS	0.2766	0.9815
NLOS	0.4486	1.2075
OtoI	0.3104	1.004

TABLE I: Parameters for the UMi angle of arrival/departure spread

Under the assumption of a uniform linear array at both the transmitter and the receiver with identical unipolarised antenna elements and the antenna spacing Δ_T and Δ_R, respectively, the correlation matrices are given by [22]:

$$\mathbf{R}_{\mathbf{rx}i,j} = \rho\left((j-i)\Delta_R, \theta_R, \sigma_R\right) \quad (9)$$

$$\mathbf{R}_{\mathbf{tx}i,j} = \rho\left((j-i)\Delta_T, \theta_T, \sigma_T\right), \quad (10)$$

where

- $\rho\left(s\Delta, \theta, \sigma_\theta\right)$ defines the fading correlation between two antenna elements having distance $s\Delta$,
- θ_T and θ_R denote the mean Angle of Departure (AoD) at the transmit array and the mean Angle of Arrival (AoA) at the receive array, respectively, and
- σ_T and σ_R is the mean AoD spread and mean AoA spread, respectively.

A Gaussian angular distribution is used in [16], implying that $\theta \sim N(0, \sigma)$. With this assumption it is shown in [23] that

$$\rho\left(s\Delta, \theta, \sigma\right) \approx e^{-j2\pi s\Delta\cos(\theta)} e^{-1/2(2\pi s\Delta\sin(\theta)\sigma)^2}. \quad (11)$$

Essentially, this model results in a correlation function which is Gaussian with spread inversely proportional to the product of antenna spacing and angle spread. Consequently, large antenna spacing and/or large angle spread lead to a small correlation and vice versa. Support of this model is given by [24], which finds by simulation that correlation reaches a maximum with both antenna arrays inline, i. e., $\theta_T = \theta_R = 0$

While the mean AoA and AoD can be derived from the receiver and transmitter positions, respectively, the spread depends on the environment. The UMi model, used for the pathloss and shadowing, also defines values for these, given in Table I.

The model differentiates between nodes close to the ground (STAs), where many close scatterer and therefore a large angle spread can be expected, and higher-elevation nodes with less scatterer and a smaller angle.

2) Post-processing per-stream SINR: To integrate the impact of the MIMO channel model into the PER calculation from Section III-B2, we extend the model from [19] with the help of the results from [25], [26] to incorporate a correlated channel.

For this, we reconsider Equation 7 including the pseudo-inverse $\mathbf{H}^\dagger$: The ZF-receiver multiplies the received signal $\mathbf{y}$ with the matrix

$$\mathbf{G}_{ZF} = \sqrt{\frac{n_{tx}}{E_S}}\mathbf{H}^\dagger, \quad (12)$$

The error vector $\mathbf{e}$ of the processed symbol stream is given by $\sqrt{\frac{n_{tx}}{E_S}}\mathbf{H}^\dagger\mathbf{n}$, resulting in a noise power on the k^{th} data stream as

$$\left[\mathrm{E}(\mathbf{e}\mathbf{e}^*)\right]_{kk} = \frac{n_{tx}N_0}{E_S}\left[\mathbf{H}^\dagger\mathbf{H}^{*\dagger}\right]_{kk}, \quad (13)$$

where $[\mathbf{X}]_{kk}$ denotes the $(k,k)^{\text{th}}$ element of the matrix $\mathbf{X}$. Hence, the post-processing SINR on the k^{th} stream is

$$\mathrm{SINR}_{\text{post},k} = \frac{E_S\left[\mathrm{E}(\mathbf{s}\mathbf{s}^*)\right]_{kk}}{n_{tx}N_0\left[\mathbf{H}^\dagger\mathbf{H}^{*\dagger}\right]_{kk}} \quad (14)$$

$$= \frac{E_S}{N_0}\frac{1}{n_{tx}}\frac{1}{\left[\mathbf{H}^*\mathbf{H}\right]^{-1}_{k,k}}. \quad (15)$$

As visible in the equation, post-processing SINR on each stream is a combination of three factors:

1) The pre-processing SINR.
2) A reduction by n_{tx}, because the transmitter has to split its transmission energy among the n_{tx} streams.
3) A post-processing MIMO loss of $[\mathbf{H}^*\mathbf{H}]^{-1}_{k,k}$.

[25] proves that the post-processing MIMO gain on each stream follows a Chi-squared distribution with $2\left(n_{rx} - n_{tx} + 1\right)$ degrees of freedom. From this fact, it is derived that the mean gain on each stream without transmit- and receive correlation is $10\log_{10}\left(n_{rx} - n_{tx} + 1\right)$ dB.

Furthermore, [25] shows that transmit correlation causes a degradation in effective SINR that can be described by

$$K_T = 10\log_{10}\left(\left[\mathbf{R}_{tx}^{-1}\right]_{k,k}\right) \quad (16)$$

on the kth stream.

[26] calculates the impact of the receive correlation as

$$K_R = 10\log_{10}\left(\left(\frac{\mathrm{tr}_{n_{tx}-1}(\lambda(\mathbf{R}_{\mathbf{rx}}))}{\binom{n_{rx}}{n_{tx}-1}\det(\mathbf{R}_{\mathbf{rx}})}\right)^{-1/(n_{rx}-n_{tx}+1)}\right), \quad (17)$$

with

- $\mathrm{tr}_l()$ the l^{th} elementary symmetric function defined as

$$\mathrm{tr}_l(\mathbf{X}) = \sum_{\{\underline{\alpha}\}}\prod_{i=1}^{l}\lambda_{x,\alpha_i} \quad (18)$$

for a positive-definite $\mathbf{X} \in \mathbb{C}^{n\times n}$, where the sum is over all ordered sequences $\underline{\alpha} = \{\alpha_1, \ldots, \alpha_l\} \subseteq \{1, \ldots, n\}$ and $\lambda_{x,i}$ denotes the i^{th} eigenvalue of $\mathbf{X}$.

- $\lambda()$ the diagonal matrix containing the eigenvalues of the matrix argument.

To visualise the impact of the antenna correlation on the post-processing SINR, the exemplary scenario in Figure 1a is used: A receiving node is positioned in a half-circle around a transmitting node; the orientation of the node remains constant, i. e., with an angle $\alpha_{rx} = 0$ to the x-axis. With different positions, the AoD and AoA varies and so do the correlation matrices. Assuming a pre-processing SINR of 30 dB, antenna spacing of 0.5 wave-lengths and angle spreads as given in Table I (LOS), the mean post-processing SINR is given by Figure 1b. As expected, a 1×1 configuration is independent of the receiver position. All other configurations result in a post-processing SINR decrease so that the 30 dBm is not reached any more; the upper bound is given by a system without correlation, i. e., $K_T = K_R = 0$. Correlation is high if the antennas face each other or if they are parallel.

3) MIMO Link Throughput: With the help of the presented MIMO model it is now possible to compute, for given node's positions and pre-processing SINR the post-processing SINR per stream and thus the per-stream BER and PER.

As in Section III-B3, a relation SINR vs. gross throughput can be derived if AoD and AoA are given. Figure 2 omits for more clarity the different MCS but shows only the enclosing hull that can be achieved with a given MIMO antenna configuration and two nodes that face each other with parallel antenna orientation.

The graph for the 1×1 case, starting at 5 dB and levelling off at 25 dB/65 Mb/s presents the basic case that would also be possible using the legacy IEEE 802.11-2007 (plus the new 64-QAM 5/6 MCS). Adding more antennas allows to receive the signal at lower SINR levels (using DIV) and increasing the throughput (using MUX), although the full throughput gain can only be reached at very high SINR, i. e., above 33 dB for the 4×4 case.

IV. Single BSS Operation

In this chapter it is assumed that only one AP exists that uses the carrier frequency f_c. Thus, no interference from other APs or STAs that do not belong to the AP's BSS exist, the SINR is simplified to the Signal to Noise Ratio (SNR).

For wireless Internet access, this theoretical case only exists if (a) the coverage area of the AP is as large as the service area and (b) the carrier frequency is licensed to the provider. While these conditions represent only a theoretical case, performance metrics of a single BSS are important for the subsequent multi-BSS evaluation, because a single BSS is the trivial upper bound for the capacity: any deployment of multiple APs will increase the interference and/or the number of orthogonal channels and thus the used bandwidth.

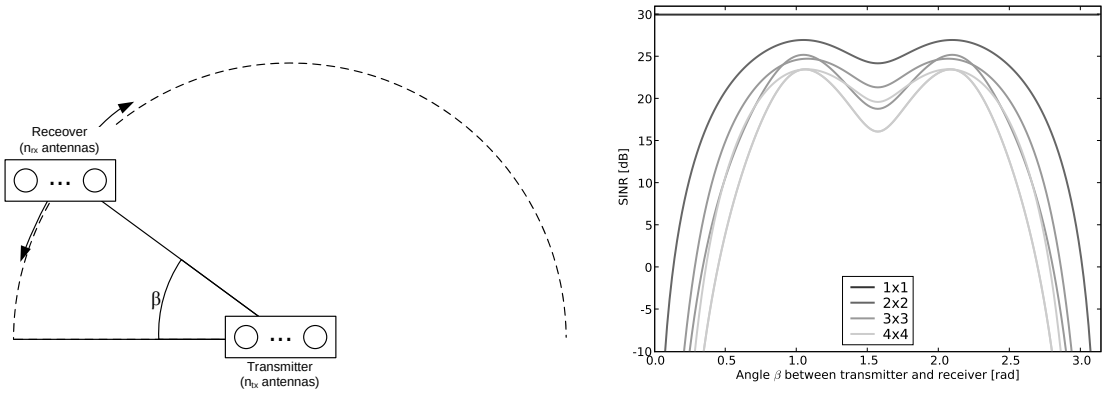
The BSS capacity is given by the maximum throughput that can be achieved in a BSS under the assumption that all STAs always have data to transmit to the APs and vice versa. The capacity depends on the link capacities of the STAs in the BSS and thus on their positions. As this differs from scenario to scenario, the evaluation assumes that STAs are positioned with a uniform random distribution over the BSS area. Thus, the capacity of the BSS becomes a random variable $\mathcal{C}$ with Probability Density Function (PDF) $p_{\mathcal{C}}$. We evaluate this capacity using a Monte-Carlo approach: In one scenario, multiple STAs are dropped randomly; their capacity is calculated using the equations for the wireless channel model, the post-processing SNR and the throughput from Figure 2. This is repeated for multiple scenarios which differ by the placement of STAs and the stochastic shadowing fading.

Figure 3a shows the Cumulative Distribution Function (CDF) of BSS capacity that is generated using the Monte-Carlo method for a 1x1 to 4x4 antenna configuration. The maximum distance of a STA to the AP is chosen such that the area covered equals to the mean BSS area used in the following multi-AP evaluation, namely $0.049\,\mathrm{km}^2$.

In the 1x1 case, the expected capacity is 37.5 Mb/s, with a probability of roughly 40% that the highest MCS with 65 Mb/s is reached. With every antenna added, one more stream can be transmitted under optimal SINR condition; the maximum capacity increases accordingly up to 260 Mb/s. However, the probability that the required SINR can be reached decreases as more streams result in a lower post-processing SINR. In the end, the probability of 260 Mb/s is only 22%. Consequently, the expected capacity scales not linearly with the number of antennas, but only by a factor of 1.71, 2.26 and 2.78 for the 2x2, 3x3 and 4x4-case, respectively.

Figures 3b and 3c show the effect of the transmit and receive antenna correlation on the capacity: If both the transmit and receive angle spread is π, antenna correlation is minimal; the only factor reducing the post-SINR is the transmit power reduction to keep the total emitted power. Hence, only the minimum MIMO loss of $10 \log_{10} \left(\left(n_{rx} - n_{tx} + 1 \right) / n_{tx} \right)$ dB is incorporated. Accordingly, as visible in Figure 3b, the probability to reach the maximum capacity increases to 40% for all antenna configurations. This scales the expected capacity by 1.85, 2.69 and 3.53 for the multi-antenna cases in comparison to the 1x1 case. A further capacity increase would be possible by using highly sensitive MCSs like 256-Quadrature Amplitude Modulation (QAM) (not part of IEEE 802.11n) in the high-SINR regions.

Figure 3c shows the expected capacity if the angle



(a) Exemplary scenario with varying SINR loss due to antenna correlation.

(b) Mean post-processing SINR per stream with varying MIMO configuration.

Fig. 1: Effect of antenna correlation on the post-processing SINR.

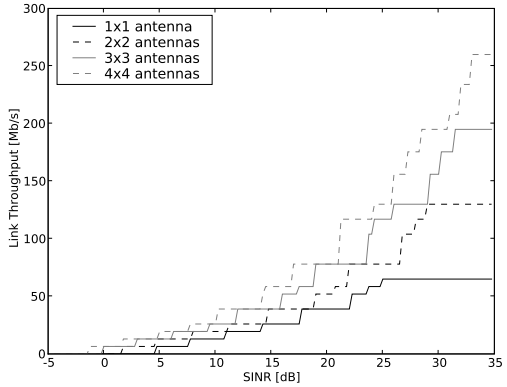


Fig. 2: Link throughput vs. SINR; assumed PSDU length is 1000 B and maximum PER 0.01.

spread is lower than in the UMi scenario: For demonstration, the angle spread from the Suburban Macro (SMa) scenario is taken (0.105 rad for the AP, 0.527 rad for STAs). As a result, the CDFs of the multi-antenna converge to the CDF of the 1x1-case. Consequently, the expected capacity increases only by 1.55, 1.81 and 2.01: Introducing complex MIMO transceivers in this scenario does not result in significant capacity gains.

V. Multi BSS Operation

The major difference between a single BSS and a WMN is the addition of interference from concurrent transmissions. In a single BSS, a concurrent transmission can only be initiated by two STAs or a AP and a STA. In the first case the AP receives two transmissions at the same time; hence, at most one signal can be decoded if transmitted using a robust MCS. In the second case, the AP is busy transmitting, there is only the chance that the downlink transmission to the STA can be received if a robust MCS is used by the AP. Both cases assume that the interference by the other STA is low, so that at least one transmission will fail and the other requires a long time due to the robust MCS. Hence, to optimise capacity, concurrent transmissions are avoided in a single BSS.

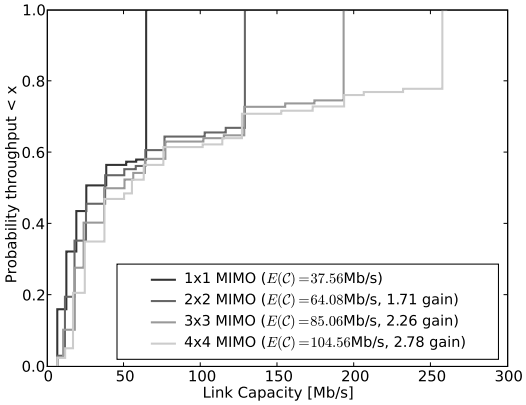
In a WMN this conclusion is not valid, because links exist that are separated from each other such that a concurrent transmission (using more robust MCS if necessary) should be preferred to a sequential operation. This is already demonstrated by the simple network shown in Figure 4, comprising four nodes and two links. It is assumed that the two links, named "1" and "2", have a maximum throughput of 30 / 22 Mb/s, if no interference is present. If both links transmit concurrently, the maximum throughput is lowered to 20 / 15 Mb/s.

If both transmitters have 1 Mb of data to transmit, it would take $1/30 + 1/22 \approx 0.079$ s until the data has reached the receivers if both links are active sequentially. Using an optimised mix of concurrent and sequential transmissions, both links transmit concurrently first; after link 1 has completed the transmission, 1-15/20 Mb are left to be transmitted at link 2, which continues at 22 Mb/s. In total, the data reaches the receivers after

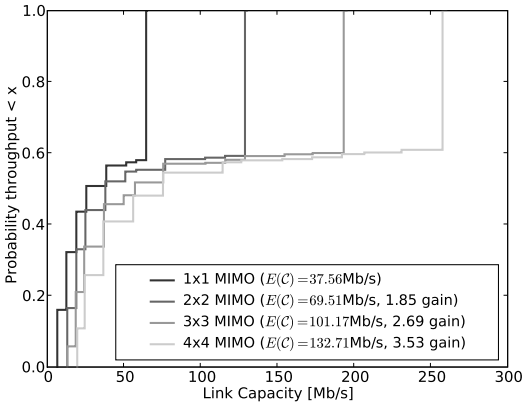
$$\frac{1}{20} + \frac{1 - \frac{15}{20}}{22} \approx 0.061\,\text{s}, \tag{19}$$

thus the required duration is shortened by 22%.

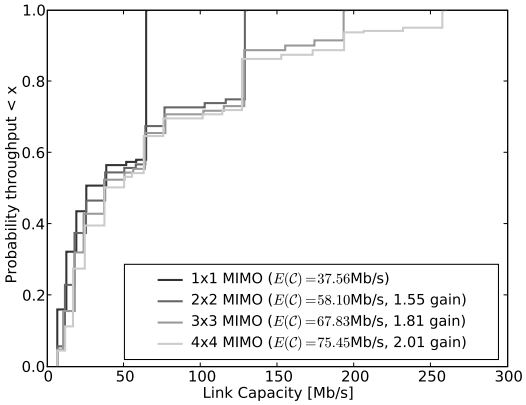
Hence, the strategy deciding which links are active at what time instance is important to determine the achievable capacity of the network. In this example the calculation of the optimal distribution between sequential and concurrent transmissions is simple because only two concurrent links exist. Every link that is added to



(a) IMT-A UMi BSS



(b) IMT-A UMi BSS without MIMO antenna correlation



(c) IMT-A UMi BSS with high MIMO antenna correlation

Fig. 3: CDF of the capacity in a BSS.

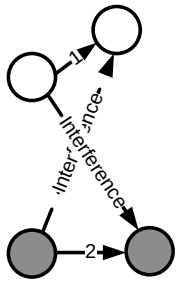


Fig. 4: Example network to demonstrate the effect of concurrent transmissions.

the network potentially doubles the number of possible combinations of active links, making the capacity calculation hard for any larger network.

A. Capacity Limits

The system model considers the restrictions and opportunities a node is constrained by and able to exploit, respectively. To find the capacity of the WMN, we apply an optimal scheduler that is able to plan concurrent transmissions optimally.

Time is assumed to be divided into fixed scheduling intervals of duration I. During one interval, a node i generates a load of l_{ij} directed to node j. This is expressed by the traffic requirement T which defines source, destination and the load. For example, the traffic requirement

$$T = \left\{ \begin{array}{ccc} \text{(source)} & \text{(rate)} & \text{(destination)} \\ N_2 & \rightarrow^{1\,\text{Mb}} & N_1 \\ N_6 & \rightarrow^{1\,\text{Mb}} & N_1 \\ N_1 & \rightarrow^{1\,\text{Mb}} & N_7) \end{array} \right\}.$$

specifies in three rows the loads from N_2 to N_1, N_6 to N_1 and N_1 to N_7, with 1 Mb each.

The task of the scheduler is to generate the sequence of transmissions such that this load is transported. This sequence is represented as the *schedule* $\left((\mathrm{S}_1, \delta_1)\,; (\mathrm{S}_2, \delta_2)\,; \ldots ; \left(\mathrm{S}_{|\mathcal{S}|}, \delta_{|\mathcal{S}|}\right)\right)$ of network states S_i and durations δ_i, with $0 \leq \delta_i \leq 1$ and $\mathcal{S}$ the set that contains all network states.

Each *network state* represents a possible combination of active links, given by transmitter, receiver, rate, source and destination of the packet flow. An example state would be

$$\mathrm{S} = \left\{ \begin{array}{ccccc} \text{(source)} & \text{(tx)} & \text{(rate)} & \text{(rx)} & \text{(destination)} \\ (N_2 \rightsquigarrow) & N_2 & \rightarrow^{54\,\text{Mb/s}} & N_3 & (\rightsquigarrow N_1) \\ (N_6 \rightsquigarrow) & N_5 & \rightarrow^{12\,\text{Mb/s}} & N_4 & (\rightsquigarrow N_1) \\ (N_1 \rightsquigarrow) & N_1 & \rightarrow^{24\,\text{Mb/s}} & N_7 & (\rightsquigarrow N_7) \end{array} \right\}.$$

This example specifies in three rows three simultaneous transmissions, one from node N_2 to node N_3 at 54 Mb/s with data originated at node N_2 and addressed to

node N_1; another from N_5 to N_4 at 12 Mb/s originated from node N_6 and addressed to N_1, etc.

A network state is *feasible* if each transmission contained is possible according to the system model. A feasible schedule must contain feasible network states only; furthermore, it must fulfil the offered traffic requirements such that if S_i is active for δ_i, $1 \leq i \leq |\mathcal{S}|$, the requirements from T are met.

The sum $\sum_{i=1\ldots|\mathcal{S}|} \delta_i$ gives the duration of the complete schedule. If this duration is larger than the duration of the scheduling interval I, more traffic is generated than what can be transported by the schedule. A schedule is called optimal if no other feasible schedule exists that has a smaller duration; let δ_i^* denote the corresponding optimal duration for network state S_i. Then the minimum resource utilisation to carry the traffic given in T using the network states $\mathcal{S}$ is

$$\mathrm{RU}(\mathcal{S}, T) = \sum_{i=1\ldots|\mathcal{S}|} \delta_i^*. \tag{20}$$

As defined in [10], the *capacity region* $\mathcal{C}(\mathcal{S})$ of a WMN with network states $\mathcal{S}$ is the set of all load settings T for which a feasible schedule exists:

$$\mathcal{C}(\mathcal{S}) = \{T : \mathrm{RU}\,(\mathcal{S}, T) \leq I\}\,. \tag{21}$$

The convex hull of $\mathcal{C}(\mathcal{S})$, i. e., the set of all T where $\mathrm{RU}\,(\mathcal{S}, T) = I$, describes the capacity limits of the WMN under any possible partitioning of resources among the (source-destination)-pairs in the network.

Therefore, the dimension of the capacity region is the number of STAs, n_{STA}, in the WMN, as each can have a different load. To reduce the number of dimensions, we compute only the *uniform system capacity* $\mathcal{C}^u(\mathcal{S})$, defined as the point of the capacity where all n_{STA} STAs have the same load l:

$$\mathcal{C}^u(\mathcal{S}) = \frac{n_{STA} \cdot l}{\mathrm{RU}\,(\mathcal{S}, T)}. \tag{22}$$

The calculation of the optimal schedule is performed $\mathcal{C}^u(\mathcal{S})$ in two steps: In step one, the set $\mathcal{S}$ of all feasible network states is computed. The second step converts T and each network state into a matrix such that the optimisation problem of finding the optimal schedule becomes an instance of Linear Programming (LP):

$$\begin{aligned} &\text{minimise } f(\delta) = \sum_{i=1\ldots|\mathcal{S}|} \delta_i \\ &\text{such that } \sum_{i=1\ldots|\mathcal{S}|} \delta_i \cdot \mathbf{s}_i = T \\ &\qquad 0 \leq \delta_i \leq 1 \qquad i = 1 \ldots |\mathcal{S}|. \end{aligned} \tag{23}$$

The complexity of both parts of the algorithm, namely the creation of the network states and the solving of the LP instance, depends on the number of network states $\mathcal{S}$ to be considered. As shown in [10], this number is expected to increase exponentially with the number of nodes, which limits the applicability to networks having less than 30 nodes.

[13] proposes heuristics to optimise the generation of the network states such that the upper bound capacity is closely approximated. Thereby, up to 150 nodes become feasible.

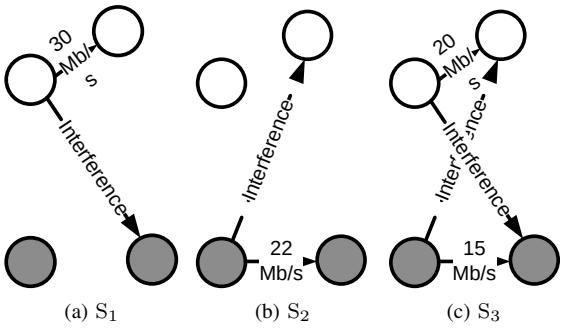


Fig. 5: The network states of the network in Figure 4.

B. Example

The example network in Figure 4 is used to illustrate network states and the resulting capacity region.

The network has two links and three network states, depicted in Figure 5: link 1 active, link 2 active, or both links active. In this network, the computation of the capacity region $\mathcal{C}(\mathcal{S})$ given in Figure 6 is simple owing to the small number of network states:

- The intersections with the x- and y-axis are given by schedules where only S_1 respectively S_2 is active.
- The throughput of a schedule where only S_3 is active cannot be achieved by any linear combination of S_1 and S_2; hence, the point $(20/15)$ is part of the hull.
- The remaining hull is a linear combination of either S_1 with S_3 if link 1 needs to transmit more than 20 Mb (and link 2 less than 15 Mb), or S_2 with S_3 in the opposite case. Any combination of S_1 and S_2 would result in lower throughput for one of the links.

The uniform system capacity $\mathcal{C}^u(\mathcal{S})$ can be found by restricting T to the points where the load of link 1 is equal to link 2. The capacity limit under this condition is given by $T = (440/27, 440/27)$, resulting in $\mathcal{C}^u(\mathcal{S}) \approx 32.6$ Mb/s.

VI. Evaluation

For the evaluation, we apply the WMN scenario creator from [27]. It is used to generate 10 scenarios of 1 km^2 each with different shadowing conditions; then, each scenario is covered with around 20 MSTAs with AP functionality so that wireless coverage and connectivity of the WMN is ensured.

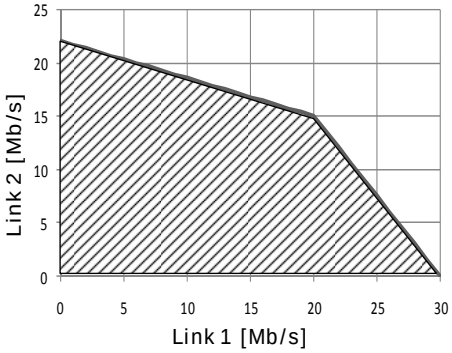


Fig. 6: Capacity region of the network in Figure 4.

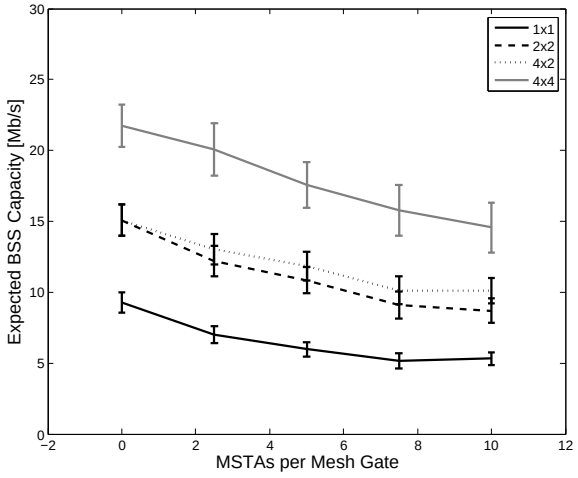
In the second step, either all, 6, 4 or 2 of the MSTAs equipped with mesh gate functions, i. e., connected to the wired backbone; this results in approximately 0, 2.5, 5 or 10 MSTAs per mesh gate, respectively. Then, costs for every link in the WMN are calculated by the maximum transmission rate on the link; this allows for creating a routing matrix using the Dijkstra all-pairs shortest path matrix.

Traffic is generated in each scenario by 100 STAs, positioned randomly and associated to the closest (in terms of pathloss) AP. Each STA requests downlink and uplink traffic from/to the Internet, divided as 90 to 10. Downlink traffic origins at the mesh gate closest to the STA (in terms of path cost), uplink traffic is destined to this mesh gate. By combining the randomly generated offered traffic and the routing matrix, a traffic requirement for each link in the network can be calculated, composing the load matrix T[1].

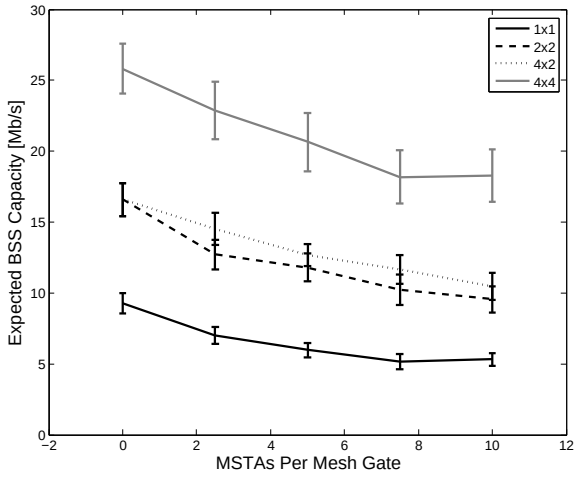
Then, the capacity calculation procedure as described above is applied in each scenario, resulting in a scenario-specific value for $\mathcal{C}^u$. Dividing this capacity by the number of MSTAs results in the mean BSS capacity of the scenario; this value allows for comparison to the values found in the single BSS case, Section IV. Finally, the scenario-specific mean BSS capacity is averaged over the 10 different scenarios, resulting in an estimation for the expected BSS capacity in a WMN. Besides this expected BSS capacity, we calculate and plot the confidence interval for a 95% confidence level of the mean capacity estimator.

Similar to Figure 3, three different settings for the MIMO model are considered: First, the default UMi values as given in Table I. Then, for comparison, the antenna correlation is minimised by setting the angle spread of the MSTAs and STAs to π. Finally, the

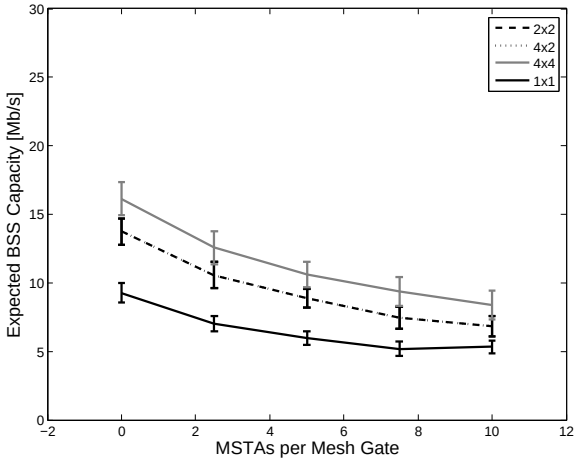
[1]Optionally, T can only contain the end-to-end loads and the optimal routes through the WMN are found automatically during the schedule minimisation. However, a STA might be associated to multiple APs and distribute its traffic over multiple routes, then. As routing is not the scope of this work, the routes are pre-calculated as described.



(a) UMi MIMO settings.



(b) No MIMO antenna correlation.



(c) High MIMO antenna correlation.

Fig. 7: Expected BSS capacity of the WMN. In axb, a is for the number of antennas of MSTAs, b for STAs

angle spread values from the IMT-A SMa are used to demonstrate a scenario setting with high correlation and low MIMO performance.

The baseline antenna configuration is a "traditional" Single Input/Single Output (SISO) equipment: All devices – MSTAs and STAs – only have a single antenna. Of course, the BSS capacity for the baseline configuration is independent of the MIMO model parameter settings, as they only impact the performance in the multi-antenna case. Nevertheless, the results from the baseline configuration are given in all figures for comparison.

The other MIMO configurations assume either 2 or 4 antennas at all devices; Additionally, a "4x2" case is given where all MSTAs are configured with 4 antennas, whereas the STAs have 2 antennas only.

Figure 7 presents the expected BSS capacity for the different MIMO configurations and model parameter settings.

Clearly, the interference from the neighbouring BSSs reduces the BSS capacity significantly below the expected capacity of a single BSS given in Figure 3. In case of a network without MSTAs, i. e., a traditional multi-AP deployment, the BSS capacity decreases by a factor of four. Interestingly, this decrease is independent of the antenna configuration and the MIMO model setting. Consequently, the capacity increase of MIMO in the single BSS is completely translated into a capacity increase in the multi BSS network, i. e., the same non-linear increase as found in Section IV is also present. For example, using the UMi MIMO model parameters, the expected capacity increases compared to the 1x1 case by a factor of 1.6 and 2.4 for the 2x2 and 4x4 configuration, respectively.

As expected, the introduction of MSTAs reduces BSS capacity. The more MSTAs per mesh gateway are deployed, the higher the mean number of hops in the WMN, which cannot be countered completely by an increase of concurrent transmissions. However, the theoretical link capacity gain of MIMO is better approached by the higher the number of MSTAs.

For the 5 MSTAs per mesh gate deployment and the UMi MIMO model setting, the expected capacity increase factors are 1.9 (2x2) and 2.9 (4x4) when compared to the 1x1 case. Without antenna correlation, these factors become 2.0 and 3.5, respectively.

The reason for a higher MIMO gain in the WMN deployment is found in the capacity distribution of links in the backbone of the WMN: According to the node placement algorithm from [27], "good" positions for MSTAs are preferred, leading to a high SINR between adjacent MSTAs in the WMN. Hence, a 1x1 link is improved more than what can be expected from the average improvement as calculated in Section IV. This improvement of few links is visible in the final results because the capacity of the WMN backbone limits the capacity of the whole WMN; hence, an improvement of few, but important links leads to an improvement of the complete network capacity.

VII. Conclusion

Both WMN and MIMO are interesting research fields on their own. In this paper, we show that the combination of both gains valuable insights: WMNs benefit significantly from the capacity increase of MIMO.

The results, based on the capacity calculation framework, represent upper bound capacities. It is not clear what remains of this capacity if a real MAC protocol, using an imperfect (distributed) scheduler, is applied: Introducing MIMO to WMNs increases the chance of the rate adaptation algorithms to apply different MCSs; consequently, more errors can be made by the scheduler, resulting in uncoordinated concurrent transmissions. Consequently, distributed scheduling of MIMO-enhanced WMNs appears to be a promising research area.

Acknowledgements

The work leading to these results has been partly funded by the European Community's Seventh Framework Program FP7/2007-2013 under grant agreement no. 213311 also referred as OMEGA project.

References

[1] S. Max and T. Wang, "Transmit power control in wireless mesh networks considered harmful," in *Second International Conference on Advances in Mesh Networks, 2009 (MESH 2009)*, (Athens, Greece), pp. 73–78, Jun 2009.

[2] IEEE, "IEEE 802.11n-2009: Standard for Information technology - Telecommunications and information exchange between systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) & Physical Layer specifications - Enhancements for Higher Throughput," Amedment 802.11n, New York, Sept. 2009.

[3] IEEE, "IEEE 802.11s/D5.0: Draft Standard for Information technology - Telecommunications and information exchange between systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) & Physical Layer specifications - Amendment 10: Mesh Networking," Amedment 802.11s, New York, Apr. 2010.

[4] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[5] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. of the IEEE Conference on Computer Communications (INFOCOM)*, Mar. 2003.

[6] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proc. of the IEEE Conference on Computer Communications (INFOCOM)*, pp. 1360–1369, Apr. 2001.

[7] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Netwoks*, vol. 11, pp. 471–487, July 2005.

[8] V. S. A. Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan, "Algorithmic aspects of capacity in wireless networks," in *Proc. of the ACM SIGMETRICS Conference*, (Banff, Canada), pp. 133–144, June 2005.

[9] E. Arikan, "Some complexity results about packet radio networks," *IEEE Transactions on Information Theory*, vol. 30, pp. 681–685, July 1984.

[10] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 736–748, July 2003.

[11] P. Bjorklund, P. Varbrand, and D. Yuan, "Resource optimization of spatial tdma in ad hoc radio networks: a column generation approach," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, pp. 818–824 vol.2, March-3 April 2003.

[12] S. Max, G. R. Hiertz, E. Weiss, D. Denteneer, and B. H. Walke, "Spectrum sharing in ieee 802.11s wireless mesh networks," *Computer Networks*, vol. 51, pp. 2353–2367, June 2007.

[13] S. Max, E. Weiss, G. Hiertz, and B. Walke, "Capacity bounds of deployment concepts for wireless mesh networks," *Performance Evaluation*, vol. 66, pp. 272–286, Mar 2009.

[14] S. Max, L. Stibor, G. Hiertz, and D. Denteneer, "On the performance of hybrid wireless/wired mesh networks," in *Proceedings of the 3rd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications WiMob 2007*, (White Plains, New York, USA), p. 8, IEEE Computer Society, Oct 2007.

[15] S. Max, E. Weiss, and G. Hiertz, "Analysis of wimedia-based uwb mesh networks," in *In Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN) 2007*, (Dublin, Ireland), pp. 919–926, IEEE Computer Society, Oct 2007.

[16] ITU, "Rep. ITU-R M.2135, Guidelines for evaluation of radio interface technologies for IMT-Advanced," report, ITU, 2008.

[17] J. G. Proakis, *Digital Communications*. Mcgraw-Hill Publ.Comp., 4. a. ed., Aug. 2000.

[18] M. Pursley and D. Taipale, "Error probabilities for Spread-Spectrum packet radio with convolutional codes and viterbi decoding," *Communications, IEEE Transactions on*, vol. 35, no. 1, pp. 1–12, 1987.

[19] J. Mirkovic, G. Orfanos, and H. Reumerman, "MIMO link modeling for system level simulations," in *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, pp. 1–6, 2006.

[20] D. Gesbert, M. Shafi, D. shan Shiu, P. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 3, pp. 281–302, 2003.

[21] J. Heath, R.W. and A. Paulraj, "Linear dispersion codes for mimo systems based on frame theory," *Signal Processing, IEEE Transactions on*, vol. 50, pp. 2429 – 2441, oct 2002.

[22] H. Bölcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 427–439, Apr. 2003. Final version 2002-11-30.

[23] D. Asztely, "On Antenna Arrays in Mobile Communication Systems," Tech. Rep. IR–S3–SB–9611, Royal Institute of Technology, Department of Signals, Sensors & Systems, Stockholm, 1996.

[24] X. Li and Z. Nie, "Effect of array orientation on performance of MIMO wireless channels," *Antennas and Wireless Propagation Letters, IEEE*, vol. 3, pp. 368–371, 2004.

[25] D. Gore, R. Heath, and A. Paulraj, "On performance of the zero forcing receiver in presence of transmit correlation," in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, p. 159, 2002.

[26] M. McKay and I. Collings, "Error performance of MIMO-BICM with Zero-Forcing receivers in Spatially-Correlated rayleigh channels," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 3, pp. 787–792, 2007.

[27] S. Max, L. Stibor, G. Hiertz, and D. Denteneer, "IEEE 802.11s mesh network deployment concepts -invited paper-," in *Proc. of European Wireless Conference 2007*, (Paris, France), Apr. 2007.

Behind-the-Scenes of IEEE 802.11a based Multi-Radio Mesh Networks: A Measurement driven Evaluation of Inter-Channel Interference

Sebastian Robitzsch, John Fitzpatrick, Seán Murphy and Liam Murphy
University College Dublin
Performance Engineering Laboratory
Dublin, Ireland
sebastian.robitzsch@ucdconnect.ie, (john.fitzpatrick, sean.murphy, liam.murphy)@ucd.ie

***Abstract*—To successfully develop IEEE 802.11a based wireless mesh network solutions that can achieve the reliability and capacities required to offer high quality triple play services the use of multiple radios in each mesh node is essential. Unfortunately, the co-location of multiple antennas in a single device leads to a number of interference problems. In this paper the impact of non-overlapping channel interference in IEEE 802.11a based multi-radio nodes is investigated. A detailed explanation of the performance decreases and their relation to radio settings is presented. The primary contribution of this paper is the discovery of a channel interference effect which is present over the entire 802.11a frequency space. This interference appears if two radios are located less than 50 cm from each other and both are attempting to operate as usual. The results were obtained by conducting experiments in a well planned testbed to produce reliable and reproducible results. The presented results incorporate multiple parameters including transmission power, modulation coding scheme, channel separation and physical layer effects such as adjacent channel interference, carrier sensing, retransmissions and packet distortion.**

***Keywords*-Wireless LAN; Measurement; 802.11a; Multi-Radio Wireless Mesh Network**

I. INTRODUCTION

In recent years Wireless Mesh Networks (WMNs) have become increasingly popular. This is primarily due to the high level of penetration achieved by Wireless Local Area Network (WLAN) as an access technology for end user devices and the widespread availability of low cost Wireless Fidelity (WiFi) hardware. Another important factor is that WiFi operates in the unlicensed Indistrial, Scientific and Medical (ISM) radio spectrum, therefore, WMNs based on this technology can be deployed without requiring the purchase of expensive spectrum licenses. Moreover, the ability of WMNs to provide last mile communication infrastructure as a number of use cases such as campus, festival or conference deployments. This flexibility has driven a number of interest groups to investigating WMN deploying issues.

As mentioned in the previous published work, [1], interference between two interfaces on the same WMN can be appear if both antennas are located relatively close to each other. Interference can still occur even if the well-known Adjacent Channel Interference (ACI) requirements, as described by Angelakis et al. [2], Nachtigall et al. [3] and Cheng et al. [4], are fulfilled by setting the radios to a channel separation larger than one.

The focus in the initial work [1] was not to provide an exact set of data to derive necessary radio and interface parameters such as Transmission Power (TxPower), Modulation Coding Scheme (MCS) and channel separation, for a network with reliable and stable links. It was rather envisaged to go one step further than [2] and [4] to provide a first survey that the channel interference phenomenon is related to the three parameters TxPower, MCS and channel separation. Even in comparison to [3], the previous work [1] showed that focusing solely on antenna separation (distance) and channel separation does not give enough detail to fully understand the radio environment. The work presented in this paper expands on this previously published work with a more detailed analysis of the evaluation of obtained Received Signal Strength (RSS) and Noise Floor (NF) values, new measurements to address the question whether data frames or their Acknowledgements (ACKs) are affected by ACI or Inter-Channel Interference (ICI) and with a more detailed evaluation and discussion of the obtained measurement results starting to quantify the different investigated interferences.

As will be described in Section III-A, a testbed environment was established including scripts to conduct a number of experiments automatically. These experiments comprised every relation of TxPower, MCS and channel separation (as long as some interference could be investigated) for antenna separations up to 60 cm in 10 cm steps.

The remainder of this paper is structured as follows. The next Section II provides an overview of related work from other groups, this is followed by Section III which gives the detailed description of the experimental setup. Section IV describes the ACI effect and its use to evaluate the testbed hardware. Section V presents experimental results which investigates the relationship of RSS and Signal to Noise Ratio (SNR), the ICI as well as the relationship of Re-Transmission Rates (RTRs), Carrier Sense (CS) and application layer losses. The paper is then concluded in Section VI which summarises the paper.

II. Related Work

In recent years the ACI effect has been well investigated by different independent researchers. However, no work appears to have been done for channel separations of more than one; this paper aims at addressing this shortcoming.

Angelakis et al. [2] qualified the effect of ACI in terms of throughput measurements within a single node equipped with multiple interfaces. In this paper the authors developed a mathematical model which showed that neighbouring 802.11a channels have a spectral overlap which produces a significant level of interference that can lead to lossy and unstable links in dual-radio equipped nodes. However, their experiments were conducted under laboratory conditions using attenuators and couplers to demonstrate the ACI effect. Due to this experimental environment the ACI effect on two transmitters with a channel separation of two could not be shown. The reason for this was based on the strength of the attenuators which they used. The level of these attenuators was too high and the transmit signal level was below the sensitivity threshold of a common WiFi card. Therefore, based on the obtained results it was not possible to conclude what level of channel separation is required to provide stable and reliable links. Even in their subsequent papers [5] and [6] this issue has not been investigated further.

Mishra et al. [7] assumed that the overlap between neighbouring channels in 802.11a is so low that it can be ignored for practical purposes. The authors conducted experiments for an 802.11b link and a channel separation of three which represents two orthogonal channel as 802.11a does. They did not observe any interference by measuring the throughput for a distance of 10 m. Additionally, Mishra et al. defined an appropriate model for partially overlapping channels which calculates the level of interference caused by all non-orthogonal channels, as they appear in 802.11b/g. However, just measuring the throughput is not sufficient to conclude that there is no interference in cases where both radios are located closer than 10 m to each other. The assumption of Mishra et al. that the only interference is due to small partially overlapping channels of 802.11a does not hold for distances below 30 cm, as will be shown later in this paper.

This can easily be verified by previously published work, such as [2] [3] [4], which investigated that ACI issues must be taken into account when conducting multi-radio measurements with Institute of Electrical and Electronics Engineers (IEEE) 802.11a hardware. One of the most important pieces of work in relation to this issue are the results of Nachtigall et al. [3]. They demonstrated that the number of available non-interfering channels depends on both the antenna separation and the Physical Layer (PHY) modulation for a dual-radio scenario. They also stated that under their conditions only one channel can be used at the same time, which is not necessarily true.

In order to verify the testbed and provide reproducible results, Burchfield et al. [8] proposed three necessary steps based on an extended set of experiments in a real environment as well as under laboratory conditions using coax cables. In summary they recommend to check first for external networks using the same Medium Access Control (MAC) protocol. Secondly, the medium should be checked also for other interference which cannot be recognized by the chipset but still senses the medium, e.g., a microwave in case of deploying a 802.11b/g testbed. Finally, Burchfield et al. recommends the use a coaxial setup to verify the system's capabilities. However, a statistical evaluation after conducting the experiment seems to be another much more reliable step in terms of proving the mean values than confirm the experimental obtained graphs by replacing the wireless links with coax cable. To verify the claim by Burchfield et al. similar coax experiments were conducted. The results showed that even if both antenna outputs were connected via coax cable with each other it was possible to see regular beacons sent from another Access Point (AP) which was located approximately 30 m away. Hence, the proposed verification by Burchfield et al. using coax cables does not seem to be 100 % accurate.

The most promising work was done by Nachtigall et al. [3] who investigated the interference among the interfaces of a multi-radio node and found that the radios located close to each other interfere with each other significantly. They even stated that under their experimental conditions, an antenna separation of 15 cm, only one channel within the entire 5.2 GHz band can be used at the same time. They concluded their work with the statement that the number of available orthogonal channels depends on the antenna separation and MCS.

III. Experimental Setup

This section provides a detailed configuration description of the experimental environment; it also provides an overview of how the logged data was processed. It also shows how the measurement process was designed in terms of duration and configuration using prior statistical evaluation, for example the method of independent replications. The use of the Fresnel formula to design the links properly is also described.

A. Testbed Configuration

The most important factor when planning the experimental testbed was to provide reliable and reproducible results with the least number of external dependencies as possible. To achieve this goal the testbed was deployed as shown in Figure 1.

All three machines (Node A, B and C) were x86 Intel based desktop machines running Ubuntu 32 bit server edition and kernel version 2.6.28. Node A was equipped with two wireless interfaces and both Node B and Node C each had

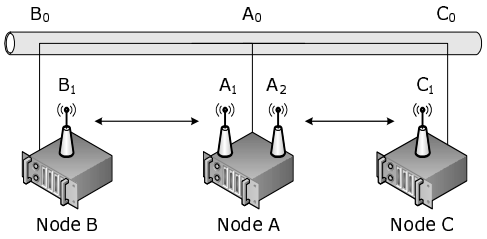


Figure 1. Experimental Setup

Table I
INTERFACE CONFIGURATION

Node	Interface	Type	IP
Node A	A_0	eth	192.168.0.1/24
	A_1	ath	10.0.2.1/24
	A_2	ath	10.0.3.1/24
Node B	B_0	eth	192.168.0.2/24
	B_1	ath	10.0.2.2/24
Node C	C_0	eth	192.168.0.3/24
	C_1	ath	10.0.3.2/24

one; each wireless interface used RouterBoard R52 wireless 802.11a/b/g mini-Peripheral Component Interconnect (PCI) cards with the Atheros chipset AR5414. The R52 card is WLAN certified by the WiFi alliance which ensures the chipsets has been designed according to the standard and produces Orthogonal Frequency-Division Multiplexing (OFDM) singals with Frame Error Rates (FERs) defined by IEEE. This mainly gives the safety that any other WiFi certified WLAN card will behave like the R52 does and will give similiar experimental results. To connect the the mini-PCI cards to the PCI slots in each machine, a RouterBoard 14 mini-PCI to PCI adapter was used. The connection between the cards and the 5.5 dBi omnidirectional antennas used 20 cm length RG-178 U.fl to N pigtails. Since the experiments being performed required that the antenna separation could be varied from 10 cm up to 60 cm, the connections from the interface cards to the antennas for interfaces A_1 and A_2 of Node A were extended using two 2 m low-loss CLF-400 coaxial cables with N connectors on both sides.

In order to obtain application layer metrics, the traffic generation and bandwidth measurement tool Iperf version 2.0.4-3 [9] was used. Metrics related to the underlying wireless technology such as RSS and NF, were obtained from the radiotab header using TShark version 1.0.4 [10] to parse the received packets. For stability and usability reasons, the Multiband Atheros Driver for Wireless Fidelity (MADWiFi) version 0.9.4 revision 4023 [11] wireless driver was favoured over the ath5k driver.

The testbed was configured as shown in Table I. In all experiments the link between $\overrightarrow{A_2C_1}$ operated on channel 36 to communicate while the link between $\overrightarrow{A_1B_1}$ was changed to achieve the variation in channel separation. The IP routing tables in the nodes were configured such that A_1 communicates exclusively with B_1 and A_2 communicates exclusively with C_1 and vice versa. Additionally, for each established interface, e.g *ath0*, a virtual monitoring interface was created on the same physical interface. This was used to obtain the radiotab header from the data packets being received.

Each machine in the testbed was also connected to a switch over wired Ethernet and configured as part of the same subnet. This setup was required since a shell script operating on the sending side of each link was responsible for the configuration of the *ath* interfaces on both the Tx and Rx sides of the link. Hence, for each pair of wireless interfaces which make up a link a shell script is used to configure the interfaces to the required settings.

Furthermore, to synchronise both scripts one of the three machines acts as a synchronisation server. Every script simply creates a text file remotely in the `/tmp` directory of the synchronisation server which indicates the current state of the remote script. This communication was done using SSH sessions over the Ethernet connection.

Synchronisation between the scripts was necessary since after reconfiguring a wireless interface the time taken for layer two connectivity to be established is variable. Therefore, before starting the Iperf client during each experimental run, the script checks that the Iperf server is reachable using a single 64 byte ping. If the server is reachable, a check is performed to verify that the remote Iperf client is also ready. This check simply reads the synchronisation file from the synchronisation server. Only if both actions are completed successfully does the script begin the experiment and measurements. Due to some inaccuracies in the time taken for both interfaces to become active ($\approx$1-2 s), the first and the last 10 IPerf samples are not considered when processing the results.

As it was required that TShark only capture data frames from the wireless interface in each node, the script automatically sets an appropriate Ethernet filter using the pcap library syntax. It is also worth noting that the shell script forces Iperf to bind on a particular IP address to ensure that the correct data is captured. This incoming packets are filtered based on the source/destination MAC addresses as well as the data type of the received packet; specifically only packets of `type data` are captured and both `type ctl` and `type mgt` are ignored. As mentioned earlier, the radiotap header is parsed to extract the required data; specifically the RSS (from the radiotap header field `dbm_antsignal`), NF (from the `dbm_antnoise` field) and the physical layer datarate (from the `datarate` field).

In all experiments performed the User Datagram Protocol (UDP) was used as the transport layer protocol. UDP was chosen over Transmission Control Protocol (TCP) and

Stream Control Transmission Protocol (SCTP) since it has no inherent congestion control mechanisms which would compensate for link degradation and hence distort the results. The UDP payload size for each experiment was set to a fixed value of 1400 octets at all times.

As recommended by Burchfield et al. [8], any research carried out in the 802.11 domain that is based on obtaining results from real deployments should be performed using their recommendations. To follow this approach, it can be confirmed that the room in which the experiments were conducted was free of any interference in the measured ISM band. This was verified in two ways; firstly by sniffing the medium for other WiFi radios. Secondly, in order to verify that there was little or no interference from non WiFi devices experiments were performed in which a single link between two 802.11 radios was set up. These experiments showed that the maximum possible UDP data rate without an RTRs could be achieved, thereby showing that there was no significant level of interference. However, the recommendation of Burchfield et al. [8] that verification of the testbed be carried out using coax cables between the radios was not conducted, as described in Section II.

B. Statistical Evaluation

In order to prove the reliability of the measured results the Confidence Interval (CI) was calculated for each sample mean using the method of independent replications, as described by Banks et al. [12]. This is shown in Equation 1 where $\bar{\mu}$ represents the sample mean, υ the Degree of Freedom (DF), α the chosen CI and $\sigma(\bar{\mu})$ the standard error or variance of the sample mean.

$$\bar{\mu} \pm t_{\frac{(1-\alpha)}{2},\upsilon} \cdot \sigma(\bar{\mu}) \tag{1}$$

As proposed by Banks et al., the Student's t-distribution was used to define υ; this is because every measured sample is independent and the common Gaussian distribution does not cover this case accurately. Since every measurement represents a set of non-normal data the number of samples used to calculate the sample mean and the corresponding CI should be at least 50 as recommended in Wang [13]. Hence, to calculate the throughput sample mean each measurement was performed over a period of 50 s with an interval of 0.5 s which gives a set of 100 values for the IPerf results. The corresponding DF of $\upsilon = 1.99$ can be derived for a probability of 5 % to exceed the critical value. Further mean values, e.g., RSS and NF, were taken directly from the radiotap header which appears per MAC frame and results in a DF of $\upsilon = 1.96$ for total number of independent measured values above 100 per sample mean and the probability of exceeding the critical value.

C. Fresnel-Zone and Free Space Path Loss

In order for the results presented in this paper to be accurate and free from external influences, it was necessary that the experimental environment had to fulfil some basic requirements. To achieve this there must be a direct Line of Sight (LOS) connection between each pair of transmitting and receiving antennas with no obstructions in the path which could cause inference due to reflections or shadowing effects. At a minimum the first Fresnel zone must be largely free from obstacles to avoid interference from reflected waves.

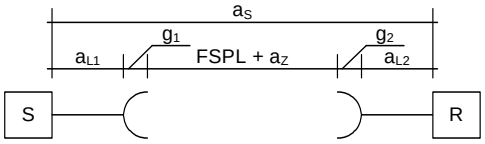


Figure 2. Loss Parameters within a System as shown in Equation 4

Equation 3 shows the simplified formula to calculate the nth Fresnel. Using this equation a radius F_n can be obtained which describes a zone that surrounds the direct LOS connection between both antennas that is completely free from obstacles, e.g., trees, hills or walls. Since in the experimental environment the only obstacle is the ground, Equation 3 just comprises the distance d between both antennas and is used to calculate the minimum antenna height such that the first Fresnel zone is clear. Therefore, the general simplified Fresnel formula is:

$$F_n = \sqrt{\frac{n \cdot \lambda \cdot d_1 \cdot d_2}{d_1 + d_2}} \tag{2}$$

$$F_n = \sqrt{\frac{n \cdot c \cdot d}{2f}} \tag{3}$$

To verify the accuracy of the results from the experimental environment the overall system loss a_S was computed analytically. A comparison between the actual attenuation experienced in the experimental set-up and the theoretical attenuation predicted was then made. To compute the theoretical attenuation that should be experienced, the following formula was used:

$$a_S = a_{L_1} - g_1 + FSPL + a_Z - g_2 + a_{L_2} \tag{4}$$

Equation 4 is made up of the cable losses from Mesh Node $(MN)_1$ to antenna a_{L_1} and from MN_2 to antenna a_{L_2}, the gain of both antennas g_1 and g_2, the free space path loss described as

$$FSPL = 92.4 + 20\log(d) + 20\log(f) \tag{5}$$

and some additional unpredictable losses a_Z, as depicted in Figure 2. Examples for unpredictable losses are interferences due to multipath propagation especially for frequencies less than 10 GHz and losses due to atmospheric absorption.

IV. Adjacent Channel Interference in Multi-Radio Nodes

There is currently a lot of interest in both the industrial and academic research communities for using IEEE 802.11a to provide backbone links for WMNs while simultaneously using IEEE 802.11b/g to provide user access. This means that each MN may have multiple 802.11a radios operating in close proximity and hence ACI issues in 802.11a are of particular importance.

Another important factor is the relatively short spacings that can be achieved between co-located antennas. It is simply not feasible to have antenna separations larger than 50 or 60 cm due to the dimensions of the MN casings and the goal of having non-intrusive and easily deployed MNs. For example it is not possible to have antenna separations of 2 m as proposed by Worldwide Interoperability for Microwave Access (WiMAX) deployment guidelines. The cumulative effect of both ACI and small antenna separations can have a significant affect on the performance of such systems and these issues must be addressed prior to deploying an 802.11a based WMN infrastructure. For this reason, the spectrum mask of the WiFi cards used in the work was examined to verify that it meets with the IEEE guidelines [14]. Figure 3 depicts the measured spectrum of a MikroTik R52 mini-PCI WiFi card. These measurements were performed using the IEEE recommended guidelines for spectrum measurements of 802.11a systems; specifically the spectrum analyser was set with a 100kHz Resolution Bandwidth (RBW) and a 30kHz Video Bandwidth (VBW) [14].

Although, Cheng et al. [4] had previously conducted this spectrum analysis, they considered a spectrum mask that should not exceed -20 dB at 11 MHz and -30 dB at 22 MHz. Furthermore, Cheng et al. stated that they used TxPower values of 30 dBm, 36 dBm and 99 dBm which were allegedly performed using the MADWiFi driver. Since the considered Power Spectral Density (PSD) limits and the chosen TxPower values do not fit to IEEE 802.11a, 802.11b or 802.11g specifications, the work and results obtained cannot be considered accurate. For operating in the 802.11a band the IEEE recommend a 20 MHz channel spacing, a maximal bandwidth of 18 MHz at 0 dBr and offsets of at least -20 dBr at 11 MHz, -28 dBr at 20 MHz, and -40 dBr at 30 MHz. Figure 3 depicts the measured PSD of the WiFi cards used for the experiments presented in this work. To accurately compare each of the different TxPower curves, the area around the centre frequency f_c of 5.2 GHz was normalised to $RSS_N = 0$. Therefore, for every TxPower curve the sample mean $\bar{\mu}_a$ between 5.191 GHz and 5.209 GHz was calculated and then added as a fixed value to the series of measurements. After applying the mean value to the whole curve the new mean value for the range between 5.191 GHz and 5.209 GHz was consistently above zero. To adjust the normalisation process to the correct mean value that will be added to the entire curve, half of the belonging standard deviation $\sigma(\bar{\mu}_a)$ was also subtracted, as shown in Equation 6.

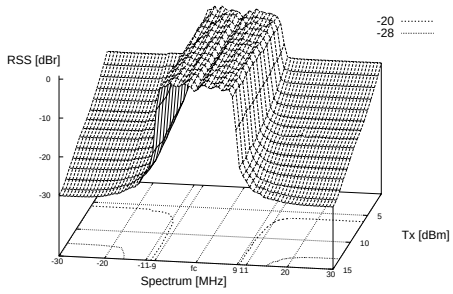


Figure 3. 3D Spectrum Analysis of a MikroTik R52 IEEE 802.11a/b/g Card for 5.2 GHz Carrier with a 20 MHz Bandwidth

$$RSS_N = x + \bar{\mu}_a - \frac{\sigma(\bar{\mu}_a)}{2} \quad (6)$$

Due to the small SNR values of the lower TxPower measurements, as shown in Figure 3, the PSDs for $f_c \pm [9\,MHz, 11\,MHz]$ could not be obtained as their values were already equal to or less than the noise floor. However, based on the obtained results shown in Figure 3 it is reasonable to say that the MikroTik R52 WiFi cards meets with the requirements of the IEEE guidelines.

V. Results and Discussion

This section presents and describes results obtained from experiments performed on the experimental testbed described earlier.

A. Received Signal Strength and Noise Floor Level Measurements

In 802.11 the term Received Signal Strength Indicator (RSSI) is used as a generic unitless signal strength metric and is the only value used to describe the signal strength of a received packet. The RSSI is an arbitrary indication of the actual RSS with a range that is defined by each vendor individually depending on the level of granularity required. The lowest possible value is 0 and goes up to the highest arbitrary value (e.g. 70 for MADWiFi).

An RSSI of 0 always represents the NF inside the radio. To calculate the actual RSSI value MADWiFi takes the RSS value from the card and adds 96 dB (NF) to it. This corresponds to a maximal RSS of -26 dBm for MADWiFi since this value is equal to an RSSI of 70. In order to easily compare the RSS and CS threshold, and to be able to make accurate assumptions about the card and driver behaviour related to the signal strength, it is essential to use the RSS instead of the RSSI. However, caution must be taken when focusing solely on the RSS, as explained below.

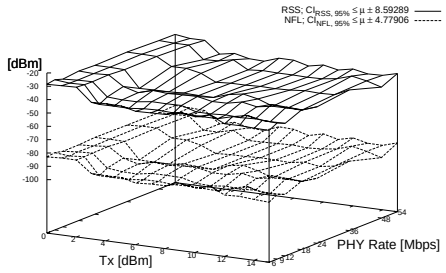


Figure 4. Received Signal Strength and Noise Floor for a Radio With Constant TxPower and an Additional Interferer

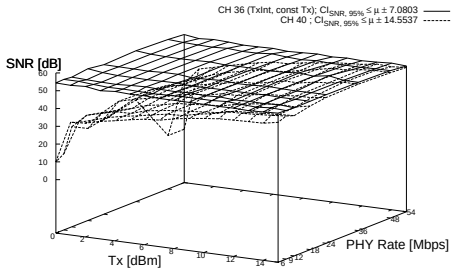


Figure 5. Derived Signal to Noise Ratio Values from Figure 4

Each WiFi chipset has an internal adaptation mechanism which adapts the noise floor level in response to the level of external interference. Specifically, the higher the level of interference, the lower the reported NF. It should be noted that this relationship is not linear and appears to use predefined thresholds which are set inside the chipset and are used to adapt the NF. This effect is depicted in Figure 4. It shows two radios placed at a distance of 10 cm from each other. The TxPower of the first radio is fixed while the TxPower of the second is incremented from 0 to 15 dBm. In Figure 4 the x axis shows the TxPower of A_1 as it increases, whereas the y axis depicts the RSS (upper surface with constant lines) measured by interface A_2 with a fixed TxPower as well as its NF (lower surface with dotted lines). From this Figure it is clear that both surfaces are equal. In particular, when the neighbouring interface A_1 increased its power from 3 dBm to 4 dBm the measured RSS and NF at radio A_2 in C_1 decreased by 12 dB.

Since the TxPower of A_2 was set at a fixed value of 15 dBm for the entire duration of the experiment, the drop in RSS is quite unexpected. Indeed it was assumed that it would report the same RSS for every received packet with only a small and relatively insignificant difference in Standard Deviation (StdDev). To investigate this issue in more detail the SNR, as given in Equation 7, for both transmissions is calculated as depicted in Figure 5.

$$SNR = RSS_m - NFL_m \quad (7)$$

As in the previous Figures the constant surface describes the radio interface sending with a fixed TxPower and the dotted line represents the radio which incrementally increased its TxPower. It is clear from Figure 5 that the SNR of the connection $\overrightarrow{A_2\,C_1}$ is constant over the entire measurement space. As expected the SNR surface of the connection $\overrightarrow{A_1\,B_1}$ increases incrementally as the TxPower of the sending interface A_1 is increased.

The relationship between the measured RSS and NF to the actual RSS can be described as:

$$RSS = RSS_m - |NFL_{def} - NFL_m| \quad (8)$$

where RSS_m is the RSS reported by the card, NFL_m is the NF defined by the vendor (i.e. -96 dBm) and NFL_m is the actual reported NF.

B. Conducted Measurements

As already shown in the previously published work [1], there are interference issues between neighbouring antennas beyond a channel separation of one. This was one of the main contributions of Robitzsch et al. and due to the importance of the problem this has been investigated further in this work. The presented results within this paper however just incorporates TxPower settings. Since the distance between every transmitter and receiver were fixed, both values, TxPower and RSS, are interchangeable. Additionally, throughput measurements are known to give a good indication of how well a link operates. However, they do not tell the whole story if they are evaluated without any further measured parameter such as RTR or application layer loss.

To discover the reason the maximum theoretical throughput cannot reached in some cases other metrics must also be considered to investigate possible sources of interference. In particular, the parameters RTR, MAC layer loss and application layer loss give a more detailed view of why the maximum throughput was not achieved. The importance of the additional metrics will become clearer after this section when a detailed description of the obtained graphs is provided. For instance, if the throughput of a connection is always 60 % lower than expected the question is whether this is due to RTRs caused by interference or whether the card backed-off a large number of times. Additionally, the relation between the MCS to the inlevel of interference gives a strong indication as well whether this MCS is suitable under these conditions and provides enough robustness.

Figure 6 to Figure 13 show a complete set of throughput measurements for an antenna separation of 20 cm including the corresponding RTRs which gives an indication of the weak link performance. As can be seen in Figure 7, radio A_2 was transmitting with a constant TxPower of 15 dBm on channel 36, whereas the TxPower of interface A_1 was

increased incrementally from 0 dBm up to 15 dBm on channel 40. To make a comparison between the different results easier, the surface of interface A_2 (fixed power surface and channel 36) was mapped to the corresponding TxPower settings of A_1. For instance, if A_1 (channel 40) was operating with TxPower 5 the corresponding measured throughput value for A_2 was mapped to the same TxPower on the x-axis, however, A_2 was still sending with 15 dBm.

From Figure 6 it can be seen that once A_1 has reached a TxPower of 2 dBm both radios achieve the same throughput. This could indicate that both radios are fairly sharing the medium as if only a single channel is being used; this occurs because the side-band from each causes the carrier sensing mechanism in the neighbouring interface to detect the channel as busy due to ACI. It could also indicate that the PSD of the side-band from each is not high enough to cause CS in the other radio but rather that there is large number of RTRs due to corrupted received frames in either B_1, C_1 or in both. To find the source of this problem the aggregated throughput is depicted in Figure 14. This shows that a throughput value of 42 Mbps for a channel separation of one and a PHY rate of 54 Mbps is achieved. This value is close to the maximum possible throughput for a single channel when using UDP and the highest MAC Service Data Unit (MSDU) size.

The RTR in Figure 7 is always around 10 % which indicates that there are quite high levels of interference. This interference causes a reduction in the maximum possible throughput of approximately 3 Mbps. Additionally, it can be confirmed that in these experiments the application layer loss was always close to zero which indicates that there was a small number of expired retransmissions. Combined these results show that it is not only the card's internal CS mechanism that causes ACI in terms of less successfully transmitted packets, but also distortion of the transmitted packets from the neighbouring radio sending a packet after the CS mechanism detected the medium to be idle. This claim can be verified by the previous work of Angelakis et al. who show the impact of the sideband of a 802.11a 20 MHz wide OFDM carrier, which can cause ACI if the antennas are located quite close, e.g., 20 cm. However, whether it was the data packet or the ACK frame that was affected by the interference cannot be obtained from this set of results. This will be examined in the next Section V-C.

It was expected that by increasing the channel separation by one the ACI effect would no longer be present and that the maximum expected throughput should be achieved on each channel. However, as shown in Figure 8, this is surprisingly not the case. In comparison to the previous Figure 6 both surfaces have less variation as if a low pass filter had been applied. However, as the TxPower of A_1 is increased the throughput of A_2 decreases significantly and when a TxPower of 8 dBm is reached both surfaces are very similar.

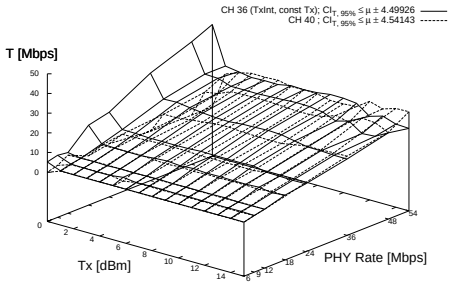


Figure 6. Application Layer UDP Throughput for Channel Separation of 1 and Antenna Separation of 20 cm

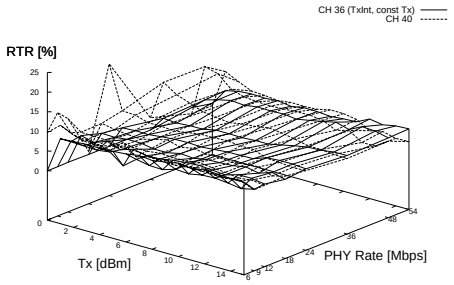


Figure 7. Retransmission Rate for Channel Separation of 1 and Antenna Separation of 20 cm

The aggregated throughput in Figure 14 confirms that T_{agg} is still significantly lower than double the maximum throughput of a single channel that would be expected when using two independent non-interfering channels. However, the throughput value for the highest TxPower and MCS reaches 47 Mbps which is just slightly above the theoretical maximum throughput of a single connection. The corresponding RTRs for this experiment, as depicted in Figure 9, indicates that after the TxPower value of 8 dBm is reached the RTR drops to almost 0 %; however, only for interface A_1 which increased its TxPower. Furthermore, interface A_2 always sent packets without requiring any retransmissions, except when the 64-Quadrature Amplitude Modulation (64-QAM) MCSs was used (i.e. 48 Mbps and 54 Mbps where the RTR goes up to 1 %). Hence, this shows that after reaching a specific ratio between a radio's TxPower and its neighbouring TxPower there is no longer interference from the neighbouring radio; there is however interference from an as yet uninvestigated source.

The channel separation was then increased to three, the corresponding throughput and RTR values are presented in Figures 10 and 11, respectively. As can be clearly derived from Figure 10, the performance of the link $\overrightarrow{A_2\,C_1}$ is no longer significantly affected by the neighbouring connection $\overrightarrow{A_1\,B_1}$. However, as shown in Figure 14 the expected total

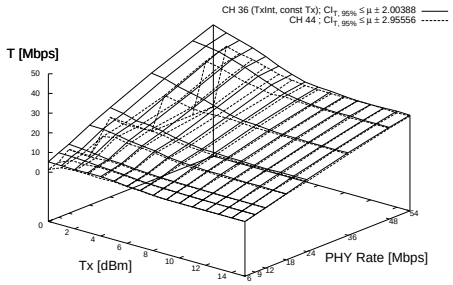


Figure 8. Application Layer UDP Throughput for Channel Separation of 2 and Antenna Separation of 20 cm

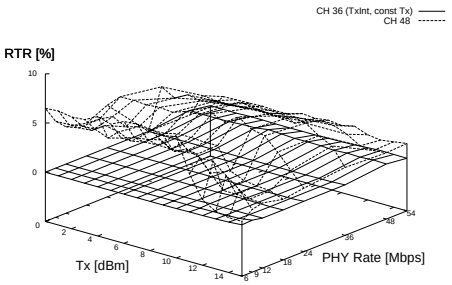


Figure 11. Retransmission Rate for Channel Separation of 3 and Antenna Separation of 20 cm

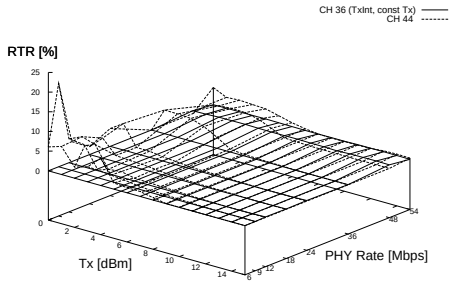


Figure 9. Retransmission Rate for Channel Separation of 2 and Antenna Separation of 20 cm

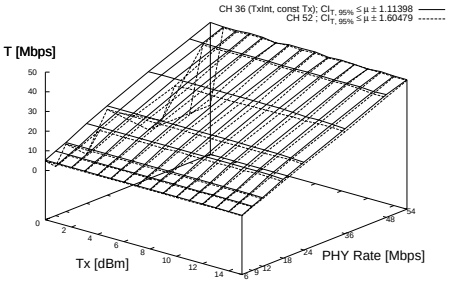


Figure 12. Application Layer UDP Throughput for Channel Separation of 4 and Antenna Separation of 20 cm

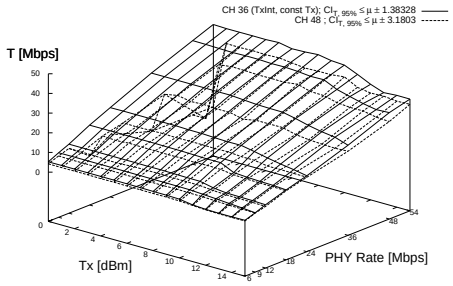


Figure 10. Application Layer UDP Throughput for Channel Separation of 3 and Antenna Separation of 20 cm

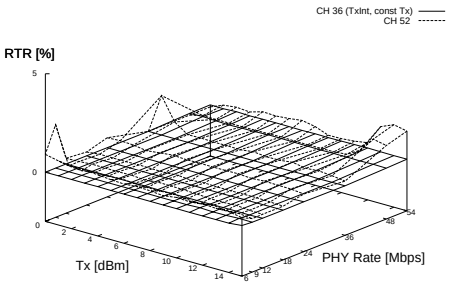


Figure 13. Retransmission Rate for Channel Separation of 4 and Antenna Separation of 20 cm

aggregated throughput is still not achieved. Rather an aggregated throughput of 62 Mbps with a TxPower of 15 dBm and a PHY rate of 54 Mbps is obtained. Taking the corresponding RTRs into account from Figure 11 it can be seen that there is still interference from A_2 in A_1 which causes packet distortion reflected by the high RTR of the link $\overrightarrow{A_1 B_1}$. Once A_1 was configured with a TxPower of 13 dBm both links share the medium equally but still with a low aggregated throughput.

Since both sending interfaces still interfere with each other and achieve throughput levels lower than the theoretical maximum, the channel separation was increased to four with A_2 operating on channel 36 and A_1 on channel 52. From the corresponding throughput and RTR results shown in Figures 12 and 13 respectively, it can be seen that both links no longer significantly affect each other and that the aggregated throughput is now almost twice the throughput of a single connection, as depicted in Figure 14.

To complete the set of experiments for an antenna separation of 20 cm and proving that both links $\overrightarrow{A_2 C_1}$ and $\overrightarrow{A_1 B_1}$ no longer interfere the missing aggregated throughputs for channel separation of five, six and seven, i.e., channel 36-56,

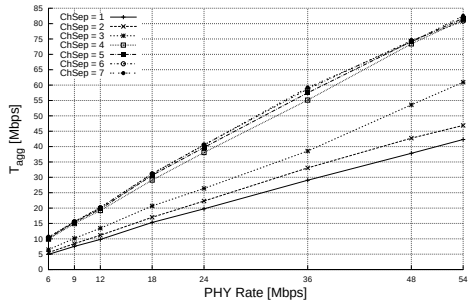


Figure 14. Aggregated UDP Throughput for Antenna Separation of 20 cm and TxPower 15 dBm

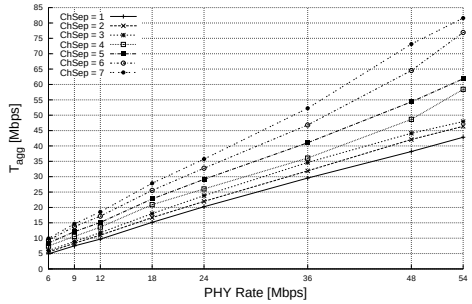


Figure 15. Aggregated UDP Throughput for Antenna Separation of 10 cm and Tx Power 15 dBm

36-60 and 36-64, respectively, are illustrated in Figure 14. As can be seen from these results, the aggregated throughput is always equal to twice the throughput of a single 802.11a UDP connection.

As mentioned earlier, experiments were conducted for an antenna separation of 20 cm; however a full set of experiments were also performed for the 10 cm case. Unfortunately, the stability and reliability of the links were extremely poor meaning that the obtained results were extremely difficult to interpret. As can be seen in Figure 15 the obtained results are however good enough to show the same general trends as seen in the 20 cm case. It can be argued that under these conditions the higher the channel separation, the better the aggregated throughput of the system. However, only when a channel separation of seven is reached - which means using channel 36 and 64 - does the aggregate throughput T_{agg} reach twice the throughput of a single UDP transmission. This shows that there are other significant factors to be considered other than just the side-band of a 20 MHz channel, as the side-band will only impact the adjacent channels. In order to explain these results it is assumed that each 802.11a chipset produces some interference over the entire 802.11 frequency spectrum that smoothely flattens out at higher channel separations to the frequency the adjacent radio is operating on.

Results for an antenna separation of 30 cm showing T_{agg} for a TxPower of 15 dBm are presented in Figure 16. As

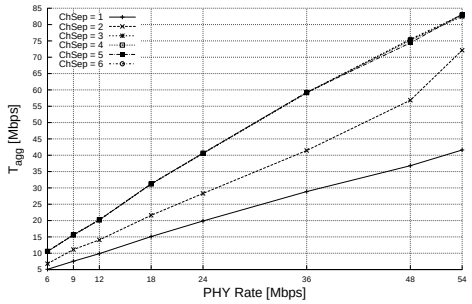


Figure 16. Aggregated UDP Throughput for Antenna Separation of 30 cm and Tx Power 15 dBm

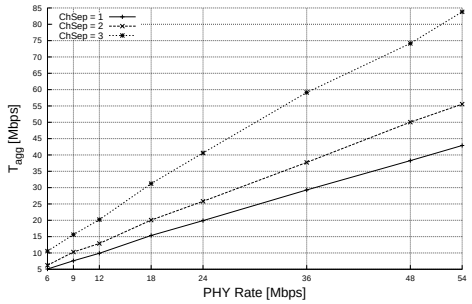


Figure 17. Aggregated UDP Throughput for Antenna Separation of 40 cm and Tx Power 15 dBm

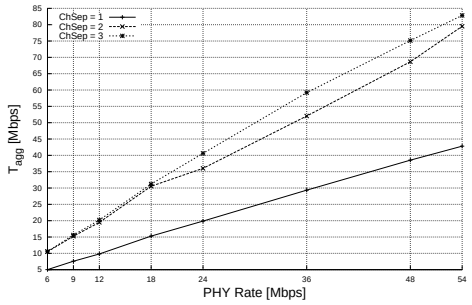


Figure 18. Aggregated UDP Throughput for Antenna Separation of 50 cm and Tx Power 15 dBm

depicted, when a channel separation of two is used the aggregate throughput achieved is equal to twice the possible throughput of a single 802.11a UDP link; this indicates that there is no more interference from the neighbouring radio as was the case for channel separations from four to seven in the 20 cm case (Figure 14).

C. Retransmissions, Carrier Sensing and Application Layer Loss

As shown in the previous Section V-B, under certain conditions two neighbouring transmitting radios can cause CS, RTRs and packet distortion. This was shown to be due to the effect of a high RTR which leads to a throughput decrease; however it is still not clear whether it is because the transmitted packet or the ACK was destroyed. To investigate

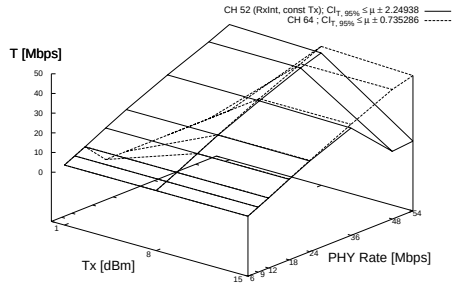


Figure 19. Throughput Measurement for TxRx Case, Channel Separation of 3 and with NoAck Policy Enabled

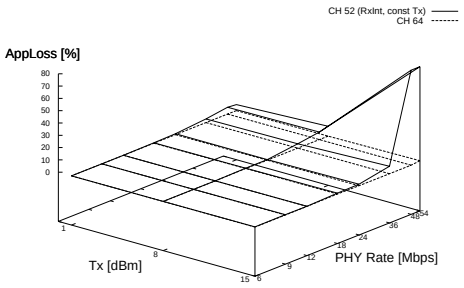


Figure 20. Application Layer Loss for TxRx Case, Channel Separation of 3 and with NoAck Policy Enabled

this issue further another experiment was conducted.

The same experimental environment as described in Section III-A was used, except in this case the channel configuration was changed to $\overrightarrow{A_2\,C_1}$ on channel 52 and $\overrightarrow{A_2\,C_1}$ on channel 64. The direction of link $\overrightarrow{A_2\,C_1}$ was reversed to $\overrightarrow{C_1\,A_2}$. Later on the term Transmission-Receiving (TxRx) will refer to the reversed testbed setting and Transmission-Transmission (TxTx) to the settings where both neighbouring antennas are transmitting data. Note, that for the sake of simplicity only TxPower values of 1, 8 and 15 dBm were considered for interface A_1. Furthermore, the No Acknowledgement (NoAck) policy of the MADWiFi driver was used to force the card to send and not wait for ACKs. Taking these settings and recording a log of the application layer loss provided by IPerf, it was possible to investigate whether the actual sent packet was destroyed or the returning ACK (in the cases where the RTR was between 5 to 10 %). To answer this question, Figure 19 and Figure 20 were produced which depict the logged data for the throughput of a TxRx use case where B_1 and A_2 were the transmitters and the corresponding application loss chart, respectively. With regard to the sent ACKs in the experiments described in the previous section, it can be clearly seen from both Figures that these management packets from B_1 to A_1 are not destroyed by the neighbouring antenna A_2 for any of the TxPower settings in A_2. This can be seen in Figure 19 where the throughput is equal to that of a single connection with no interference for the lowest MCS. Furthermore, the application loss presented in Figure 20 shows no losses at all for channel 52. Additionally, it can be seen that if both connections use the same TxPower of 15 dBm that by using 64-QAM with both Coding Rates (CRs) 2/3 and 3/4 A_1 destroys close to 80 % of the sent packets of C_1 to A_2. What has still not been investigated so far is the question whether both radios affect the data packets sent by the neighbouring interface for higher channel separations than one. In order to do so further experiments have been conducted. Figure 21 shows the chart for a TxTx set-up where A_2 again stays constant at a TxPower of 15 dBm

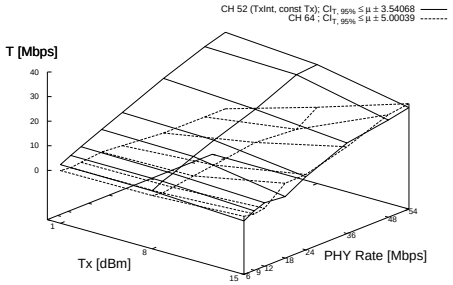


Figure 21. Throughput Measurement for TxTx Case, Channel Separation of 3 and with NoAck Policy Enabled

and A_1 changes its TxPower to 1, 8 and 15 dBm. Taking the corresponding application loss Figure 22 into account, it is worth noting that as A_1 increases its TxPower it got more often access to the medium which leads to a higher throughput. At the same time the throughput of A_2 drops down to the same level as A_1 when both operating with the same TxPower. Therefore, they consequentially share the medium equally over all TxPowers, as depicted in Figure 23 which shows the aggregated throughput. That no packet was destroyed on either A_1's or A_2 side can be confirmed by taking the application loss, as depicted in Figure 22. It is clearly observable from this figure that almost no packet was affected except for the case where A_1 sent at 1 dBm and with MCS 16-Quadrature Amplitude Modulation (16-QAM). However, it can be confirmed that by switching to 64-QAM there was no throughput value logged at all by IPerf. Therefore, both MCSs 16-QAM and 64-QAM are not robust enough to not get affected by the neighbouring interface if the neighbour's Equivalent Isotropically Radiated Power (EIRP) is much high than the outgoing power from the own antenna. Since the previous provided aggregated throughput Figures 15 to 18 show solely TxPowers of 15 dBm this application loss for 16-QAM and 64-QAM MCSs can be ignored.

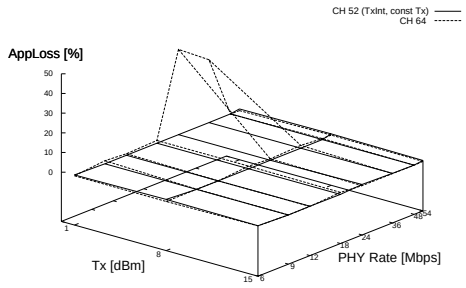


Figure 22. Application Layer Loss for TxTx Case, Channel Separation of 3 and with NoAck Policy Enabled

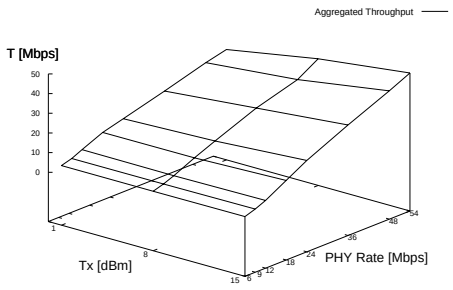


Figure 23. Aggregated Throughput for TxTx Case, Channel Separation of 3 and with NoAck Policy Enabled

VI. Conclusion and Future Work

The use of multiple radio equipped nodes within a WMN is the most promising approach for significantly increasing network performance.

A key problem with this approach however, is that antenna separations of less than 50 cm have a significant impact on the performance which can be achieved. The presented work here shows that it is not only ACI which has an impact but also ICI (channel separations of more than one). This problem does not appear to have been investigated previously and is the primary contribution of this paper.

The results were obtained based on experiments performed in a real testbed environment which was evaluated to produce reliable and reproducible results. This evaluation used CIs calculations and prior offline planning by using the Fresnel formula and statistical methods to design the testbed. The results presented show that by increasing the channel separation between co-located radios that the level of ICI decreases. All presented results take the radio parameters TxPower, MCS, channel separation and physical layer effects into account to explain the performance degradation due to carrier sensing, packet distortion and backing-off.

The results obtained will be used to develop an algorithm that takes the radio parameter settings, external dependencies and some prior knowledge as an input and provides the optimal global configuration of nodes in a WMN so that ICI will be minimised.

Acknowledgements

The authors of this paper would like to thank Mathias Kretschmer and Christian Niephaus for their initial work on the awarded paper.

The support of the Irish Research Council for Science, Engineering and Technology (IRCSET) and the Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

This work was partially funded by the European Commission within the 7th Framework Program in the context of the ICT project Carrier-Grade Mesh Networks (CARMEN) [15] (Grant Agreement No. 214994). The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the CARMEN project or the European Commission.

References

[1] Sebastian Robitzsch, Christian Niephaus, John Fitzpatrick, and Mathias Kretschmer. Measurements and Evaluations for an IEEE 802.11a Based Carrier-Grade Multi-radio Wireless Mesh Network Deployment. In *2009 Fifth International Conference on Wireless and Mobile Communications*, pages 272–278. IEEE, 2009. ISBN 978-1-4244-4679-7. doi: 10.1109/ICWMC.2009.52.

[2] Vangelis Angelakis, Stefanos Papadakis, Vasilios Siris, and Apostolos Traganitis. Adjacent channel interference in 802.11a: Modeling and testbed validation. In *2008 IEEE Radio and Wireless Symposium*, pages 591–594. IEEE, 2008. ISBN 978-1-4244-1462-8. doi: 10.1109/RWS.2008.4463561.

[3] Jens Nachtigall, Anatolij Zubow, and Jens-Peter Redlich. The Impact of Adjacent Channel Interference in Multi-Radio Systems using IEEE 802.11. In *2008 International Wireless Communications and Mobile Computing Conference*, pages 874–881. IEEE, August 2008. ISBN 978-1-4244-2201-2. doi: 10.1109/IWCMC.2008.151.

[4] Chen-Mou Cheng, Pai-Hsiang Hsiao, H. T. Kung, and Dario Vlah. Adjacent Channel Interference in Dual-radio 802.11a Nodes and Its Impact on Multi-hop Networking. In *IEEE Globecom 2006*, pages 1–6. IEEE, 2006. ISBN 1-4244-0357-X. doi: 10.1109/GLOCOM.2006.500.

[5] Vangelis Angelakis, Nikos Kossifidis, Stefanos Papadakis, Vasilios Siris, and Apostolos Traganitis. The effect of using directional antennas on adjacent channel interference in 802.11a: Modeling and experience with an outdoors testbed. In *6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and*

Wireless Networks and Workshops, pages 24–29. IEEE, 2008. doi: 10.1109/WIOPT.2008.4586029.

[6] Vangelis Angelakis, Konstantinos Mathioudakis, Emmanouil Delakis, and Apostolos Traganitis. Investigation of Timescales for Channel, Rate, and Power Control in a Metropolitan Wireless Mesh Testbed. In Paul Cunningham and Cunningham Miriam, editors, *ICT-MobileSummit 2009 Conference Proceedings*, 2009. ISBN 978-1-905824-12-0.

[7] Arunesh Mishra, Vivek Shrivastava, Suman Banerjee, and William Arbaugh. Partially overlapped channels not considered harmful. In *SIGMETRICS '06/Performance '06: Proceedings of the joint international conference on Measurement and modeling of computer systems*, pages 63–74, New York, NY, USA, 2006. ACM. ISBN 1-59593-319-0. doi: http://doi.acm.org/10.1145/1140277.1140286.

[8] R. Burchfield, E. Nourbakhsh, J. Dix, K. Sahu, S. Venkatesan, and R. Prakash. RF in the Jungle: Effect of Environment Assumptions on Wireless Experiment Repeatability. In *2009 IEEE International Conference on Communications*, pages 1–6. IEEE, 2009. doi: 10.1109/ICC.2009.5199421. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5199421.

[9] NLANR and DAST. Iperf. URL http://iperf.sourceforge.net/.

[10] Gerald Combs. TShark - The Wireshark Network Analyser. URL http://www.wireshark.org.

[11] MADWiFi. Multiband Atheros Driver for WiFi. URL http://madwifi-project.org/.

[12] Jerry Banks, John Carson, Barry L Nelson, and David Nicol. *Discrete-Event System Simulation, Fourth Edition*. Prentice Hall, December 2004. ISBN 0131446797.

[13] F K Wang. Confidence interval for the mean of non-normal data. *Quality and Reliability Engineering International*, Volume 17:257–267, 2001. doi: 10.1002/qre400.

[14] IEEE. IEEE 802.11-2007 Wireless LAN Medium Access Control and Physical Layers Specifications, June 2007.

[15] CARMEN. Carrier Grade Mesh Networks. URL www.ict-carmen.eu.

IEEE 802.16 Wireless Mesh Networks Capacity Assessment Using Collision Domains

Rafal Krenz
Faculty of Electronics and Telecommunications
Poznan University of Technology
Poznan, POLAND
e-mail: rkrenz@et.put.poznan.pl

***Abstract*— Wireless Mesh Networks (WMN) are considered an attractive alternative to the traditional wired backbone networks for broadband Internet access. However, their capacity is limited due to the nature of the radio channel, which must be shared by the nodes forwarding the traffic from and to the gateway. Therefore, estimating the capacity of WMNs is an important question. The capacity analysis proposed for ad hoc networks can not be directly applied to WMN due to some fundamental differences, e.g. a different traffic pattern and node density. The main contribution of this work is the application of collision domains concept to the IEEE 802.16 based WMNs. We consider a simple chain topology but the method can be extended to any arbitrary topology and the real world impairments (interference, fading, etc.) can be easily incorporated in the analysis. The presented results may have important implications for 802.16 mesh networks planning.**

***Keywords*- capacity analysis; collision domains; IEEE 802.16; mesh network**

I. INTRODUCTION

Broadband wireless internet access is becoming more and more popular nowadays. This is especially true since the introduction of IEEE 802.16 standard for local and metropolitan area network, called WiMax, in 2001. However, all the deployed WiMax installations use point-to-multi-point (PMP) mode of operation. The revision of 802.16 standard published in 2004 specified an optional mesh mode, where the nodes operate not only as hosts but also as routers, forwarding packets on behalf of other nodes that may not be in the range of the base station. WMN may form a self-configured and self-organized wireless backhaul network, which can be deployed incrementally, one node at a time, as needed, replacing a more costly wired backbone. However, multihop wireless communication is a relatively new idea and requires much research effort to analyze and optimize its performance.

In this paper we will concentrate on capacity aspects of WMN [1], with special emphasis on 802.16 standard [2]. Recently, a lot of research has been carried out to investigate the capacity of ad hoc networks, but their results can not be directly applied to WMNs due to several reasons which will be explained in Section II. Section III shortly describes 802.16 MAC protocol and the specific features of the mesh mode of operation. In Section IV we will show how the concept of Collision Domains, presented by Jun et al. in [3], can be applied to 802.16 mesh networks to estimate the capacity. This will be followed by discussion of nominal and effective load of the 802.16 Collision Domains as well as construction of collision domains in multi-channel and multi-radio configurations. We will show numerical results obtained using the approach presented before and their verification by means of computer simulation in Section V. Finally, the work will be concluded in Section VI, where the possible directions for future work will be presented as well.

II. RELATED WORK

In the past decade a lot of research have been devoted to determining the capacity of wireless ad hoc networks. In the fundamental work by Kumar and Gupta [4] the analytical lower and upper bounds of stationary network capacity have been derived and it has been shown that the throughput capacity per node reduces significantly when the node density increases. In [5] the authors analyzed ad hoc networks allowing node mobility and showed, that, if long delays are tolerated, the capacity remains constant with the number of nodes. The other interesting results related to this work have been presented in [6] and [7].

However, most of the results valid for ad hoc networks can not be directly applied to mesh networks due to some fundamental differences. They have been identified in [3] and are discussed below:

A. In ad hoc network the traffic flows between any arbitrary pair of nodes while in WMN practically all the traffic is gateway oriented. WMN's BS acts as a hot spot and may be a bottleneck of the whole network's capacity.

B. Topology of the WMN is rather stable, with new nodes occasionally joining or leaving the network, while an ad hoc network can change dynamically in both, number of nodes and number of links/connections.

C. There are no energetic constraints, nodes have access to external power sources.

D. As a consequence of C. nodes can have multiple radios which can increase throughput capacity significantly.

E. The number of nodes and the required bandwidth in WMN may be higher than in ad hoc network.

F. Most of the results focused on the theoretical analysis for the asymptotic case. The resulting capacity bounds do not reflect the exact capacity of the WMNs with a given number of nodes.

Consequently, another methods of WMNs capacity estimation must be developed, which will be discussed in the next sections.

III. OVERVIEW OF 802.16 MESH MAC PROTOCOL

The mesh mode of operation, introduced in 802.16d standard, is an important extension to the original PMP mode, with the advantage of less path loss, coverage and robustness improving exponentially as nodes are added to the network and larger user throughput over multi-hop paths [8], [9].

The TDMA MAC protocol designed for the mesh mode supports both centralized and distributed scheduling. In the centralized mode the mesh base station (BS), providing the connectivity to the wired backbone, is responsible for collecting bandwidth requests from subscriber stations (SS) and managing resource allocation. In the distributed mode, transmissions are scheduled in a fully distributed manner, without requiring any exchange of control information between the SS's and BS. Since decisions are taken locally by nodes, based on their current traffic load and channel conditions, the distributed mode is more flexible and responses quickly to the network requirements. Therefore, we will focus on the distributed mode only.

The TDMA frame structure used in the mesh mode is illustrated in Fig. 1. It is divided in the control and data sub-frames. The control sub-frame consists of 16 slots (transmission opportunities) and the data sub-frame is divided into multiple mini-slots. The control slots are accessed by nodes based on the distributed election procedure. Every node competes for the transmission opportunity using its neighbors' scheduling information and the procedure ensures that in a two-hop neighborhood there is only one node which can transmit its control message at a time.

The control slots are used to convey several types of control messages. Bandwidth negotiation is performed using MSH-DSCH (Mesh Distributed Schedule) message, which contains the schedule and data slots allocation of the broadcasting node and its neighbors. Consequently, each node can obtain scheduling information of its two-hop neighbors and data packet transmission is collision-free in the entire extended neighborhood. A three-way handshake procedure is used for data slot reservation. The negotiation phase consists of three steps:

1. the transmitting party sends out a request,

2. the receiving party responds with a grant,

3. the requester then confirms the indicated grant.

Such a mechanism is required since not all nodes are in the same transmission range in a mesh network [10].

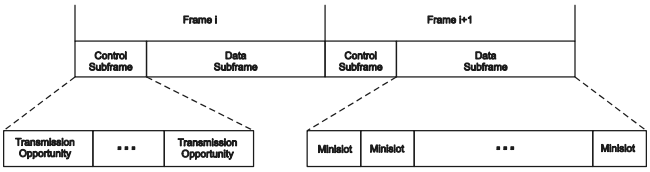


Fig. 1 IEEE 802.16 mesh mode frame structure.

IV. COLLISION DOMAINS IN 802.16 WMN

A. *Definition of collision domains in 802.16 WMN*

The concept of collision domains was applied to WMNs capacity calculation for the first time by Jun at al. in [3]. The method is based on the fact that the existence of gateways in WMNs introduces hot spots in the network that act as bottlenecks. Identifying the bottleneck collision domains allows computing exactly the minimum and maximum data rates available for each node for a given network topology and link layer protocol. The concept was further developed by Aoun and Boutaba in [11], by considering fairness to ensure proper operation of WMNs.

However, all the previously listed research considered 802.11 based WMNs only or did not take into account the specification of the MAC protocol at all. One of the key strength of the collision domains approach is the ability to include any MAC layer implementation by redefinition of collision domain. This is simply done by imposing a set of constraints (specified by the MAC protocol) on the links between nodes communicating in the mesh network.

The main contribution of this work is the application of collision domains concept to the 802.16 based WMNs, as specified in the 802.16 standard [2]. For clarity, let us

consider a simple chain of $N = 8$ nodes (SSs) receiving and forwarding traffic from the gateway (BS). We will assume that the traffic is unidirectional (downlink), the bidirectional case will be treated later in Section VI. We define the collision domain CD_4 for link $k = 4$ (between SS3 and SS4) as follows (see Fig. 2):

- the requester (SS3) broadcasts a *Request message*, notifying nodes SS2 and SS4 of its request – we include links 3 (SS2 → SS3), 4 (SS3 → SS4) and 5 (SS4 → SS5) in the collision domain CD_4,

- the receiver (SS4) responses with a *Grant message*, indicating the granted data slots to nodes SS3 and SS5 – we add link 6 (SS5 → SS6) to the collision domain CD_4,

- finally, we add link 2 (SS1 → SS2) to the collision domain CD_4, this step is done since node SS2 advertises its availability before node SS3 sends its grant confirmation (because of the cyclic way the control schedule is designed) and node SS1 is aware of the pending transmission in its extended neighborhood (and specifically between nodes SS3 and SS4),

- additionally, node SS1 will not accept any requests from BS (since SS1 is in the interference range of SS3) and, therefore, we add link 1 (BS → SS1) to the collision domain CD_4.

Unfortunately, the multi-hop hidden terminal problem has not been effectively eliminated in 802.16 MAC protocol. Referring again to Fig. 2, due to the uncoordinated channel access of node SS6, which is outside the extended neighborhood of node SS3 and may cause collisions at node SS4, we must include link 7 (SS6 → SS7) in the collision domain CD_4.

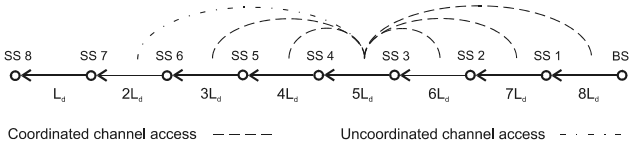


Fig. 2 Simple chain topology of WMN with 8 nodes.

B. Nominal and effective load

The definition of the collision domain presented in previous section allow us to compute the traffic to be forwarded within the collision domain. Assuming that each node in the chain (see Fig. 2) downloads from the gateway the traffic of $L_d\ [bit/s]$, a link which is closer to the gateway has to carry more traffic, e.g. link 1 (BS → SS1) has to carry the load of $8L_d\ [bit/s]$ while link 6 has to carry $3L_d\ [bit/s]$. Since each collision domain has to be able to forward the total load of its links, the collision domain CD_4 defined in the previous example forwards $8L_d + 7L_d + 6L_d + 5L_d + 4L_d + 3L_d + 2L_d = 35L_d$. If the bandwidth (the capacity of the MAC layer) for each link in the collision domain is constant and equal to $B[bit/s]$, the throughput L_d available to each node in the chain is limited to $L_d < B/35$ (for the considered collision domain CD_4). In order o find the collision domain which is limiting for the network (so called bottleneck collision domain) we have to identify the collision domain and its load for every link in the network and find the minimum throughput L_d [3].

The load of collision domain presented so far leads to a pessimistic value of the throughput L_d and is called the nominal load [11]. Instead, the effective load gives a more accurate estimate of the throughput by considering the spatial channel reuse, which is typical for multi-hop links in WMNs. Due to the spatial separation of links in the collision domain simultaneous transmissions are possible and should be deducted from the total load of the collision domain. Referring again to Fig. 2 we find out that link 2 (SS1 → SS2) can transmit simultaneously with link 6 (SS5 → SS6 - node SS2 is outside the interference range of node SS5) and we reduce the nominal load of collision domain CD_4 by the load of the lower loaded link in the pair. Similarly, link 1 (BS → SS1) can transmit simultaneously with link 5 (SS4 → SS5) and link 3 (SS2 → SS3) can transmit simultaneously with link 7 (SS6 → SS7). The effective load of the collision domain CD_4 is now equal to $35L_d - 4L_d - 3L_d - 2L_d = 26L_d$ and the throughput L_d available to each node in the chain is limited to $L_d < B/26$ (25% gain over the nominal load).

C. Impact of link adaptation (MCS)

The physical layer of 802.16 mesh mode is based on WirelessMAN-OFDM/TDD and features link adaptation (Modulation and Coding Scheme - MCS) for better utilization of radio resources [2]. Consequently, the raw data rates may vary from 2.40 Mb/s to 26.18 Mb/s (for 7 MHz channel), depending on the receiver SNR, which, in turn, is a function of propagation conditions as well as network topology.

With link adaptation every link in a specific collision domain may apply different MCS, which impacts the

collision domain load – now the load of every link must be scaled by the inverse of its bandwidth. Let us consider a simple example of a WMN consisting of two SSs downloading equal traffic $L_d\ [bit/s]$ from the BS. Assuming identical bandwidth B for both links the collision domain load is equal to $L_d + 2L_d = 3L_d$ and the throughput L_d is limited to $L_d < B/3$. However, if the bandwidth of the link 1 (BS → SS1) is twice that of the link 2 (SS1 → SS2), i.e. $2B$, the load of the collision domain is calculated as $\frac{L_d}{1} + \frac{2L_d}{2} = 2L_d$ and the throughput L_d is now bounded by $L_d < B/2$.

Generally, for the collision domain CD_k including N_k links characterized by load L_i and bandwidth B_i the following equation must be fulfilled:

$$\sum_{i \in CD_k} \frac{L_i}{B_i} = 1$$

The quantity L_i/B_i defines the percentage of time available for link *i*, since the transmission time (in other words the available resources) must be shared among all links forming the collision domain to carry all its load.

D. Capacity increase in multi-channel mode

The 802.16 mesh MAC protocol is designed primarily for multi-hop networks operating in a single channel. However, the nodes can employ up to 16 multiple non-interfering channels [2] for data transmission to increase the available throughput for nearby nodes, which can not exploit spatial reuse. Assigning additional channels is known to be one of the most effective ways to increase WMN capacity [8].

The collision domain concept presented so far can be easily extended to incorporate the WMN operating in multiple channels. This requires re-defining collision domains, having in mind both the specific MAC protocol, as well as the existence of multiple frequency channels for parallel transmissions. Consequently, the construction of collision domains has to be performed for each specific channel assignment separately.

Let us refer to the example discussed in section IV.A. Assuming the specific channel assignment, as shown in Fig. 3, we can remove from the collision domain CD_4 (defined for link 4, which communicates over channel B) links 1 (BS → SS1), 2 (SS1 → SS2) and 6 (SS5 → SS6), since they use channel A and do not interfere with link 4 (SS3 → SS4). Although link 5 (SS4 → SS5) operates on channel A as well, it can not be removed from CD_4 due to the single-radio configuration of nodes (the node can not receive and transmit simultaneously). This reduces the nominal load of the collision domain CD_4 to $6L_d + 5L_d + 4L_d + 2L_d = 17L_d$. Taking into account spatial channel reuse, the effective load is further reduced to $17L_d - 4L_d - 2L_d = 11L_d$ (pairs 5-7 and 3-5 can transmit in parallel).

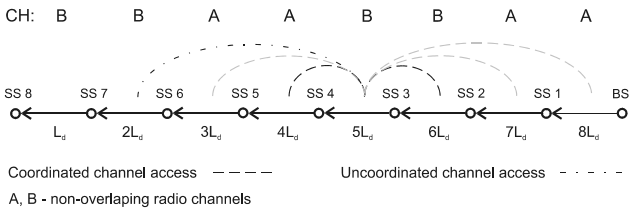


Fig. 3 Simple chain topology of WMN with 8 nodes configured in multi-channel mode.

E. Capacity increase with multi-radio nodes

As we could see in the previous section assigning additional channels to the link chain can substantially decrease the load of the collision domain, increasing the available throughput of the WMN. However, even if the number of available channels is very high, the collision domain can not contain less than three links in a single-radio configuration. This is caused by the fact that for a given link *i* both the transmitting node $i-1$ and the receiving node *i* can not receive (node $i-1$) nor transmit (node *i*) simultaneously with link *i*.

This limitation do not exist in a multi-radio configuration any more. Since cost of radios and battery consumption are not limiting factors in a WMN, multiple radios can be placed in nodes to increase the capacity of WMN.

Introducing multi-radio nodes releases the constraint limiting the capacity in multi-channel mode – now simultaneous reception and transmission (on different channels) is possible in every node. We refer again to the example discussed in section IV.A. If the specific channel assignment shown in Fig. 4 is applied, the collision domain CD_4 consists of links 3 (SS2 → SS3), 4 (SS3 → SS4) and 5 (SS4 → SS5), the nominal and effective load equals to $6L_d + 5L_d + 4L_d = 15L_d$ and $15L_d - 4L_d = 11L_d$, respectively in the single-radio configuration. By adding radios to nodes SS3 and SS4 the collision domain CD_4 reduces to the link 4 (SS3 → SS4) with load $5L_d$.

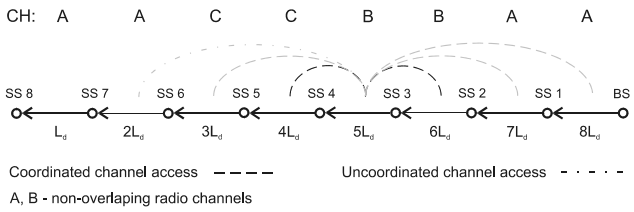


Fig. 4 Simple chain topology of WMN with 8 multi-radio nodes.

However, as we will see in the next section, multi-radio configuration requires a careful selection of channel assignment scheme to be really effective. When the number of channels is limited, multi-radio configuration may have no impact on WMN capacity in the worst-case scenario.

V. NUMERICAL RESULTS

A MATLAB® script was used to evaluate the method and we considered a chain topology similar to that presented in Section IV. The script identifies the collision domain for every link and calculates the available throughput. The results were verified using a custom-coded 802.16 mesh mode MAC simulator, written in C++.

A. *Simple Chain Topology – Unidirectional and Bidirectional Traffic*

Let us first analyze a single node downloading data from the gateway through a chain of forwarding nodes. Assuming the bandwidth B normalized to 1, the throughput available to the downloading node changes as $1/N$ for $N \le 4$, and reaches 0.25 for $N > 4$ (Fig. 5). The behavior is similar to 802.11 based chain [11], however, for 802.16 chain the nominal load of the bottleneck collision domain is $7L_d$ and the effective load is equal to $7L_d - 3L_d = 4L_d$.

If all nodes in the chain download the same traffic from the gateway the situation changes dramatically. In this case, for $N \ge 3$ the collision domain CD_4 (SS3 → SS4) is the most congested and forms a bottleneck. If the traffic is unidirectional (downlink or uplink) the throughput decreases to 0.1 (per node) for $N = 4$ and becomes as low as 0.03 for $N = 10$ (see Fig. 6). For bidirectional asymmetric traffic with $L_u = 0.1L_d$ (a value typical for ADSL links) the same throughput of 0.03 is reached for $N = 9$. If the traffic is symmetric ($L_u = L_d$), a similar throughput is obtained for $N = 6$.

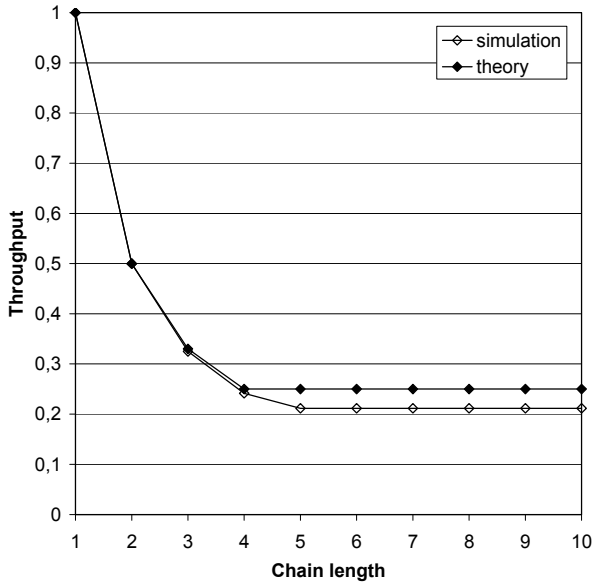


Fig. 5 Throughput vs. position of single downloading node.

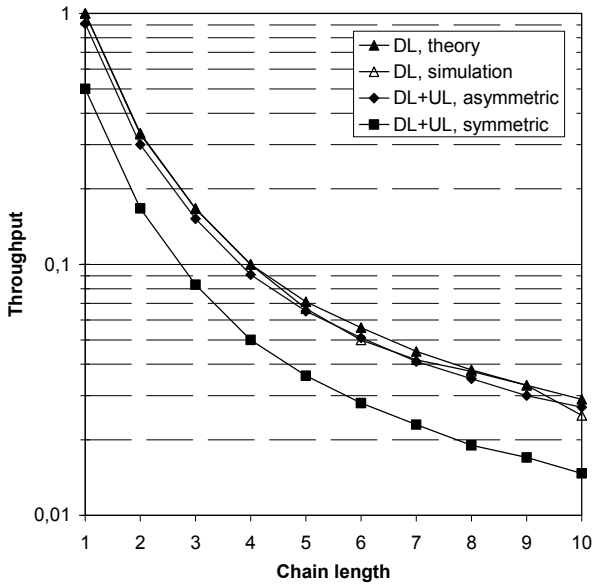


Fig. 6 Throughput vs. number of downloading nodes, unidirectional and bidirectional traffic.

The theoretical result for the downlink were compared against simulation and showed a good agreement (Fig. 6). However, the bidirectional case was not verified by simulation due to some limitations of the simulator, which currently supports unidirectional traffic only.

B. *Effects of Adaptive Coding and Modulation*

802.16 OFDM PHY layer used in the mesh mode features an adaptive coding and modulation scheme [10]. Using a modified version of the method presented earlier in this paper we can easily show the impact of the available bandwidth, which can vary from link to link, on the overall performance of the mesh network.

Doubling the bandwidth on the links 1 to 5 increases the throughput almost linearly (Fig. 7), however, modifying subsequent links does not increase it any more. The most important is the fact that the substantial change of the bandwidth impacts the location of the bottleneck collision domain, which can be observed by increasing the bandwidth of some links to $4B$. On the other hand, limiting the bandwidth to $B/2$ or $B/4$ on the links 1 to 4 decreases the throughput up to 57% and 85%, respectively.

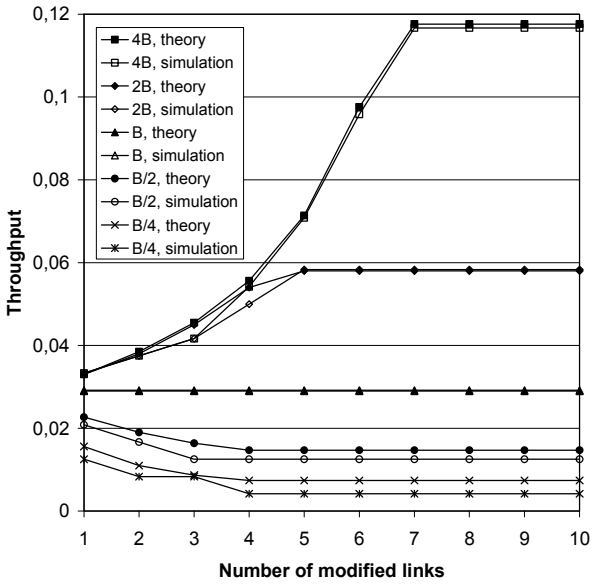


Fig. 7 Effect of adaptive coding and modulation – throughput vs. number of affected links in the chain.

From the above we can conclude that the performance of the 802.16 WMN depends heavily on the bandwidth available on the links closer to the gateway, while the bandwidth of the other links do not impact the performance of the network substantially.

Fig. 8 shows how the throughput is affected by changing the bandwidth of a single link in the chain. It should be noticed that the mesh network is more susceptible to the reduction of the bandwidth of the links than to its increase. This is important for 802.16 WMN planning, since the reduction of bandwidth of the links closer to the gateway reduces considerably the throughput available to all nodes. In the worst case scenario, limiting the bandwidth of link 1 to $B/2$ or $B/4$ decreases the throughput available to all nodes by 28% and 57% respectively.

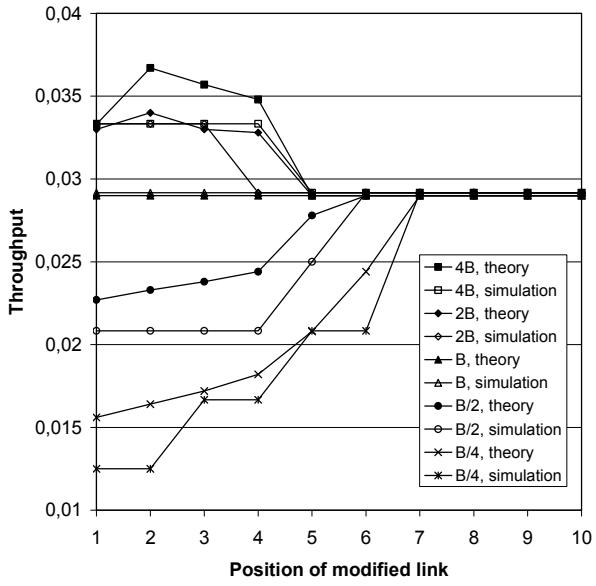


Fig. 8 Effect of adaptive coding and modulation – throughput vs. position of affected link in the chain.

C. *Multi-channel mode*

As mentioned in the previous section, assigning additional channels is one of the most effective ways to increase WMN capacity. However, networks exploiting multiple frequency channels require very careful selection of number of channels as well as their spatial configuration. We analyzed several configurations with 2, 3 and 4 channels, and some of the results are presented below.

A comparison of different channel assignment schemes in 2-channel configuration is shown in Fig. 9 and Fig. 10. From this figures we can find out, that the load of the collision domain is minimized (and the throughput maximized) when the channels alternate as often as possible (AABB – throughput increased by 100% for $N=10$), however, due to the limitations of the single-radio configuration the assignment ABAB should be avoided (see Fig. 11 and Fig.12).

Even if more channels are available for the 802.16 WMN at a given location, the throughput can be at most doubled in the single-radio configuration. The channel assignments schemes ABC and ABCD (Fig. 11 and

Fig. 12) as well as AABBCC (Fig. 13 and Fig. 14) performs identically as the previously analyzed AABB configuration. From this figures we can conclude, that in the simple 802.16 WMN chain and single-radio configuration the throughput can be doubled using 2 radio channels and assigning additional channels does not increase the throughput any more. However, the results may differ considerably for the WMN with more realistic (complex) topology and interference model and we plan to investigate the issue more deeply in the future.

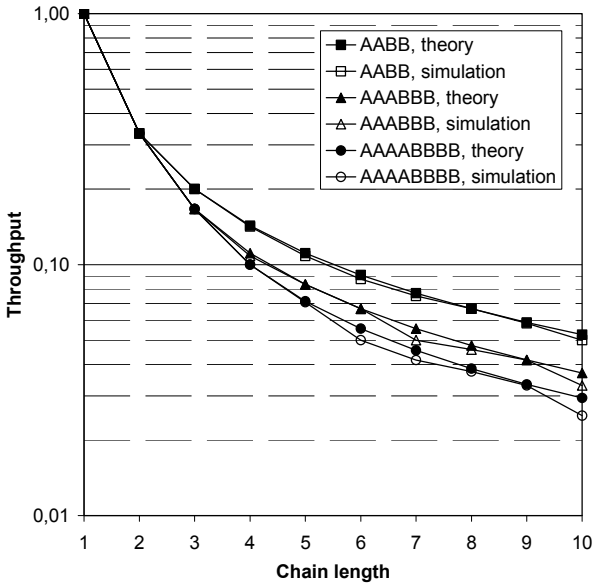


Fig. 9 2-channel configuration – comparison of 3 possible channel assignment schemes.

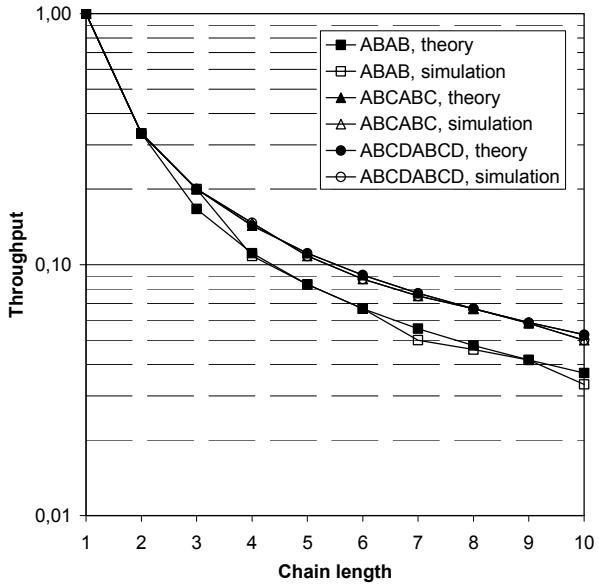


Fig. 11 Comparison of 2, 3 and 4 channel configurations.

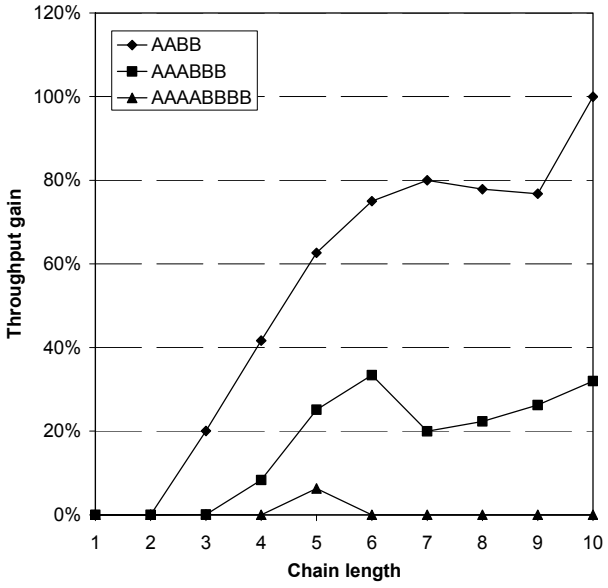


Fig. 10 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 9.

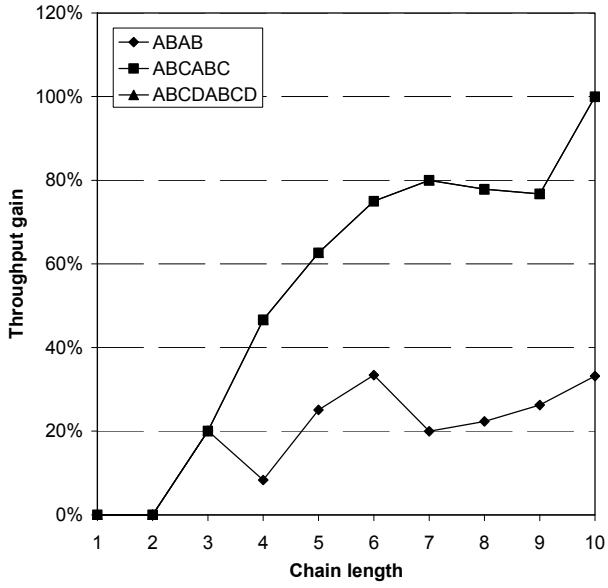


Fig. 12 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 11

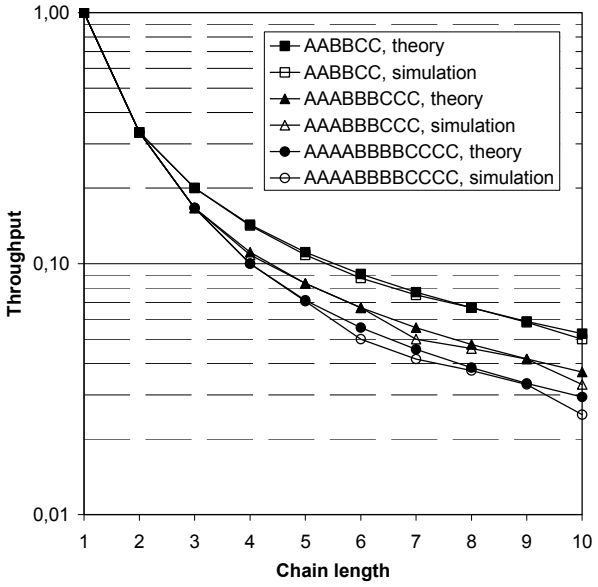


Fig.13 3-channel configuration – comparison of 3 possible channel assignment schemes.

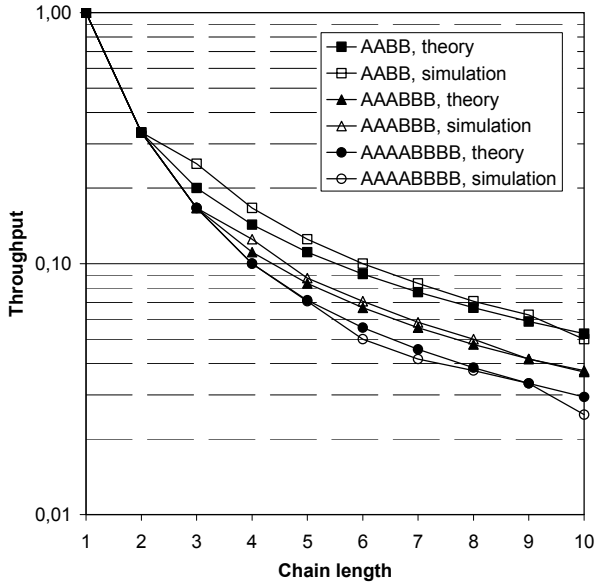


Fig. 15 Multi-radio configuration and 2-channel assignment schemes.

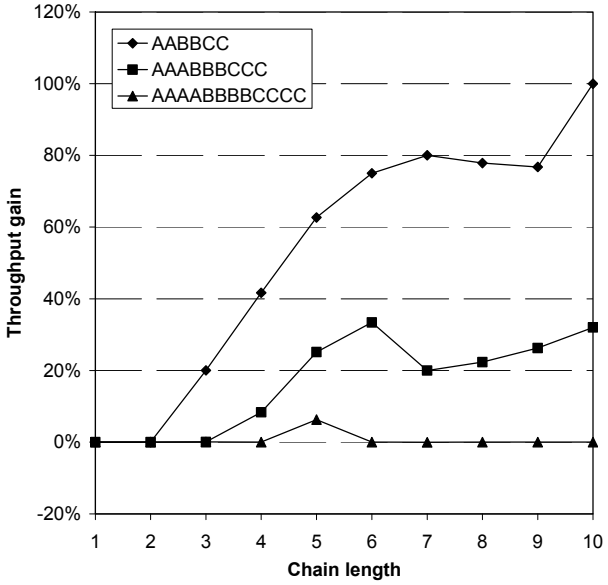


Fig. 14 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 13

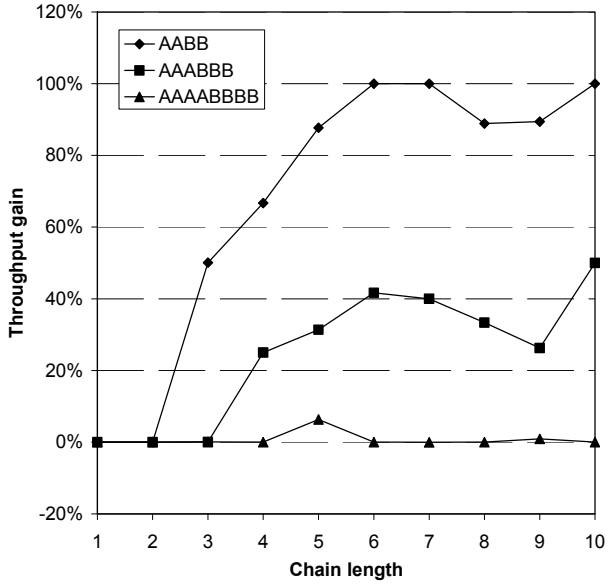


Fig. 16 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 15.

D. Multi-radio configuration

The further reduction of 802.16 collision domains load, resulting in the increased throughput, requires introducing multi-radio nodes in the network. As argued in Section IV.E the collision domain can be reduced to the single link with the sufficient number of radio channels in this case.

With two radio channels available to the WMN the throughput can be increased by 100% (Fig. 15 and Fig. 16) in the multi-radio configuration. We obtained similar results in the single-radio configuration, however with multi-radio nodes the throughput is doubled for shorter chains, i.e. 6 nodes vs. 10 nodes in multi-channel case.

The advantages of the multi-radio configuration are fully exploited when more than two radio channels are available (Fig. 17). Adding one channel increases the throughput by 130% for N=10 (ABCABC - Fig. 18), while four fold (300%) increase is observed for the 10-node chain configured with four channels (ABCD).

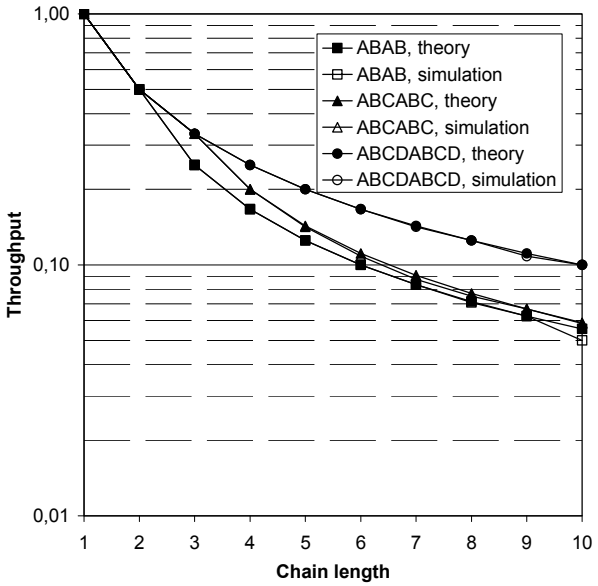


Fig. 17 Comparison of 2, 3 and 4 channel configurations with multi-radio nodes.

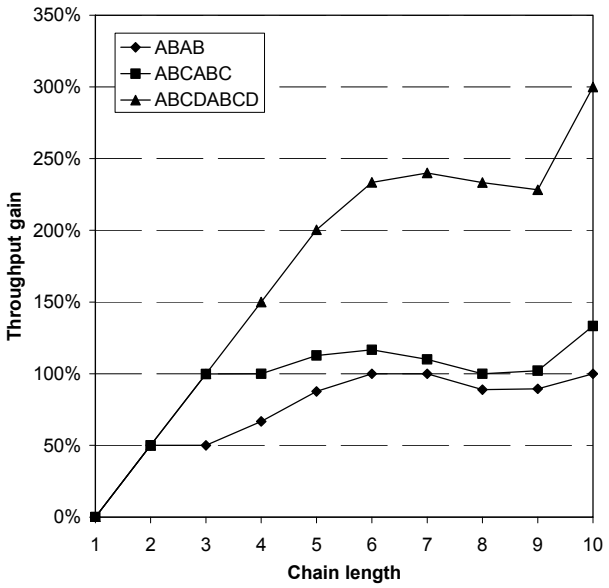


Fig. 18 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 17.

A comparison of 3-channel multi-radio configurations is presented in Fig. 19 and Fig. 20. Unlike the single-radio configurations, adjacent links should avoid operating in the same channel since the multi-radio nodes can receive and transmit simultaneously, if assigned non-interfering channels.

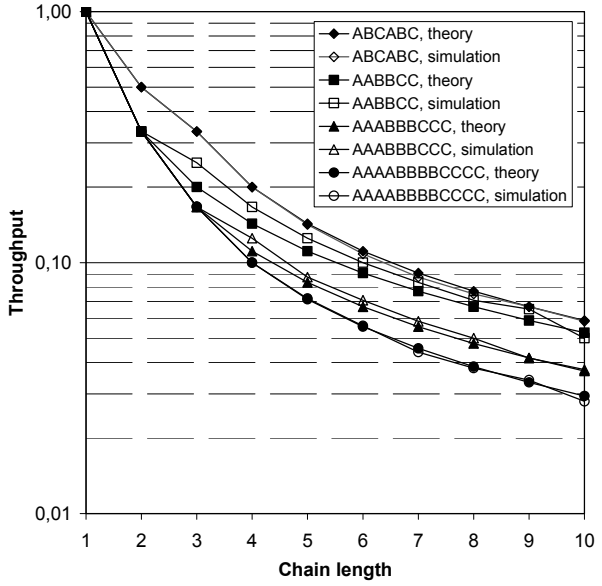


Fig. 19 3-channel configuration – comparison of 3 possible channel assignment schemes.

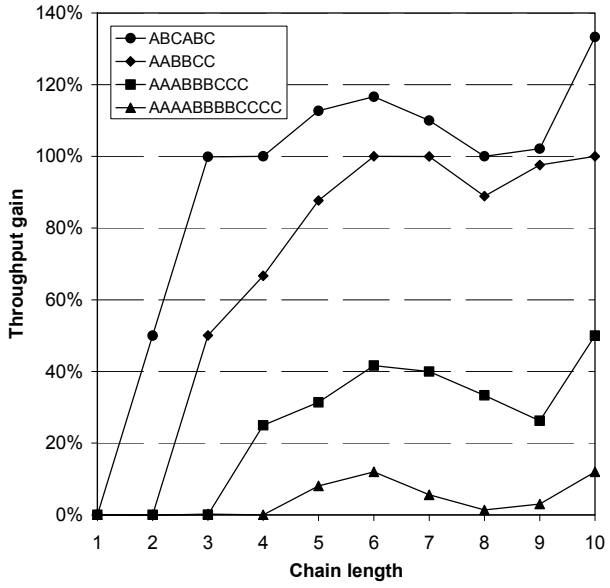


Fig. 20 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 19.

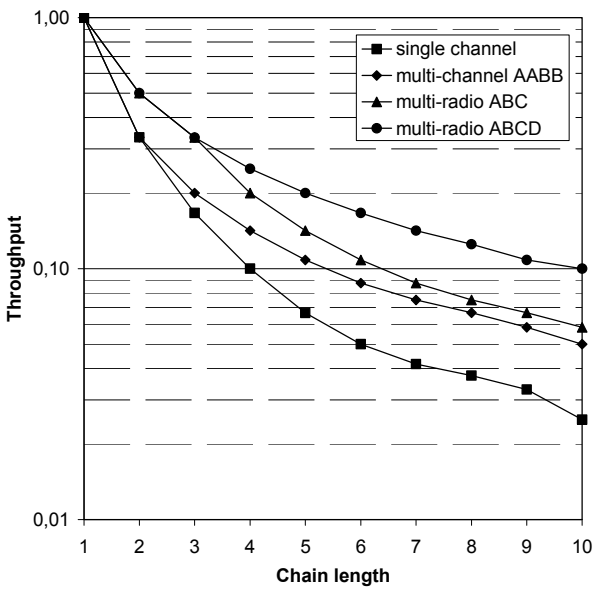


Fig. 21 Comparison of selected multi-channel an multi-radio configurations of the 802.16 WMN.

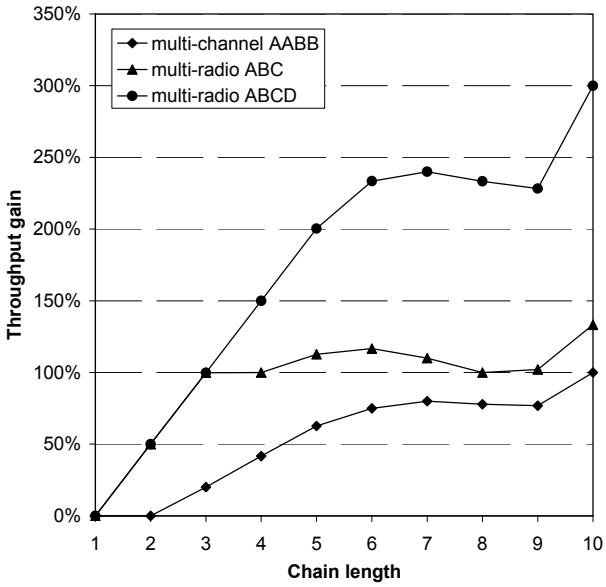


Fig. 22 Throughput increase (relative to the single channel) for the channel assignments shown in Fig. 21

Fig. 21 and Fig. 22 summarize the benefits of the multi-channel and multi-radio modes of operation of the 802.16 WMN. We can see from these figures that the simplest but effective method of increasing the throughput requires adding the additional (and properly configured - AABB) radio channel. The further throughput increase is possible by assigning another channels (ABC, ABCD) and replacing single-radio nodes by multi-radio devices.

VI. CONCLUSION

In this paper we showed how the concept of collision domains can be applied to 802.16 WMN capacity calculation. The definition of collision domain is strongly affected by the MAC protocol used. Therefore it must be redefined for every mesh network standard.

The collision domain concept proved to be exact, the results obtained by simulation of the 802.16 MAC layer are close to the theoretical analysis based on collision domains. We extended the method by allowing the variable bandwidth of each link, implemented in the 802.16 standard by the application of adaptive coding and modulation. The other extensions, like multi-channel and multi-radio terminals were included as well. The results obtained using the modified method may have important implications for WMN planning.

However, the simple chain topology and the unrealistic interference model considered throughout the paper does not allow for any generalization of the results. We plan to adapt the method to the arbitrary topology of the 802.16 WMN. Since the collision domains are defined based on the interference among links, this can be done only after implementation of the more realistic propagation models in our collision domain identification algorithm.

Finally, the mesh mode is very likely to be replaced by the relay solution in the new version of the 802.16 standard. Therefore, we will consider the application of the collision domain concept to the new mode as well as to the forthcoming IEEE 802.11s standard.

ACKNOWLEDGEMENT

This work has been supported by the VII Framework Programme European project NEWCOM++ (Network of Excellence in Wireless Communications) ICT-216715.

REFERENCES

[1] R. Krenz, 802.16 WMN Capacity Estimation Using Collision Domains, Proc. IARIA Int. Conf. on Advances in Mesh Networks (MESH), June 2009, pp. 115-119

[2] 802.16 IEEE Standard for Local and metropolitan area networks, October 2004.

[3] J. Jun, M. L. Sichtiu, The Nominal Capacity of Wireless mesh Networks, IEEE Wireless Communications, Oct. 2003, Vol. 10, No. 5, pp. 8-14.

[4] P. V. Gupta, P. R. Kumar, The Capacity of Wireless Networks, IEEE Trans. on Information Theory, March 2000, Vol. 46, No. 2, pp. 388-404.

[5] D. N. C. Tse, M. Grossglauser, Mobility Increase the Capacity of Ad Hoc Wireless Networks, IEEE/ACM Transactions on Networking, Vol. 10, No. 4, pp. 477-486.

[6] J. Li, C. Blake, D. S. J. De Couto, H. I. Lee, R. Morris, Capacity of Ad Hoc Wireless Networks, Proc. ACM Annual Int. Conf. on Mobile Computing and Networking MOBICOM, 2001, pp. 61-69.

[7] F. Eshghi et al., Performance Analysis of Ad Hoc Wireless LANs for Real Time Traffic, IEEE/ACM Trans. On Networking, Feb. 2003, Vol. 21, No. 2, pp. 205-216.

[8] I. F. Akyildiz, X. Wang, W. Wang., Wireless mesh networks: a survey, Computer Networks, 2005, Vol. 47, pp. 445-487.

[9] N. Nandiraju et al., Wireless Mesh Networks: Current Challenges and Future Directions of Web-in-the-Sky, IEEE Wireless Communications, Aug. 2007, Vol. 14, No. 4, pp. 79-89.

[10] C. Eklund et al., WirelessMAN – Inside the IEEE 802.16 Standard for Wireless Metropolitan Networks, IEEE Press, New York 2006.

[11] B. Aoun, R. Boutaba, Max-Min Fair Capacity of Wireless Mesh Networks, Proc. IEEE Int. Conf. on Mobile Ad-Hoc and Sensor Systems (MASS), 2006, pp. 21-30.

Simulation of Multihop Energy-Aware Routing Protocols in Wireless Sensor Networks

Adrian Fr. Kacsó
Computer Science Department
University of Siegen
57068 Siegen, Germany
Email: adrian.kacso@uni-siegen.de

***Abstract*—This paper provides and evaluates many simulation results concerning the energy consumption in WSN under different assumptions for various scenarios, including the impact of the routing strategy, broadcast delay, data aggregation, data rate, and the significant effect of MAC selection and configuration of its parameters. For simulations we analyze first the node's behavior in terms of energy consumption and investigate the impact of different parameters on it.**

To that aim, we use our sensor network framework (SNF), a flexible tool to build various protocols by combining existing building blocks at different layers (i.e., we provide also complete energy-efficient MAC modules). In this simulation environment routing protocols can be rapidly developed, closely inspected and the effects of changing configuration parameters and their impact on the performance better investigated and analyzed. We illustrate this in case of a two phase adaptive energy-aware routing protocol (with several routing metrics) as well as an enhanced, energy-aware directed diffusion and provide here various experimental results. After simulation and evaluation we are able to give guidelines for suited routing metrics and strategies, composition of protocols at different layers and how joint optimizations with MAC protocols increase the efficiency of routing.

***Index Terms*—wireless sensor network (WSN); routing protocols; energy-aware; simulation framework; modeling**

I. Introduction

A wireless sensor network (WSN) consists of a large number (hundreds, thousands) of sensor nodes that are randomly and densely deployed in a geographical area. Each of the distributed nodes in the WSN is able to collect large amounts of information, analyze and/or preprocess them and communicate them to a base node (sink). The nodes operate unattended and are forced to self-organize themselves as a result of frequent topology changes and to adjust their behavior to current network conditions. Typically, a sensor node has restricted communication (radio range) and computation capabilities, limited energy and memory. The communication is unreliable, messages can be lost or corrupted and sensor nodes can be damaged. The network topology changes also due to node transient failures, addition or depletion.

A very challenging aspect in query-driven WSNs is to determine the way the messages (query and data) are forwarded between the sink and sources (nodes able to deliver the requested data) using data-centric approaches. In such *data-centric routing* schemes the destination node of messages is specified by tuples of attribute-value pairs of the data carried inside the packets and not using globally unique identifiers (node address). WSN applications are usually interested in the kind of data and it is less important which node sent the data. When the distance between source(s) and sink is large, intermediate nodes forward the messages from hop to hop until they reach the intended destination, leading to several possible multihop paths. Determining which set of intermediate nodes to select in order to establish a path with the aim to prolong the network lifetime (by conserving the energy of the nodes as long as possible) is not trivial. Besides the energy-efficiency requirement a routing protocol for large WSNs must be reliable and scalable.

The paper is an extension of [1] and is structured as follows. Section II presents the state-of-art and the motivation behind designing energy aware routing protocols for WSNs. Section III describes the node software communication architecture. Section IV discusses how diverse routing protocols are built using our framework. Section V illustrates the performance of these protocols by giving various simulation results.

II. Related work, motivation and objectives

For WSNs, where multiple source nodes send data to a sink node (many-to-one communication pattern), establishing reverse paths is a very used scheme [2][3][4][5]. Many of the algorithms use distance-based forwarding, where the number of hops serves as a distance metric. Here each node selects the neighbor with the lowest hop counter to forward the packet. Since in most WSNs applications the battery of a sensor node is not replaceable, an important objective for routing protocols is the energy-efficiency. The biggest energy drain results from transmission of packets. Shortest-path routing improves the overall energy consumption since the energy needed to transmit a packet from source to final destination is correlated to the path length. Unfortunately, algorithms which minimize the path length will heavily load nodes on the path and those nodes drain off sooner, thus creating holes in the network, or worse, lead to disconnected networks.

Techniques to balance the load among all forwarding nodes are thus required [6][7][8][5]. One approach is to choose routes such that the variance in battery levels between different routes is reduced. By taking the amount of node's remaining energy into account we prevent nodes from choosing the same route often and thus increase the lifetime of frequently used

nodes on common used routes [9][6][10]. Minimizing the variance of the remaining energy of all nodes in the network is used in the lifetime prediction routing protocol [8], where the network lifetime is maximized. The lifetime of a node can be predicted based on the residual battery capacity and the rate of energy discharge.

In [7], Nurull et al. propose a route selection method that considers both the routing cost and the network lifetime metrics, achieving in this way a good tradeoff between these conflicting goals. Using a least cost route in order to optimize some cost (such as hop count, energy, delay, link quality, etc.) impacts on the network lifetime since nodes with higher communication demands might die soon.

According to Aslam et al.[9], the network lifetime maximization problem can be viewed as a max-min optimization problem. The proposed max-min zP_{min} algorithm selects routes that achieve a balance between the energy consumed by a route and the minimum residual energy at the nodes along the selected route. The basic idea is to select a route that uses at most $z \cdot P_{min}$ energy, where P_{min} is the energy required by the minimal energy route, and z is an adjustable parameter ($z \geq 1$). Between the routes with the total power consumption per path below $z \cdot P_{min}$, the route with the maximal minimum residual energy is selected.

Similarly, GBR [6] improves Directed Diffusion [2] by uniformly balancing the traffic inside the network using traffic spreading and in-network processing (aggregation).

Moreover, energy efficiency can be achieved by using greedy forwarding schemes which are aware of the geographic coordinates of the nodes as in [11][12]. For energy-saving approaches at MAC layer the reader is referred to [13].

Besides the energy constraint, a major concern in the design of WSN protocols is a reliable delivery of the data packets to sink. Radio communication links are known to have transitional region with widely varying degrees of packet loss and high error rates [5][14]. Although the successful packet reception decreases with the distance, there might be cases where distant nodes may have smaller loss than nearby nodes. That means that establishing energy-aware reverse paths using hop counter metrics might not always keep the overall packet transmissions energy to a minimum. In terms of energy, it might be more efficient to establish longer paths exhibiting low loss instead of shorter ones with poor link qualities where more energy has to be spent for successful packet delivery. In such cases new metrics are required such as the ratio of delivery rate and energy costs as in [5] or metrics that incorporate packet delivery rate and link asymmetry as in [14].

Besides communication links failures, sensor nodes can be damaged for a shorter or longer period of time. Routing protocols must tolerate such unreliable links and node failures. In the latter case, the routing protocol must react quickly to topology changes, especially when a route is affected by the failure of a node. The robustness to different types of failures can be improved by multipath routing [15][3][4][5], where multiple paths between source(s) and sink are established. Different strategies to construct disjoint and partially disjointed (meshed) paths and the tradeoff between energy-robustness are discussed in [15]. In Gradient Broadcast (GRAB) [4] and Minimum Cost Forwarding [3] packets travel to sink by descending a cost path. The cost is defined as the minimum energy needed to forward packets to the sink along previously established routes. All nodes receiving a packet with smaller cost forward it, meaning that the packet can reach the destination along several routes. This improves the reliability but increases the energy consumption since the packet is transmitted (superfluously) on more paths. To improve the energy consumption one can use, for example, multi-link energy-efficient forwarding as in [5]. In contrast to single-link forwarding, where the sender sends packets to one forwarder, the multi-link forwarding exploits the broadcast characteristics of the wireless shared channel. The idea is to broadcast the packet to a predetermined potential forwarder set. If the first node in the ordered forwarding set does not acknowledge the packet, the sender polls the next node in the set.
Note that reliability needs to be implemented at upper layers (as in Directed Diffusion [2]) when the transmission does not employ a MAC layer reliability mechanism such as the RTS/CTS/ACK handshake [16].

The behavior of the nodes between source(s) and sink(s) depends on several factors such as the network topology and connectivity, the number of active requests in the network, the position of source(s) and sink(s), communication pattern, MAC protocol parameters (e.g., listen and sleep times), application parameters (e.g., interest refresh rate) and so on. The cumulated impact of such a large amount of parameters on the (routing) behavior of individual nodes is difficult to be predicted. Our main goal is to simulate reliable routing protocols for WSNs able to prolong the network lifetime by conserving the energy of the nodes as long as possible. Therefore, we look after new metrics based on a sensor node's residual energy or other attributes to be able to take appropriate routing decisions in order to extend the lifetime of the WSN. Moreover, the energy consumption must be optimized at each layer of the node communication architecture and the cooperation between layers should be improved. We proposed in [17] a modular, energy-aware network architecture of a sensor node as a flexible approach to design and plug-and-play various protocols at network and MAC layers, and to combine and analyze the impact of different strategies (inside the protocols) on the performance and lifetime of the WSN. We implemented the SNF simulator using the OMNeT++ 3.4b2 discrete event simulation package[1] [18] and its Mobility Framework (MF[2]) 2.0p3 [19]. Part of the MAC protocols and the automatic energy component of the framework were described in [20]. Moreover, we extend the framework with new features such as add/delete, move (drag-and-drop), disconnect/reconnect nodes

[1]OMNeT++ is a public-source, component-based, modular and open-architecture simulation environment with applicability in the simulation of communication networks (http://www.omnetpp.org/)

[2]extension intended to support wireless and mobile simulations within OMNeT++

during simulation, which allow to analyze the impact of dynamic topology changes.

In the present paper we focus on alternatives to design energy-aware routing protocols using metrics that prolong the network lifetime and we illustrate the performance of these distributed algorithms using our automatic evaluation tool (based on our statistic component). The quantitative effort to design new protocols and to integrate them into a complete protocol architecture is considerably reduced since we promote modularity (in design) and code reusability. We illustrate here this by implementing two different routing protocols.

III. Node architecture

The common approach used to structure a communication protocol is layering; separate protocols are building the protocol stack where each protocol accesses functions of the lower layer protocol. Since for a resource constrained node strict layering is inappropriate [21][22], we employ a cross-layer design [23] by allowing exchange of information mainly across application, routing and MAC layers in order to optimize them. For routing protocols such optimization may improve the energy consumption during communication, extend the spectrum of routing decisions and adapt the communication to tolerate different kinds of failures or to avoid local congestion.

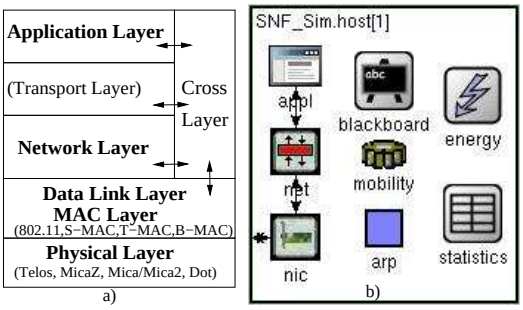


Fig. 1. a) Components of a sensor node b) in simulator.

The software comunication architecture of a sensor node is illustrated in Figure 1 and consists of application, network and NIC layer; the latter incorporates the MAC layer, the physical layer and the radio. We provided full implementations for the application and MAC layers, we extended the mobility component and added energy and statistics modules (see [20]). The existence of the blackboard component (module) allows cross-layer interactions between the layers.

Recall that we are looking for energy-aware routing protocols for long term query driven applications under the assumptions that the message sequence is not known in advance and several sinks inject requests in a large random sensor network.

A. Application layer

The application layer provides the user a general way to send his request to the network. In a data-centric approach, a request or *interest* is a sequence of attribute-value pairs such as *type=temperature, interval=100ms, area=[(x,y),(u,v)]* which is converted to an application packet. A request can be sent to a given area if the user specifies a rectangular area (defined by the coordinates of the two points). The application layer contains and simulates the sensing unit of a sensor node and sends responses or *data events*. Upon receiving a request, the node checks if it is a source of the requested data; if yes, the application layer starts the requested sensor to gather the data and packs it in a response message to be sent to the request initiator (sink). To that end, we employed one application layer than can be used by all routing protocols.

B. MAC layer

At MAC layer, the main energy consumer in a sensor node is the transceiver. To accommodate a low energy consumption, the main idea is to turn off the transceiver most of the time and to activate it only when necessary, meaning that it works at a low duty cycle. Thus, we provide implementations (as complete NIC modules) for energy-aware MAC protocols like S-MAC [24], T-MAC [25], Preamble Sampling [26] and IEEE 802.11 WLAN standard that can be used below a network layer routing protocol (including support for collision detection). To the best of our knowledge, so far there are no implementations for different MAC protocols (except a first attempt of the IEEE 802.11 WLAN) which can be embedded in the OMNeT++ sensor node architecture. This contribution will be made publicly available for download and is not the focus of this paper.

C. Network layer

In order to quickly build and experiment with different routing protocols, the network layer should be designed to allow fast reconfiguration and code reusability.

At network layer the possibilities to reduce the energy consumption are to communicate rarely and to reduce the volume of communication, i.e., the number of transmitted packets by using in-network processing (aggregation, compression, etc). In WSNs the network layer provides a (best-effort) connectionless multihop communication abstraction to the upper layers. Typically, its functionality includes packet forwarding towards one or many destinations, creation and maintenance of routing structures, retransmissions and acknowledgments. Note that the general functionality of a routing protocol remains the same, what is changing from one protocol to the other are the protocol logic and the strategies which use different metrics to route the packets. In order to better exploit this, we implemented in SNF the network layer architecture proposed in [17]. We shortly describe here the architecture (Figure 2), since we will have to refer to most of its components and their interactions.

It consists of three components: the in-out unit (IOU), the forwarding unit (FU) and several routing units (RUs). The main task of IOU is to guide the packet flow. The FU forwards the packets to the appropriate RU according to the packet type and the direction they are coming from. It also maintains and provides access to internal cache structures (Interest, Gradient and Neighbor Tables, etc.). The *Interest Table* is a dynamical, user-programmable table containing request relevant information. Since we allow more sinks to inject concurrent requests

into the network the interest is uniquely identified by the attributes: a node identifier, the type of requested data and the area under observation. Each interest entry contains a reference to a dynamical, user-programmable *Gradient Table* which contains the direction (node id) the packet came from and other routing relevant information. Unlike the Gradient Table, the *Neighbor Table* contains neighborhood information that is independent of the interest (energy, link quality or times when neighbors start and end their active period according to the used MAC protocol, etc.).

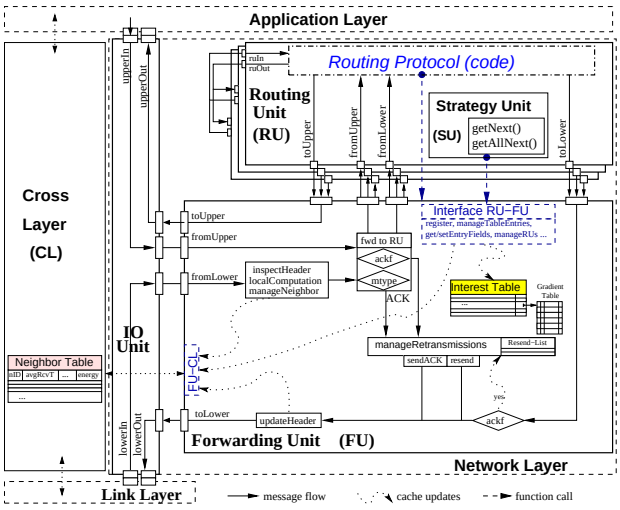


Fig. 2. Interaction of the components at network layer.

The RU is a dynamically exchangeable component implementing (part of) the routing protocol with the main task to forward packets by determining their next hops. The SU is an optional component inside the RU with the aim to modularly separate the policy of determining the next hops for packets inside the routing protocol. Since the architecture is not subject of the presented paper the reader is referred to [17] for more details.

D. Other components

Furthermore, we integrated in the node architecture:

- an energy component that models automatically the energy consumption in a sensor node (including also the energy consumption due to collisions, in order to have a realistic energy model),
- a statistics component with automatic visualization of relevant data, which enables a fast analysis of different performance criteria,
- a cross layer enabling interchange of significant information directly among the different layers of the protocol stack (without additional messages).

IV. Energy-aware multihop routing

Here we concentrate on two routing protocols that we implemented at the network layer: a two phase multihop routing and enhanced directed diffusion.

A. Two phase multihop routing

This is an adaptive energy-aware routing protocol based on two phases: the *interest propagation* and the *data transmission*, where the data is sent along the reverse paths established during propagation of the interest. The main idea of this routing is to find out what is the relevant routing information that shoud be spread to the nodes in the network without sending explicit routing messages.

During the first phase the interest packet propagates throughout the network, the cost field (hop count or other metrics) is established at each node and the gradient tables are created and initialized. Finally, each node has determined its minimal cost to sink and depending on the size of its gradient table, it knows a subset or all of its neighbors and their energy.

In the *data transmission* phase, the data packets are routed from source(s) to sink according to the minimum cost forwarding principle. The choice of the next hop is based solely on information available inside the gradient table and therefore the corresponding routing algorithm is sender initiated, where each involved node selects the ”best” next hop and sends data as unicast. This best next hop can be chosen according to several strategies (metrics).

The first strategy was to route the data packets on the shortest path between source and sink (hop count metrics, denoted in the sequel `hc`).

The second strategy, denoted `hc,E`, combines the hop count metrics with the neighbor residual energy. More precisely, each node records the neighbor with smaller hop count than itself and among these the node with the maximal residual energy is selected as its best next hop. However, choosing the neighbor with the highest energy level does not guarantee that the path to the sink along this node contains only relay nodes with high residual energy. It can occur that this path contains a bottleneck energy node, which should be avoided whenever possible.

In order to overcome this we consider a third strategy, denoted `Δhc/E`, which combines the hop count and the residual energy of each node on the entire path between source and sink. To that end each node u computes its cost as $M_u = \min\{M_v + 1/E_u | v \in Neighbor(u)\}$, where E_u is the residual energy of node u and M_v is the summation of the costs on the path from the sink up to and including node v.

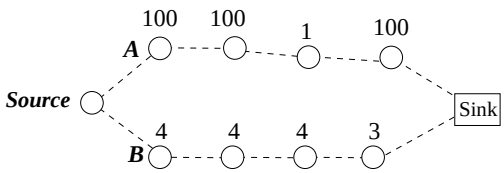


Fig. 3. Two disjoint paths from source to sink. The residual energy for each node is given.

This additive path cost function represents now a quantitative characterization for the goodness of the entire route. For example, if a relay has three alternatives nodes A, B and C with costs $1/20$, $1/10$ and $1/5$ respectively, it will choose the

route going through node A, since this obeys the minimum-cost forwarding principle. This third metrics contributes to better balance the energy consumption of the network by redistributing the traffic load more uniformly on the nodes. Still, there are particular scenarios where this strategy does not avoid a bottleneck node, as illustrated for a source with two disjoint long paths in Figure 3. The cost on the path through A is $1/100 + 1/100 + 1/1 + 1/100 = 1.03$ wheras through B is $1/4 + 1/4 + 1/4 + 1/3 = 1.083$. Thus Δhc/E selects the path with the smaller cost through A although it contains a bottleneck node with residual energy 1 (which will be soon depleted).

Alternatively, in order to avoid such node with low residual energy we introduce a forth strategy, denoted hc,cE, which employs uncorrelated the hop count and the critical energy (cE) on the entire path between the sink and source (i.e., the least residual energy on this path). Each node u computes $cE_u = \min\{E_u, \max\{cE_v | v \in Neighbor(u) \ \wedge \ hc_v \leq hc_u\}\}$, hence the hop count plays the main role in selecting the next hop, while the critical energy on the path just refines the decision. In order to enable this, each node maintains in its cache tables (§III-C) information about its neighbors including hop count, critical energy and the timestamp of last received critical energy message. To forward a data packet to sink a node uses the hop count and critical energy metrics (where the weight of each factor is adjustable). The node selects its next hop candidate (*cand*), out of three sets: nodes with smaller (*Set1*), equal (*Set2*) and higher (*Set3*) hop count (*hc*) by executing the pseudocode:

```
cand = cand1
if (cand2 ∈ Set2)
   if ((cand2.cE - cand.cE) ≥ eDiff1)
      cand = cand2
if (cand3 ∈ Set3)
   if ((cand3.cE - cand.cE) ≥ eDiff1*(1+(cand3.hc-cand.hc-1)*k)
      cand = cand3
```

eDiff1 and *k* are configurable threshold values. Note that for *Set3* we relax the condition $hc_v \leq hc_u$ to enable selecting neighbors that are one hop further away from the sink than the current node. The drawback here is the possibility of creating loops in the routing process since we lose the "right direction" information kept inside the hop count.

In order to avoid the drawbacks of strategies three and four we propose a further new strategy, referred as hccE, which correlates both the hop count and the critical energy on the path. Therefore, each node computes and forwards the pair: [hop count distance to sink; critical energy on path], as $(hc_u; cE_u) = (hc_v; cE_v) \oplus (1; E_u)$, where $(hc_v; cE_v)$ is the hop count and critical energy pair corresponding to node $v = arg\min\{hc_v/cE_v | v \in Neighbor(u)\}$ and the operator $\oplus$ is defined for each term as $hc_u = hc_v + 1$ and $cE_u = \min\{cE_v, E_u\}$.

Additionally, to alleviate the problem of excessive broadcasts during flooding caused by the fact that a node broadcasts instantly after receiving a lower cost without knowing whether this cost is minimal we introduce a waiting time T_w. A node will wait for a time T_w which is chosen either constant or directly proportional with the received cost field. During this period, the node extracts from all received packets the minimal cost field and if this is better than its own, it updates its local cost. Then it broadcasts the packet with its cost. It is obvious that if T_w is large enough the node broadcasts only once the minimal cost, but this introduces latency in the transmission.

To guarantee reliable delivery (per hop), data packets are unicasted using RTS/CTS/ACK handshake at MAC layer or they are marked as relevant at network layer (which enables the network layer reliability mechanism).

Note that the interest is refreshed (broadcasted) at configurable intervals depending on the data generation interval. The refreshes are necessary in order to notice changes in the topology (new or failed nodes) and to propagate the path critical energy information in the network.

B. Enhanced directed diffusion

To illustrate the usefulness and the possibilities of the proposed SNF we present here also how a complex routing protocol such as directed diffusion (DD) can be realized. The protocol we implemented, referred as enhanced, energy-aware directed diffusion (EDD), is a variant of the original DD [2], including energy-awareness mechanisms (metrics which consider the residual energy of nodes and geographic coordinates [11]). The directed diffusion protocol is considered complex since it combines discovery, querying and routing mechanisms in a single protocol. Therefore, we decided to decompose it in phases, which can be modularly embedded in our architecture of the network layer. We discuss how we decomposed the protocol in phases (including our changes to the original) and show how its complexity can be reduced by employing simple routing units, which later can be exchanged and variably configured (by reusing the code) to achieve an energy-aware routing.

We briefly describe the four phases and the different types of messages used; we name the phases according to the operations the sensor node executes in each of them. Generally, the operations correspond to some known traffic patterns (given below for each phase in parenthesis) in the WSN:

Phase 1: Interest propagation (flooding, 1:all)
A sink aiming to subscribe to certain events creates an interest and injects it in the sensor network. The original interest has a low data rate and is broadcasted to the one hop neighbors (Figure 4.a). Intermediate nodes receiving the interest create and add a gradient entry in their gradient table, containing the interest attributes and the sender node from where the interest was received. These gradients allow the node to route back data matching the interest (a node knows only the one hop neighbor which sent the interest, not also the initiator of the interest). If the interest has not yet been seen, it is rebroadcasted by each intermediate node. This way the message is forwarded hop by hop in the entire network.

Phase 2: Path establishment (convergecast, k:1)
Once the interest reaches potential sources, the sources reply with an exploratory data message at the (low)

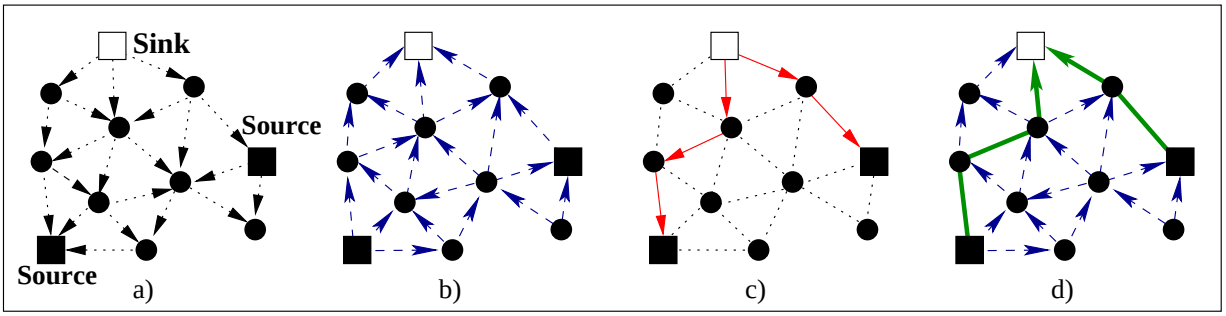


Fig. 4. EDD phases a) interest propagation, b) exploratory data convergecast, c) reinforcement of one neighbor to each source d) data propagation

requested rate, in order to find paths to the sink. Using the gradients (corresponding to this interest) from phase 1, the exploratory data messages are sent back hop by hop (via multiple paths) to the sink (Figure 4.b). This phase is referred in original diffusion as initial gradient setup phase.

Phase 3: Reinforcement (multicast, 1:k)

The sink receives one or more exploratory data messages. It selects the *best*[3] exploratory data message and a reinforcement message is sent towards the neighbor sending this message. Upon receiving a reinforcement message, each node creates a reinforced gradient for that neighbor, chooses the best next neighbor towards the source and resends the reinforcement message to the selected neighbor (Figure 4.c). Recall that the selection process (i.e., which node should be reinforced) is a local decision based on the contents of the data cache. This data cache (in our architecture the `DataInitiator Table`) is similar to the gradient cache but in the opposite direction, to route messages (reinforcement) from sink to source nodes. To be able to reinforce a given neighbor each sensor node stores, for each known interest, a data cache with recently received data messages. If the same data message is received several times, it is silently discarded. Thus, an intermediate node knows only two things: where to forward the incoming data message and which neighbor has been reinforced.

Phase 4: Data delivery (restricted convergecast)

When the reinforcement message reaches a source, a complete path of reinforced gradients exists between source and sink (Figure 4.d). Data messages can now be routed from each source to the sink using exactly one path.

In each phase a special message type is forwarded inside the network. There are several relevant messages types:

- **Interest** – Each time a request is injected at the sink an interest message is created. The interest is a set of attribute-value pairs containing at least the type of the requested data, a data generation interval and the expiration deadline for the interest. The original interest message is flooded, each intermediate node maintaining a gradient entry towards the sender of the interest. The sink refreshes the interest message periodically, in order
  - to announce that it still wants the data,
  - to update the node state and to discover topology changes (i.e., new/depleted sensor nodes),
  - to reach all nodes in case one or more of the previous interest messages were lost (due to collisions, communication failures, etc.).

  Additionally, if the network has been partitioned, the absence of the refresh message notifies the source that the sink cannot reach it. Correspondingly, the source may decide to delete the request after a given time. Between the original interest and the refresh interest there is no difference.
- **Exploratory data** – The exploratory data message is used by the source to explore or discover a path to the sink, as the sink and the source nodes do not know each other. When a node receives an interest message, it inspects the request and, if the data information that it can deliver matches the requested data, it declares itself a potential source for that interest. Data messages are initiated by a source and are forwarded (with the identity of the source) to *all* gradients in order to reach the sink. This creates multiple exploratory data messages reaching the sink. These exploratory data messages are cached at each node; they are refreshed periodically (with a lower data rate) for the same reasons as the interest message.
- **Reinforcement** – The positive reinforcement message is a modified interest, usually with a higher data rate, sent by the sink upon receiving one or more copies of an exploratory data message. In order to reduce redundancy of data messages, the sink selects one or *several* preferred neighbors as starting points for path(s) to reach the source(s). The selection criterion is usually latency, that means the fastest[4] neighbor (the first node that sent a not yet known exploratory data message) is selected. Checking the initiator of these data messages is possible only by caching the exploratory data messages and being able to distinguish among them.

  The reinforcement message is forwarded by each in-

[3]The meaning of the term varies according to the goal of routing. For example, when low latency is required the fastest neighbor is selected.

[4]Other possible criteria are: energy, hop-count, etc.

termediate node, based on the same next hop selection criterion until it reaches the source(s). Additionally, each intermediate node upon reception of the reinforcement message marks the reinforced gradient.
The **negative** reinforcement message is a path maintenace message. If a new improved path is discovered, the node has the possibility to use the better path and to diminish the importance or to deactivate the previous one. The old path can be used as backup path or becomes a normal gradient.

- **Data** – The data messages contain the actual data. A source initiates with the interest's requested rate periodically a data message, which has to be forwarded along the reinforced path. Each intermediate node can perform in-network processing on the data message received. The most common processing is a store and forward function, where some kinds of data reduction (e.g. aggregation) or filtering are performed before forwarding it.

According to the identified phases we can now split the protocol in several routing units as illustrated in Figure 5.

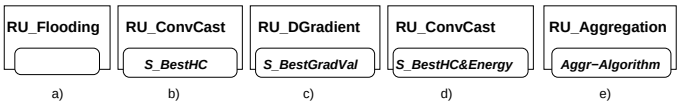


Fig. 5. a) Interest propagation, b) Convergecast using a subset of the gradients, c) Reinforcement using a subset of data-gradients, d) Data delivery and e) Aggregation.

We employ five routing units, the first four corresponding to the four phases of the protocol and the fifth one is an aggregation unit responsible to control the aggregation of data messages. Each of the routing units (excepting the `RU_Flooding` unit) has an attached strategy unit which specifies its policy (to achieve its goal).
The `RU_Flooding` unit, responsible for the interest propagation, does not need a strategy since it uses flooding (without any decision to be taken). The communication traffic cannot be reduced during this phase (except for geographic flooding, see §V-I), but the gathered information will be employed to considerably reduce the traffic in the subsequent phases.

Unlike the original directed diffusion where data messages are forwarded to all gradients (still flooding in phase 2), in our enhanced version we restrict the set of gradients to several of them according to a customizable criterion. We implement this phase in the `RU_ConvCast` routing unit and allow the use of any metrics (from simple one like hop count, latency, energy, etc. to combined ones like in §IV-A). Furthermore, the sources start to send their data exploratory messages only after receiving several refreshes for the same interest, in order to give the network time to stabilize.
The second and forth routing units are the same since both phases propagate data messages (either exploratory or real data messages). The only difference consists in the strategy (unit) used: in Figure 5 the `S_BestHC` strategy forwards the (exploratory) data messages in phase 2 (only) to neighbors with smaller hop count and the `S_BestHC&Energy` strategy to neighbors that besides a smaller hop count have also enough residual energy. It is also possible to consider for the second phase a strategy that consider principally the energy on the path without the hop count. In this way one get maybe longer paths and in the fourth phase one can use a hop count strategy to select the shortest path between them.

The third routing unit `RU_DGradient` uses data gradients (gathered in the second phase) and reinforces the nodes on the path according to the `S_BestGradVal` strategy. This strategy uses a path additive metrics (similar to the $\Delta hc/E$ strategy from §IV-A), where the path along nodes having better values are selected. Since the reinforcement process works always the same, but the decision which path to reinforce (the fastest one, the path with highest residual energy, etc.) the user can choose among all these (loadable) `SU`s the one more appropriate for his application.
Due to the fact that we generally keep several reinforced neighbors (according to the metrics) also in phase 4 we offer flexibility in chosing an energy-aware policy in order to better balance the data transmission load among the reinforced paths.

Hence, our enhanced version of directed diffusion allows a much better tuning during the protocol, which offers more flexibility and leads to a gradually and considerably diminuation of the communication traffic.

As each cost-field approach, directed diffusion scales well for large networks since the number of gradients kept in nodes depends on the number of requests and density (neighborhood) not on the number of sensor nodes.

V. Simulation results

The ultimate goal of running a simulation is to provide results and to get some insight into the behavior of the sensor network by analyzing the obtained results. We start with the two phase multihop routing protocol given in §IV-A using a 48 nodes WSN, continue in Section §V-I with a simulation scenario for EDD and conclude with some remarks about the reconfiguration of our simulator.

A. Impact of the routing strategy on the energy consumption

The application requirements assume that the data generation interval is set to $200ms$ and the request is refreshed at a $5s$ interval. We assume one sink (node 21) and a zone with only one source (node 16) as illustrated in Figure 8. Simulation results for the impact of different routing strategies on the energy consumption are given in Table I.

48 nodes net Energy consumed	**Strategy1** hc (hop count)	**Strategy3** Δhc/E (energy)	**Strategy4** hc,cE (critical energy)	**Strategy5** hccE combined
Max [mJ]	5.520	4.957	4.402	4.935
Max-Min[mJ]	3.625	3.075	*2.507*	*3.038*
Std. dev.[mJ]	*1.140*	977	*797*	*989*
Total [mJ]	*145.689*	147.972	150.557	148.769

TABLE I
Impact of strategies on the energy consumption.

As can be seen, an energy-aware approach leads to a better overall behavior of the network. More precisely, the results show that the better metrics are combinations of hop

count with critical energy (strategies 3,4 and 5); hereby the overall energy consumption is slightly larger (less than 3.3% in comparison with strategy 1), but the consumption is better balanced among nodes (up to 30% smaller standard deviation than for strategy 1), which extends the network lifetime.

When using only the hop count metrics we notice a very unbalanced energy consumption of the nodes (see standard deviation, the difference Max–Min); nodes on minimal hop count paths will soon be depleted while nodes with enough energy on comparable paths are not employed at all.

Figure 6 illustrates the *sorted* energy consumption of all nodes using T-MAC protocol with all the four strategies given before. The energy is sorted in order to better visualize the distribution of the energy consumption on the nodes. One can notice that strategies 4 and 5 using both the hop count and critical energy have the positive effect that they balance the overall energy consumption on more nodes. This can be observed by comparing the plots of *Tmac-hc,CE* and *Tmac-hcCE* with *Tmac-hc*; in the first two plots the curve is smoother than the last one, with no more than 3.3% overall energy increase (see Table I).

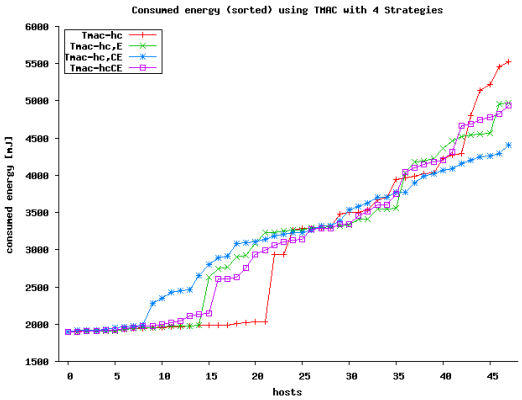


Fig. 6. Energy consumption using different strategies (same T-MAC).

To analyze the energy consumption on each individual node one can plot the energy unsorted as illustrated in Figure 7.

One can separate the nodes according to their position to the established path in three classes: a) nodes on the path (relay data messages according to the requested data interval), b) 1-hop neighbors to the path (receive data message updates and react protocol dependent on receptions for which they are not the intended receivers) and c) nodes more than 2-hops away from the path (not involved in data message transmissions, they receive interest refreshes or synchronization messages). The energy consumption decreases according to these classes. For example, for the first strategy using the hop count metrics, the nodes on the path, namely 1-7-2-11-20-21 in Figure 8, consume more energy than the other nodes.

B. Cumulative impact of the broadcast delay and strategy

We illustrate further the impact of the broadcast delay (T_w) on total energy consumption. It would be expected that for larger values of T_w the energy would decrease. This is true for MAC protocols without or with a fixed active-sleep regime.

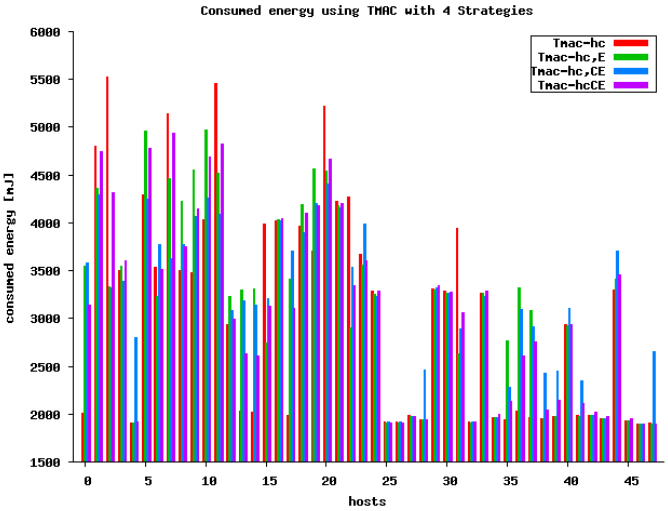


Fig. 7. Energy consumption of each node using the four strategies.

In case of low duty-cycle protocols, like T-MAC, the situation can be somewhat different. Adjusting T_w to achieve a better energy consumption remains difficult and we give some results for simulations configured with T-MAC. Nevertheless, the cumulative impact of different link (MAC, radio), network and application parameters should be further studied/analyzed to find out if an optimal value for T_w exists.

Since we set the listen time and frame time for T-MAC to 30ms and 600ms respectively, we run the simulation for three different values for T_w: 0.4ms, 20ms and 600ms. The simulation results for `hc` and `hccE` strategies are given in Table II, where the total energy consumption in WSN is given for each value of T_w. The simulation time was 2min and each of the 3 sources generates 5 data packets/s.

Energy [mJ]	Broadcast delay (T_w)		
Strategy	0,4ms	20ms	600ms
`hc`	85.403	84.759	85.221
`hccE`	91.868	89.964	90.964

TABLE II
IMPACT OF T_w ON ENERGY CONSUMPTION.

The best energy consumption is achieved with a broadcast delay of 20ms, thus not for the highest value of T_w.

We next check the impact of these values on the number of rebroadcasts.

Rebroadcasts	Broadcast delay (T_w)		
Strategy	0,4ms	20ms	600ms
`hc`	1.067 (2)	1.053 (4)	1.052 (4)
`hccE`	1.534 (2)	1.268 (0)	1.063 (1)

TABLE III
IMPACT OF T_w ON THE NUMBER OF REBROADCASTS.

In order to count how often each node rebroadcasts an interest we stop the refreshes after 112s (the simulation runs until 120s). In this way, we guarantee that each interest refresh reaches all the nodes (no one is still propagating). That means that each node should broadcast at least 22 times (interest

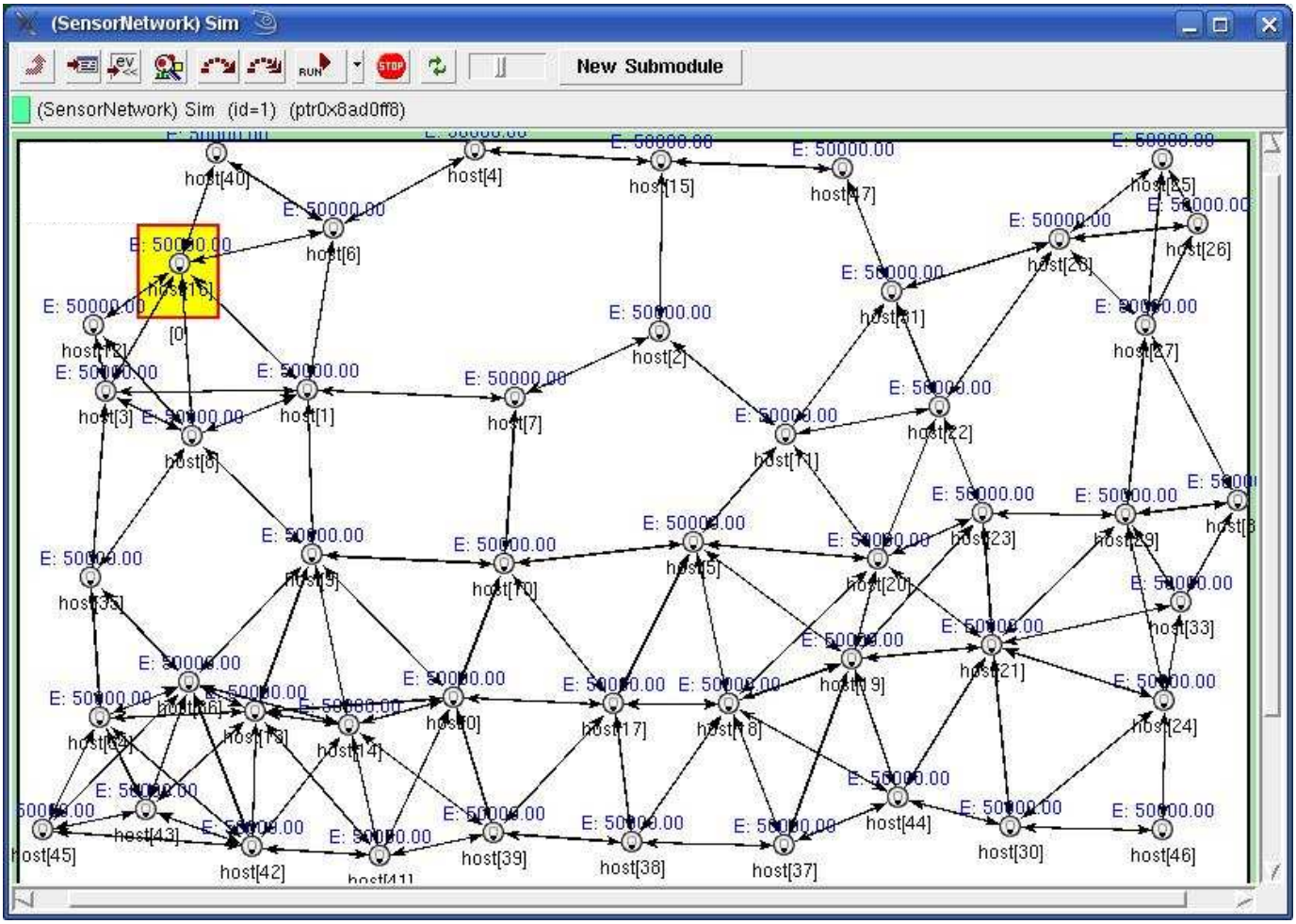


Fig. 8. The standard network with 48 nodes used in most simulations (to identify easily the nodes referred in the text; later snapshots are relatively small).

refresh rate is 5s), i.e., for 48 nodes this gives a total of 1056 times (optimum). The total number of rebroadcasts is given in Table III, with the number of missed refreshes in parenthesis (due to collision not all refreshes are received by all nodes).

From Table III one can observe that for the hccE the total number of rebroadcasts improves (near optimum) as T_w increases. For a broadcast delay of 0.4ms the number of rebroadcasts is high and therefore by using hccE strategy a T_w greather than 20ms is recommended. The distributions of rebroadcasts on each node for both the hc and hccE strategies are visualized in Figure 9 and 10, respectively.

Note that the value of T_w plays an important role for the hccE strategy, since here the metrics changes faster. Even though the number of rebroadcasts for T_w=600ms is near optimum, the energy consumption does not improve. This suggests that besides the number of rebroadcasts there are other factors that affect the energy consumption. We supposed that this can be caused by a higher number of collisions. Therefore, we measured the number of collisions and we found out that both strategies have slightly the same number of collisions for the same T_w (excepting hccE strategy with 0,4ms delay). That

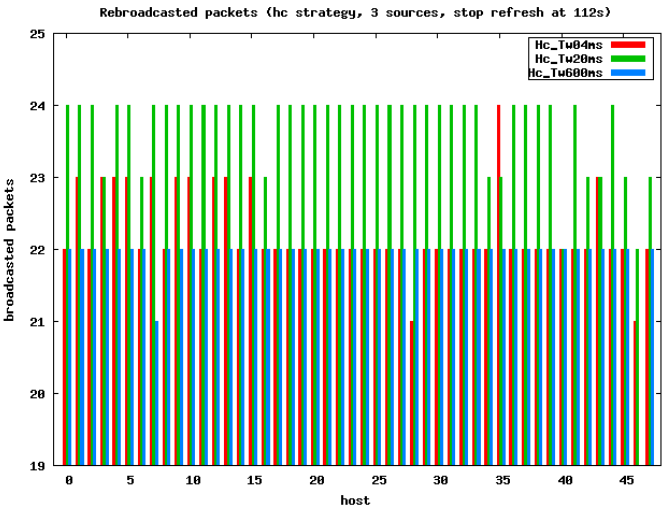


Fig. 9. Rebroadcasted interests for hc strategy.

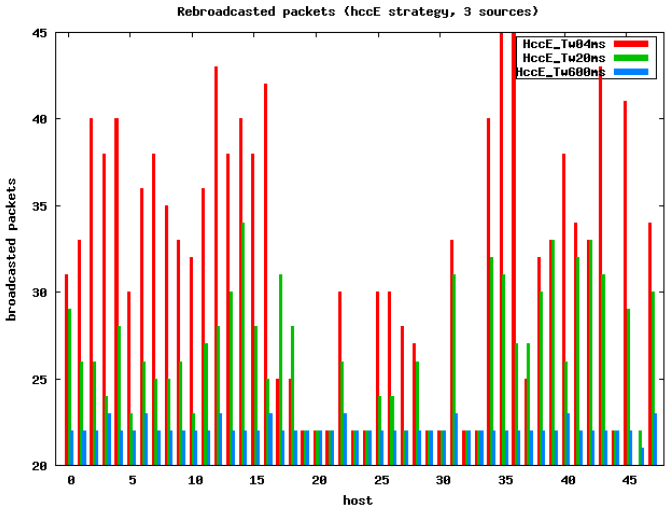


Fig. 10. Rebroadcasted interests for hccE strategy.

means that the better energy consumption for T_w=20ms is inherent to T-MAC's aggressive time-out policy. Since T-MAC extends its listen period at each send/receive event, the total time the node is in idle state is longer for a 600ms delay than for a 20ms delay.

C. Data aggregation

Since data readings are most of the time correlated, one can use in-network processing in order to reduce transmission. We consider now a simulation scenario with 3 sources (placed inside the rectangle in Figure 11), each one sending 500 packets at an interval of 200ms, with a simulation time of 2 min. By letting 3 sources to send simultaneously, the traffic load increases and each source tries to send its data packets on the shortest path (using hc strategy) or on an optimal path with the greatest critical energy (hccE strategy).

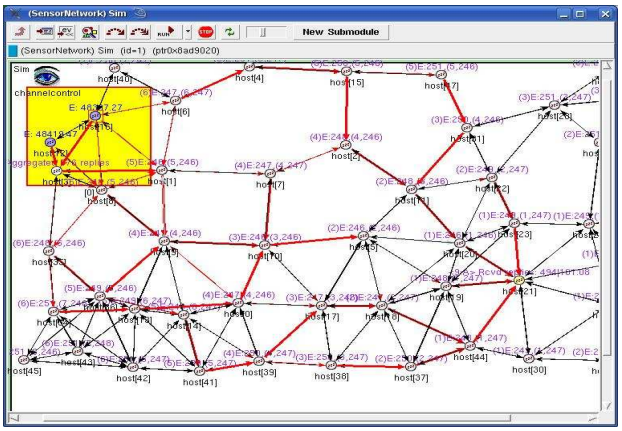


Fig. 11. Snapshot of the network with 48 nodes running strategy 5. The routes followed by the aggregated data messages are highlighted (red arrows).

For the hc and hccE strategies the energy consumption of nodes with and without aggregation (green and red curves, respectively) are given in figures 12 and 13. The energy gain is 32% for the hc strategy and 35% for the hccE strategy, respectively. One can notice that when aggregation is enabled not only nodes along the used paths consume less energy but also their neighbors.

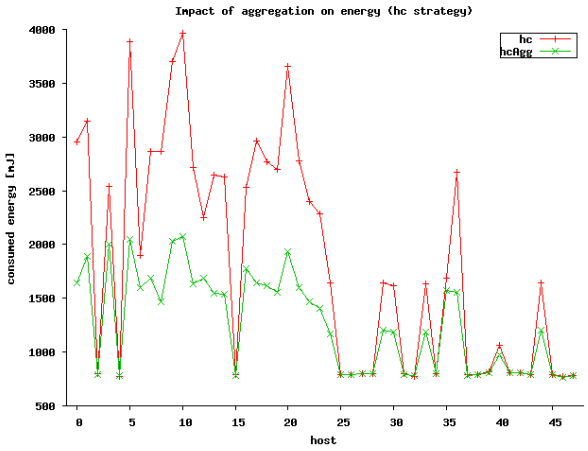


Fig. 12. hc: energy consumption of the nodes with/without aggregation.

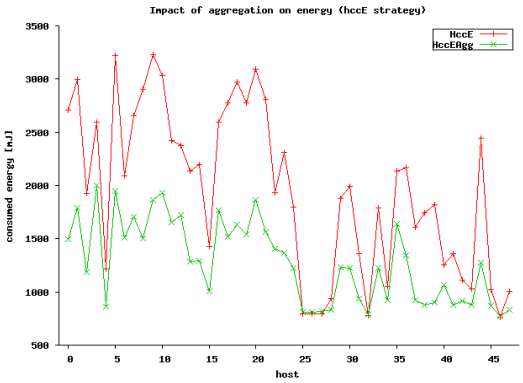


Fig. 13. hccE: energy consumption with/without aggregation.

The number of data packets sent by the sources and received by the sink reveals that no data packet (3 x 500) was lost on the way to sink (for place reasons we omit the result).

To aggregate the data messages the sources are building an aggregation tree (the aggregation algorithm is implemented by a different RU) inside the zone. The best positioned node becomes aggregator, it waits for data messages from the two sources and sends one aggregated message.

The routes selected by the hccE strategy for sending such aggregated messages are also illustrated in Figure 11. This shows how the strategy balances the packets' transmissions and the adaptivity of the routing protocol to find all possible paths between sources and sink.

D. Source-sink latency

Using the framework we can also determine the source to sink latency. In Figures 14-15 we illustrate comparatively the source to sink latency with aggregation enabled for the hc and hccE routing strategy, respectively. As expected, the source-sink latency is greater for the hccE strategy than for hc.

Since the first strategy selects also longer paths to balance the transmission load, the underlying T-MAC may need an extra active frame time (of 0,6s) until the data packets reach the sink.

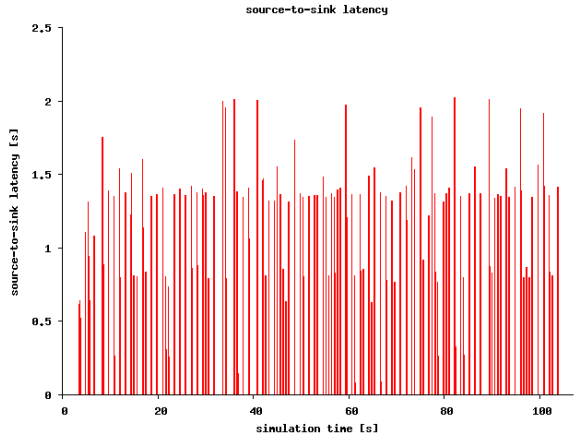


Fig. 14. hc: source to sink latency.

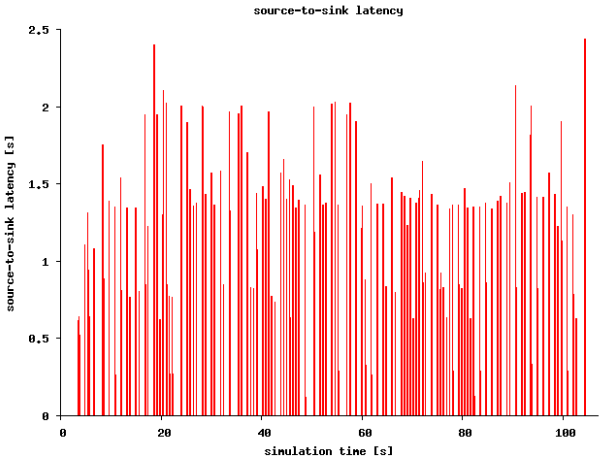


Fig. 15. hccE: Source to sink latency.

E. Latency MAC queue

Moreover, we can plot the latency for a packet in the MAC queue (namely the time between entering and leaving the queue) for each node, the number of packets in the MAC queue or (average) the number of collisions per node. For example, the latency for the aggregator (node 3) is illustrated in Figure 16.

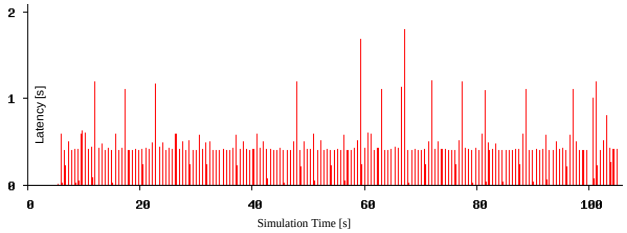


Fig. 16. Latency in MAC queue.

F. Impacts of data rate on the energy consumption

To illustrate the impact of data rate on the energy consumption in larger networks, we build another scenario with a test network consisting of 103 nodes with a distance of 9 hops (on the shortest path) between source and sink. At network layer we configure the simulator to route according to the hccE strategy. We first use T-MAC and set up three runs for different data intervals of 200ms, 500ms and 1000ms, respectively, which is equivalent to a data rate of 5 pkts/s, 2 pkts/s and 1 pkt/s. For T-MAC we set the listen time to 15ms and the frame time to 600ms. We repet the measurements for B-MAC, where the listen time is set to 50μs and the sleep time to 5ms.

Energy [mJ]	Data generation interval [ms]		
MAC Protocol	200	500	1000
T-MAC	273,48	238,08	213,67
B-MAC	264,33	220,65	197,75

TABLE IV
ENERGY CONSUMPTION USING DIFFERENT DATA GENERATION RATE.

Table IV gives the total energy consumption for both configurations with T-MAC and B-MAC.

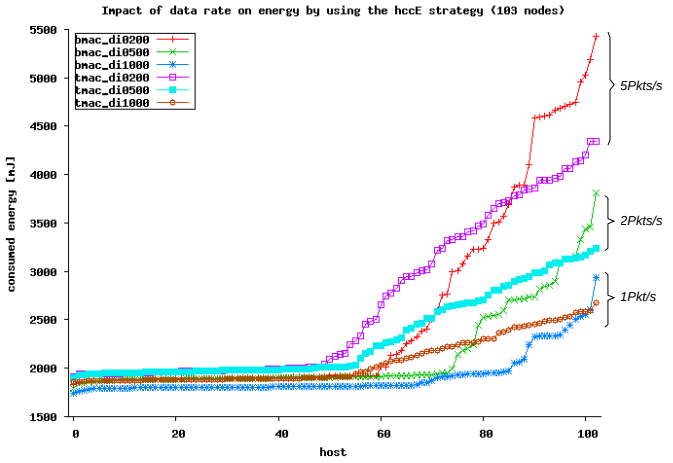


Fig. 17. Impact of using data rates on energy (hccE strategy).

Figure 17 illustrates the sorted energy consumption. If we compare the curves for a data generation interval of 200ms the energy consumption of the nodes on the path for B-MAC is still high, even though the total energy consumption is comparable with the one of T-MAC. By decreasing the data rate the consumption of the nodes on the path decreases considerably and B-MAC outperforms T-MAC.

G. Impact of unknown higher traffic depletion time

We consider the WSN as in Figure 11; additionally to sink 21 (data interval of 700ms, interest refresh is 5s) with 3 sources (right red rectangle) a new sink, node 41, was added which requests data from the zone containing node 2 (middle green rectangle) at a data interval of 500ms and sends refreshes at 4s. At MAC layer we use T-MAC protocol with listen time set to 60ms and frame time 600ms.
We set a very low initial energy for several nodes: $700mJ$ (equivalent to 3eU) for node 10 and $1000mJ$ (5eU) for nodes 0 and 7. The energy of nodes is converted in a scale between 0 and 255, which are called *energy units* (eU). Additionally we

multiply the energy consumption of the node (in all its states) with a factor of 10, to achieve a shorter simulation time. Such a situation could result from a long run of a previous interest, for example from sink 39, (or 41, 42) to sources 2, 7 and 10 (or subset of them).
The goal is to show the impact of the two strategies on the time when the nodes run out of energy (the depletion time).

By using the hc strategy, when both interests are active, the shortest route for the second interest goes through the nodes 7, 10, 0 which are depleted relatively fast by the first interest and the data packets are then obliged to travel along longer routes, e.g. 11-5-17-39-41. By using the hccE strategy the zone containg the nodes 0,7,10 is avoided. Since this is the shortest path from source 2 to sink 41 it is used for a short time at the beginning until the source gets information about its neighorhood (nodes 11 and 15). In our SNF it is easy to identify and visualize during simulation the route that a packet follows to reach the corresponding sink. Moreover, one can plot the number of data packets sent sent by the sources and the number of data packets received by the sinks in order to verify if all packets reach their destination. Under our scenario this is the case for both routing strategies.
The depletion times for the nodes configured with less energy are given in Table V.

Depletion time [s]	Nodes		
Strategy	Node 10	Node 0	Node 7
hc	40.32	61.27	74.43
hccE	55.87	76.81	82.85

TABLE V
IMPACT OF TRAFFIC AND STRATEGY ON DEPLETION TIME.

The hc strategy has preferred to route along shortest path and when the path was no longer available it used the next shortest path. On the other side, the hccE strategy has preferred to use longer paths in order to omit the nodes with lower energy reserve. Therefore the energy consumption by using the hc strategy must be smaller. It is noteworthy to remark that for this network scenario under the given settings the procentual difference between the total energy consumption of both strategies is below 2%, while the depletion time increases for the hccE strategy with a procentage between 12%-38% (achieved for the nodes 7 and 10, respectively).

The simulation and the results show that using the hccE strategy the lifetime of the nodes and thus of the network can be prolonged without significant penalties in the total energy consumption.

H. Impact of MAC

Besides comparing energy consumption of nodes, min/max latency in a node and between source and sink our framework allows also an analysis of the impact of a complete MAC protocol.

a) We investigate first the effects of different MAC protocols and their configuration parameters on the energy consumption of the nodes. Fig.18 illustrates the *sorted* energy consumption of all nodes using S-MAC and T-MAC protocols with two routing strategies, namely the hc and hccE. The energy is sorted in order to better visualize the distribution of the energy consumption on the nodes. The application requirements assume that the data interval generation is set to 200ms and the request is refreshed at a 5s interval. We use the following abbreviation, e.g., *Smac 60 600* employs S-MAC with an listen (active) time and frame time of 60ms and 600ms respectively and the hop count as default routing strategy, while *Tmac 15 600ce* uses T-MAC with an active time and frame time of 15ms and 600ms respectively and the hccE routing strategy.

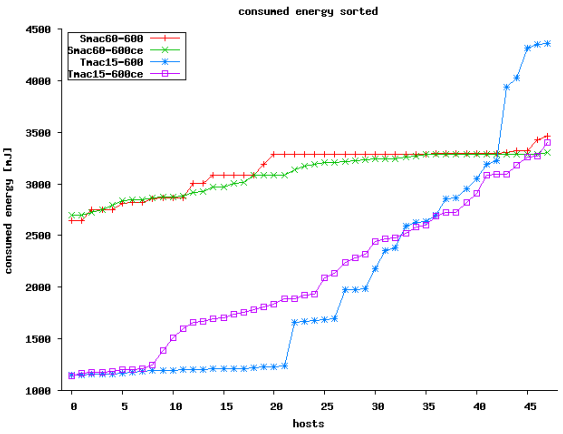


Fig. 18. Energy consumption using different protocol stacks (S-MAC and T-MAC with different parameter settings and two routing strategies).

It can be observed that in case of *Smac 60 600* (with the hc strategy) and *Smac 60 600ce* (with hccE), the influence of the strategy on the energy consumption is minimal, since the data traffic is relatively small and all the nodes are most of the time in idle listening and are consuming almost the same amount of energy. However, when employing *Tmac 15 600*, the choice between the hc and hccE strategy has a relevant impact on the energy consumption, the balancing policy of hccE is reflected by the smoother *Tmac15-600ce* curve than the *Tmac15-600* one for the hc strategy.

We chose here for S-MAC a higher active time than for T-MAC (60ms vs. 15ms) since for higher data rates S-MAC collapses (fails to deliver) than T-MAC as we will see next.

b) In the sequel we investigate the limits of the MAC protocols. The goal of this analyze is to find out when the MAC protocol is overloaded and is not able to deliver successfully data messages. We consider the MAC protocol overloaded if it discards more than 10% of the data messages. The broadcast messages (interest and its refreshes) are not taken into account, since we consider them not relevant for the application. The impact of different data rates on the behavior of the MAC (with a bounded queue) can be illustrated in Figure 19 by counting the total number of dropped frames.

Reasons to discard frames are either that the MAC queue is full or a transmission failure occurs (the maximal retries threshold to send the same frame was hit). One can observe that all MAC protocols have a point in the graphic from where the number of discarded frames increases steeply. If we consider the 10% limit as the point from where the MAC protocol is considered overloaded (unreliable), one can observe

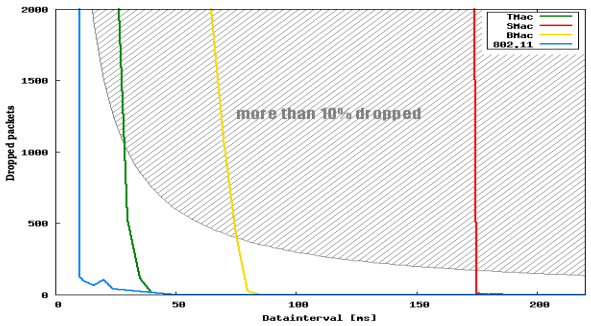


Fig. 19. Impact of the data generation interval on different MAC protocols.

that S-MAC hits this limit at a data generation interval of 180 ms (aprox. 5 frames/s), B-MAC at 75 ms (aprox. 13 frames/s) and T-MAC at 30 ms (aprox. 33 frames/s). Opposite to them, the WLAN 802.11 is characterized by a high efficiency, since even at 10 ms (aprox. 100 frames/s) it discards only few frames.

c) We study next what influence has the MAC protocol and the data generation interval on the source to sink latency. For time-critical sensor applications this behavior is an important aspect in choosing the appropriate MAC protocol. In order to analyze the latency we have taken 20 measurements with randomly placed sink and a constant distance of six hops between the source and sink. For different data generation intervals the results are illustrated in Figure 20.

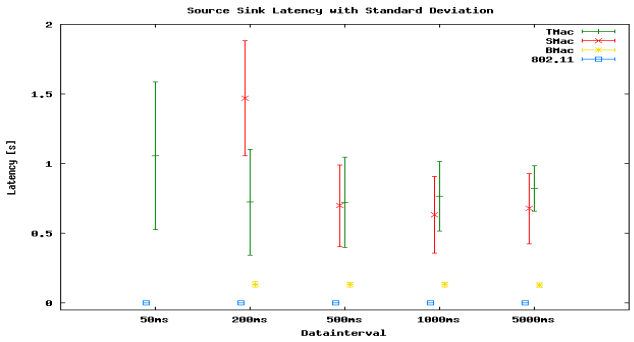


Fig. 20. Source-to-sink latency for different data rates and MACs.

In case of T-MAC the source to sink latency is relatively constant, increasing slightly at higher data rate (for 50ms), since the protocol nears to its collapse limit. The relatively high standard deviation shows that the number of active cycles that a packet needs to reach the sink is variable.
S-MAC experiences a relatively high latency (aprox. 1.5s) beginning with a data interval of 200ms, and collapses at a data interval of 50 ms (the protocol reaches its limits[5] see Figure 19). At lower data rate the latency is almost constant in the range of a frame time (0.6s). The number of hops that a packet travels per listen time is approximately fix and depends on the actual setting of this time. If the path has more hops than a packet can travel per listen time, it is cached in the MAC queue and waits the next listen period. That leads to a latency with a variation of one listen period.
B-MAC has a low latency without variation. This is due to its very low duty cycle compared to the one of S-MAC and T-MAC. Usually a packet needs here one (duty) cycle pro hop. The WLAN 802.11 has very low latency, without variation, since the protocol has no sleep state.

[5]Appropriate configuration of the listen period allows a correct function at 50 ms, but then the energy consumption increases

d) Finally, the dependency of the source-sink latency on the number of hops between the source and sink is illustrated in Figure 21. For this simulation the data generation interval is set to 500 ms and we take 20 measurements with a variable distance between source and sink in the range from 1 to 10.

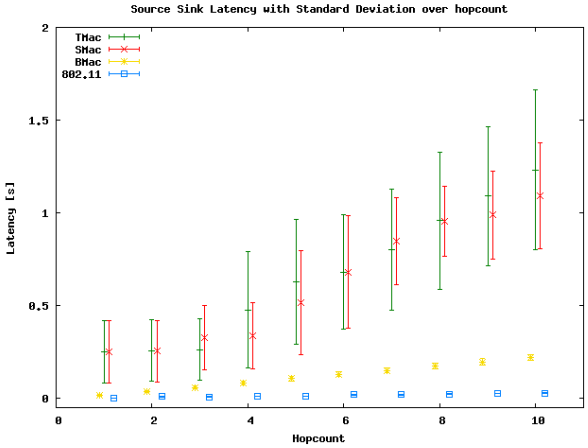


Fig. 21. Impact of the number of hops on the source-sink latency.

T-MAC and S-MAC have a similar behavior, they forward a packet 2-3 hops in one listen period and correspondingly the source-sink latency is small. After that the latency increases constantly with each new hop. Notice the slightly smaller latency (and variation) of S-MAC compared to T-MAC. The cause is the fact that at this lower data rate the listen time of T-MAC is seldom extended and therefore packets must wait a sleep period until the next listen period. In contrast, S-MAC is able to forward the packets more hops during its fixed listen period. B-MAC has a small, slightly increasing latency, while WLAN 802.11 has a very small latency even at higher hops.

As a concluding remark for the simulations involving the two phase multihop protocol we can state that both the choice of the MAC protocol and the choice of routing strategy with/without aggregation influence the energy consumption, and thus the network lifetime, significantly.

The same simulations and measurements can be carried over for the **enhanced directed diffusion (EDD)** protocol. Instead of presenting similar results here we choose to exemplify here the advantages of our decomposed, modular design of this complex routing protocol.

I. EDD scenario

To illustrate the exchange of a complete routing unit (RU) we implemented EDD in 5 RUs, each of them implementing one phase of the protocol, namely: interest propagation, path establishment, reinforcement, data event gathering and an additional RU to control the aggregation.

Here we want to illustrate the impact on energy consumption resulted by exchanging different RUs and SUs. We will exchanged, for example, for the first phase the flooding RU with a geographic flooding RU and for the third phase the strategy used at reinforcement. Instead of reinforcing the fastest neighbor that delivered the data event we chose the neighbor with the most residual energy able to deliver it.

For our simulation we use a more dense network (with more than 50 nodes, initially with a fixed residual energy of 2500 mJ), with node 0 as sink, placed in a central position of the network. The sink injects an interest requesting a data generation at 1s (refresh it at each 10s) in a zone with 8 sources, placed in the right side of the network as illustrated in Figure 22.

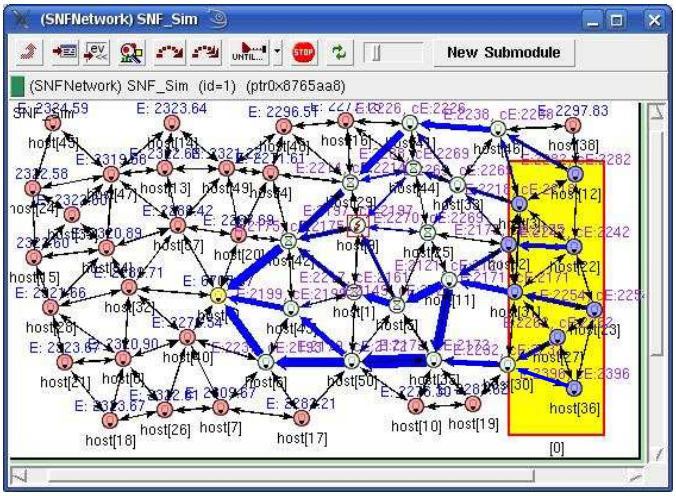


Fig. 22. Flooding and data flow routes at the end of simulation.

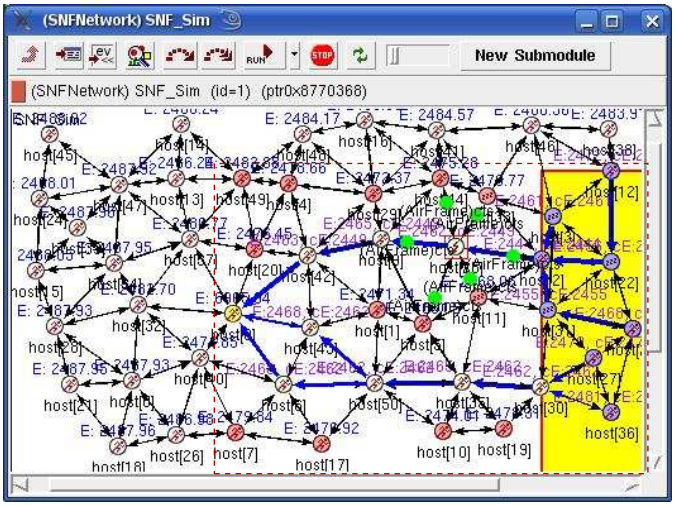


Fig. 23. Geographic flooding.

The interest is flooded in the whole network (each node that hears the interest is colored with light red). Figure 22 also shows the preferred paths (see the thick blue arrows[6]) that data packets have used to reach the sink. One can also remark that during the simulation several paths have been reinforced (the white nodes) due to energy consideration reasons.

We changed now the configuration by replacing the flooding used in the interest propagation phase with a geographic flooding unit (i.e., we just exchanged the corresponding RU). Since the network is dense enough the forwarding zone for the interest (and its refreshes) is restricted to the dotted rectangle determined by the sink and the destination zone (Figure 23).

As a result of our energy-aware strategy for the gradient reinforcement phase we obtain different data paths, namely the thick blue ones. Note that the data traffic is reduced, since the source 2 aggregates now the source nodes 3, 22 and 31, the last two aggregates source 12 and 33, respectively. Nodes 27 and 36 send as before their data directly.

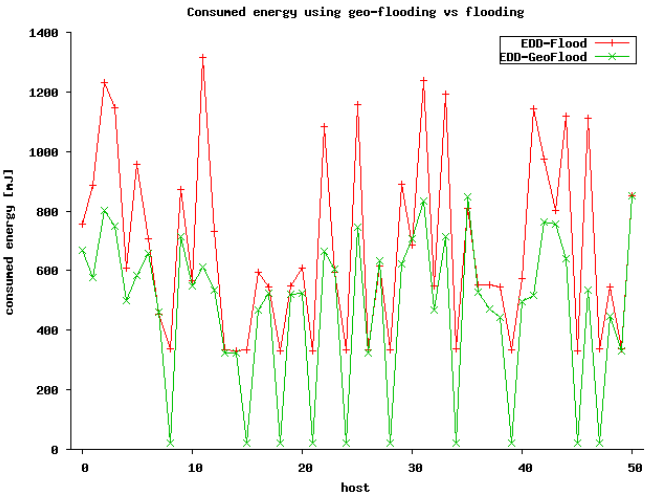


Fig. 24. Energy consumption by using flooding and geographic flooding.

The energy consumption for a 2 minutes run is plotted in Figure 24, where green stands for geographical flooding and red for flooding. Note the significant decrease in energy consumption in case of geographic flooding and also the fact that a lot of nodes have almost no energy consumption (the ones outside the forwarding region).

Of course, one can run simulations for enhanced directed diffusion in combination with different MAC protocols and illustrate all kinds of performance criteria (energy efficiency, latency, depletion time, collisions, etc) like in case of two phase routing protocol.

J. Simulator reconfiguration

The overhead required to reconfigure the simulator (in order to combine different building blocks of protocols) is small. For example, in order to modify a routing strategy we only need to

[6]the thickness of a link is according to the number of data packets that used that link

edit the strategy name. To activate the aggregation process one needs to edit the aggregation RU name and to set a flag. To exchange a complete MAC protocol one has to edit two lines in the configuration file of the node to include the new submodule and to import the corresponding code. Protocol's parameters (e.g., active and frame times) have corresponding parameters in the configuration file which can be edited. A recompilation step for the whole code is not necessary (assuming that all MAC protocols have been compiled) in order to start the simulation and to visualize the results.

VI. Conclusion and future work

In this paper we investigated several factors that impact on the performance of routing protocols used in resource constrained wireless sensor networks. The main performance criteria we are interested in are the energy consumption, the network lifetime and also the latency of the network in delivering replies to users requests. We analyzed the impact of factors such as various routing strategies, different MAC protocols and their configuration parameters, link and node failures, changes in the network topology, in-network processing, and fluctuating traffic. A node's behavior in terms of energy consumption is difficult to be exactly predicted, since it depends on a large amount of parameters, which may have adverse or unexpected effects, and thus also its impact on the evolution and performance of the entire network. Hence, an adequate modeling and simulation framework was needed in order to achieve a fine tuning of all these parameters and to better inspect the cumulative impact of their behavior on the sensor network.

To that aim we employed our SNF, a flexible tool to design and combine various protocols at application, network and MAC layers and to analyze the impact of different routing strategies and factors mainly on the energy consumption of the WSN. The framework can automatically visualize various performance criteria to enable a fast evaluation and comparison of protocols.

For routing protocols we proposed and compared several energy-aware routing metrics (§IV) by employing local and more global information concerning the residual energy of nodes. The main challenge here is to decide what is the relevant routing information that should be spread to the nodes in the network without sending explicit routing messages, in order to balance the load of forwarding the data packets on all the nodes by using different routing strategies. We exemplified this by describing two concrete protocols: a two phase multihop routing and enhanced, energy-aware directed diffusion.

We showed that an energy-aware routing can significantly contribute to better balance the communication load among nodes, and thus to prolong the network lifetime with minor penalties in the total energy consumption. Local neighborhood knowledge turned out to be insufficient to achieve this.

Aggregation is very useful to reduce the energy consumption of the nodes on the path and, additionally, it can reduce the traffic in the network avoiding in this way congestions and induced collisions (§V-C and §V-I).

Delaying the request (interest) broadcast is recommended to optimize further the energy consumption, but finding an optimal value is not trivial, especially when the underlying MAC protocol does not have a fixed schedule (§V-B). In such cases a closer exchange of information between the MAC protocol and the routing protocol is required. This can be accomplished by using the cross-layer component.

The simulation results have shown that the choice of the MAC protocol, especially its duty cycle, has a major impact on the energy consumption in the network. Thus, whenever the application requirements are known it is essential to select the MAC protocol appropriately. Moreover, the choice of the combination of the MAC and routing protocol influences the behavior of the network in terms of energy. An interesting observation is that the adaptive characteristic of the strategies combined with topology knowledge can be exploited when the MAC protocol is based on a preamble sampling scheme (like B-MAC) (see §V-F).

More important is the fact that the main impact on the energy consumption of the nodes is given by the MAC protocol and only secondary by the routing protocol.

Currently we provide implementations for different routing protocols and several low duty-cyle MAC protocols (S-MAC, T-MAC, Preamble Sampling), including support for collision detection and radio switch times).

As future work we intend to quantify the programming overhead needed to develop and integrate new protocols. Furthermore, since the MAC protocol has the main impact on the energy consumption, we intend to provide more MAC protocol implementations and to compare their performance. Carrying out comparative analysis between different MAC protocols and their interaction with network layer protocols will reveal surely other promising aspects that can bring optimization at both layers.

Additionaly, we strive for more modularity at MAC layer, mainly to embed at MAC layer more customizable services like the receiver-based contention (or other innovative ideas from new MAC protocols), which in our opinion gives another perspective to the interlayer communication and would improve the energy efficiency of routing.

References

[1] A. Kacsó and R. Wismüller, "Modeling and simulation of multihop routing protocols in wireless sensor networks," in *Proc. 5th Int. Conf. on Wireless and Mobile Communications (ICWMC'09)*. Cannes, France, August 2009, pp. 296–302.

[2] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion a scalable and robust communication paradigm for sensor networks," in *Proc. ACM MobiCom*. Boston, 2000, pp. 56–67.

[3] F. Ye, A. Chen, S. Lu, and L. Zhang, "A scalable solution to minimum cost forwarding in large sensor networks," in *Proc. 10th Int. Conf. on Comp. Comm. and Networks*. Arizona, 2001, pp. 304–309.

[4] F. Ye, G. Zhong, S. Lu, and L. Zhang, "Gradient broadcast: A robust data delivery protocol for large scale sensor networks," *Wireless Networks/Springer, The Netherlands*, vol. 11, no. 2, pp. 285–298, 2005.

[5] M. Busse, T. Hänselmann, and W. Effelsberg, "Energy-efficient forwarding schemes for wireless sensor networks," in *Proc. Int. Symp. on WoWMoM*. New York, USA, June 2006, pp. 125–133.

[6] C. Schurgers and M. Srivastava, "Energy efficient routing in wireless sensor networks," in *Proc. MILCOM on Comm. for Network-Centric Operations: Creating the Inform. Force*. Virginia, 2001, pp. 357–361.

[7] H. Nurul1, M. Hossain, S. Yamada, E. Kamioka, and O.-S. Chae, "Cost-effective lifetime prediction based routing protocol for manet," in *Proc. of ICOIN*. Springer-Verlag Berlin, 2005, pp. 170–177.

[8] M. Maleki, K. Dantu, and M. Pedram, "Lifetime prediction routing in mobile ad-hoc netowrks," *Proc. IEEE WCNC*, vol. 2, pp. 1185–1190, March 2003.

[9] J. Aslam, Q. Li, and D. Rus, "Three power-aware routing algorithms for sensor networks," *Wireless Comm. and Mob. Computing*, vol. 3, no. 2, pp. 187–208, 2003.

[10] E. Shih, S. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan, "Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks," in *Proc. 7th Ann. Int. Conf. on Mob. Comp. and Netw.* Rome, Italy, July 2001, pp. 272–287.

[11] Y. Yu, R. Govindan, and D. Estrin, "Geographical and energy aware routing: a recursive data dissemination protocol for wsns," in *UCLA/CSD-TR-01-0023s*. LA, California, USA, May 2001, pp. 171–180.

[12] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*. John Wiley & Sons, 2005.

[13] G. Halkes, T. Dam, and K. Langendoen, "Comparing energy-saving mac protocols for wireless sensor networks," *Mob. Netw. Appl.*, vol. 10, no. 5, pp. 783–791, 2005.

[14] M. Zuniga and B. Krishnamachari, "Analyzing the transitional region in low power wireless links," in *Proc. IEEE SECON*. Santa Clara, CA, October 2004.

[15] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly-resilient, energy-efficient multipath routing in wireless sensor networks," in *Proc. 2-nd ACM Int. Symp. on Mobile ad hoc networking and computing*. Long Beach, CA, USA, October 2001.

[16] V. Bharghavan, A. Demers, S.Shenker, and L. Zhang, "Macaw: A media access protocol for wireless lans," in *Proc. of SIGCOMM Conf.* London, UK, September 1994, pp. 212–225.

[17] A. Kacsó and R. Wismüller, "A framework architecture to simulate energy-aware routing protocols in wireless sensor networks," in *Proc. IASTED Int. Conf. on Sensor Networks*. Greece, 2008, pp. 77–82.

[18] A. Varga, *User Manual*. OMNeT++ Version 3.2, 2006.

[19] M. Löbbers and D. Willkomm, *Mobility Framework for OMNeT++ (API ref.)*. http://mobility-fw.sourceforge.net: OMNeT++ Ver.3.2, 2006.

[20] A. Kacsó and R. Wismüller, "A simulation framework for energy-aware wireless sensor network protocols," in *Proc. 18th Int. Conf. on Computer Communications and Networks, Workshop on Sensor Networks*. San Francisco, CA, USA, August 2009.

[21] J. Polastre, J. Hui, P. Levis, J. Zhao, D. Culler, S. Shenker, and I. Stoica, "A unifying link abstraction for wireless sensor networks," in *Proc. 3rd ACM Int. Conf. SenSys*, November 2005, pp. 76–89.

[22] C. Ee, R. Fonseca, S. Kim, D. Moon, A. Tavakoli, D. Culler, S. Shenker, and I. Stoica, "A modular network layer for sensornets," in *Proc. 7th Symp. OSDI*. Seattle, WA, USA, 2006, pp. 249–262.

[23] I. Akyildiz, M. Vuran, and O.Aka, "A cross-layer protocol for wireless sensor networks," in *Proc. CISS*. Princeton, NJ, March 2006.

[24] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated, adaptive sleeping for wireless sensor networks," *IEEE/ACM Trans. on Netw.*, vol. 12, no. 3, pp. 493–506, 2004.

[25] T. Dam and K.Langendoen, "An adaptive energy- efficient mac protocol for wireless sensor networks," in *Proc. 1st Int. Conf. on Embedded Networked SenSys*. LA, California, USA, 2003, pp. 171–180.

[26] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proc. 2nd ACM Int. Conf. on Embedded Networked SenSys*. NY, USA, November 2004, pp. 95–107.

TeraPaths: End-to-End Network Resource Scheduling in High-Impact Network Domains

Dimitrios Katramatos[1], Xin Liu[1], Kunal Shroff[2], Dantong Yu[1], Shawn McKee[3], Thomas Robertazzi[4]

[1] Computational Science Center, Brookhaven National Laboratory, Upton, NY 11973, USA
[2] National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973, USA
{dkat, xinliu, shroffk, dtyu}@bnl.gov
[3] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
smckee@umich.edu
[4] Department of Electircal and Computer Engineering, Stony Brook University, Stony Brook , NY 11794, USA
tom@ece.sunysb.edu

***Abstract*— The TeraPaths project at Brookhaven National Laboratory is pioneering a framework that enables the scheduling of network resources in the context of data-intensive scientific computing. Modern wide area networks, such as ESnet and Internet2, have recently started providing network resource reservation capabilities in the form of virtual circuits. The TeraPaths framework utilizes these circuits and extends them into end-site local area networks, establishing end-to-end virtual paths between end-site hosts. These paths are dedicated to specific users and/or applications and provide guaranteed resources, minimizing or eliminating the adverse effects of network congestion. In this article, we present an overview of TeraPaths and examine issues raised by the end-to-end resource reservation-based networking paradigm as well as implications and benefits for end users and applications. We also discuss scalability issues and optimization techniques for wide area network circuit reservations.**

Keywords—End-to-end QoS networking, hybrid networks, network virtualization, virtual circuit reservation optimization.

I. INTRODUCTION

This article is an extended and revised version of the INTERNET 2009 conference paper entitled: "Establishment and Management of Virtual End-to-End QoS Paths Through Modern Hybrid WANs with TeraPaths" [1].

Modern data intensive scientific applications, including high energy and nuclear physics, astrophysics, climate modeling, nanoscale materials science, and genomics, will soon be capable of generating data on the order of exabytes per year [2]. This data must be transferred, visualized, and analyzed by geographically distributed teams of scientists, imposing unprecedented demands on computing and especially networking resources. While such applications can capitalize on modern high-performance networking capabilities, they can also be critically sensitive to the adverse effects of unpredictably occurring network congestion. Because network capacity is finite, competition among data flows may cause applications to suffer severe performance degradation and eventual disruption. When data delivery must conform to specific deadlines or application components need to interact in real time, the standard best-effort networking model may not always be sufficient. To work effectively, these applications may require resource availability guarantees. In the case of network, the requirement primarily translates to bandwidth guarantees, however, other Quality of Service (QoS) parameters may also be included, i.e., delay, jitter, etc. The Department of Energy (DOE) Office of Science identifies QoS as one the five top ranked issues essential to the success of distributed science [3].

The next section discusses the motivation behind TeraPaths, while section 3 describes two key projects that constitute the framework for the advance resource reservation model. Section 4 focuses on the differences between the two kinds of dedicated network paths through WAN domains supported by this framework, while Section 5 presents techniques necessary for the effective utilization of these dedicated WAN paths. Section 6 examines fault tolerance issues and Section 7 discusses related work. Finally, Section 8 presents our conclusions and future work directions.

II. HIGH-IMPACT NETWORK DOMAINS

As noted in the title, TeraPaths targets "High-Impact" network domains (sets of related users and systems connected by networks) and so we provide some background on what we mean by this. Typical network use for a given system characteristically utilizes a few-to-many, small bandwidth, short duration network flows: email, web browsing, and the occasional file-transfer are common examples. However, there is a much smaller set of systems which regularly transfer large amounts of data over the network. Typically, this may involve bandwidth-intensive applications or large files (data, movies, games, HD video-conferencing, etc.) and may use a significant fraction of the available bandwidth along a network path. More importantly, some of these large flows may have additional requirements regarding packet loss, delay, and jitter, as well as overall deadline scheduling needs that are critical to the specific user or application. We characterize high-impact domains as those sets of users and systems who need to transfer large amounts of data through the network and who may require additional control over network related characteristics of their critical flows (such as "real-time or interactive flows", e.g., video-conferencing, real-time

instrument control, conference audio/visual streaming, etc.).

The high-impact domains TeraPaths envisions supporting are in the e-Science area where significant amounts of data need to be shared across wide-area networks (WANs) and additional important considerations regarding timeliness of some data transfers and their corresponding flow characteristics are important to the success of the applications involved [4]. In particular, grid-computing infrastructures in science are already broadly deployed and could be considered synonymous with high-impact domains. Virtual organizations built upon grids would significantly benefit from end-to-end predictability of network paths interconnecting their shared resources [5]. While small in number (by relative count of users or end-sites), these domains can have a disproportionally disruptive effect on the network and thus are "high-impact".

We would further make the case that not all large-scale flows are of equal importance or criticality. On today's Research and Education networks one may see large scale flows corresponding to high-energy physics data transfers, eVLBI astronomy, bio-informatics and life sciences as well as peer-to-peer traffic sharing movies, applications, music, and other multimedia content. Even within a networked collaboration of users, some large scale transfers may have significantly different importance but are currently treated equivalently by the best effort network. Part of the motivation behind TeraPaths is to give researchers the tools they need to most effectively utilize the resources they have access to.

III. BACKGROUND

Several available networking technologies, such as the Differentiated Services (DiffServ) [6], Integrated Services (IntServ) [7], Multi-Protocol Label Switching (MPLS) [8], and Generalized MPLS (GMPLS) [9] architectures, have the capability to address the issue of providing resource guarantees. In practice, however, the scope of network connections utilized by distributed applications spans multiple autonomous domains. These domains typically have different levels of heterogeneity in administrative policies and control plane and data plane technologies, making it difficult or impossible to provide network QoS guarantees using a single architecture across all domains. For example, Differentiated Services Code Point (DSCP) packet markings, used in the DiffServ architecture, are by default reset at ingress points of network domains. As such, the DiffServ architecture is ineffective across domains without prior inter-domain Service Level Agreements (SLAs) in effect and proper configuration of involved network devices.

Recent networking research and development efforts [10] – [13] adopt a hybrid solution to the problem, with individual network segments utilizing different underlying technologies. From the end user perspective, however, these technologies are seamlessly tied together to ensure end-to-end resource allocation guarantees. This hybrid solution creates a new networking model that transparently co-exists but fundamentally differs from the standard best-effort model. Under the new model, it is possible to allocate network resources through advance reservations and dedicate these resources to specific data flows. Each such flow (or flow group) is steered into its "own" virtual network path, which ensures that the flow will receive a pre-determined level of QoS in terms of bandwidth and/or other parameters. Virtual paths can comprise several physical network segments and span multiple administrative domains. These domains need to coordinate to establish the virtual path. Coordination takes place by means of interoperating web services. Each domain exposes a set of web services that enable the reservation of resources within a domain's network. Authorized users of these services, which can be another domain's services, can reserve network resources within the domain and associate them with specific data flows. When reservations activate across all domains between a flow's source and destination, a dedicated end-to-end virtual path spanning these domains is assembled. This path offers to the flow of interest a predetermined level of end-to-end QoS. The coordination of multiple network domains through web services is essentially a loosely coupled Service Oriented Architecture (SOA) for the network control plane, a network "service plane" [14].

End-to-end virtual paths can be viewed as consisting of three main segments: two end segments, one within each end site Local Area Network (LAN), and a middle segment spanning one or more Wide Area Network (WAN) domains. In this article, we consider the establishment of end-to-end virtual paths from the perspective of end sites. User applications run on end site systems, communicate with the rest of the world through end site LANs, and are subject to end site administrative policies. In the standard networking model, traffic through the WAN is subject to pre-existing SLAs between adjacent network domains. In the new advance resource reservation model, such SLAs are essentially dynamic, allowing end sites to utilize and – indirectly – manage WAN capabilities in a way that maximizes the benefit to the end user.

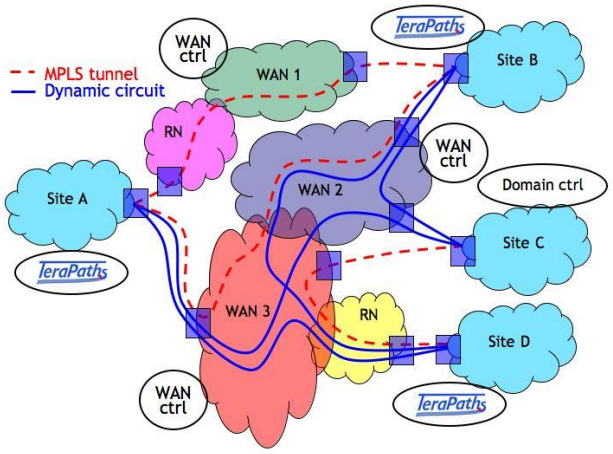


Figure 1. The framework for establishing end-to-end paths; TeraPaths-controlled sites are interconnected with WAN MPLS tunnels and/or dynamic circuits; some paths pass through regional networks that have long-term static configurations to accommodate QoS.

The framework for establishing end-to-end QoS-aware network paths encompasses web service-based systems that properly configure end site LAN and WAN domains (see Figure 1). The capability for advance resource reservation is currently available between sites interconnected through the ESnet [15] and Internet2 [16] networks. In this section we give background information on the two projects that constitute this framework, the TeraPaths project and the OSCARS project.

A. *The TeraPaths Project*

The DOE-funded TeraPaths project [10] at Brookhaven National Laboratory (BNL) combines DiffServ-based LAN QoS with WAN MPLS tunnels and dynamic circuits to establish end-to-end (host-to-host) virtual paths with QoS guarantees. These virtual paths prioritize, protect, and regulate network flows in accordance with site agreements and user requests, and prevent the disruptive effects that conventional network flows can bring to one another.

Providing an end-to-end virtual network path with QoS guarantees (e.g., guaranteed bandwidth) to a specific data flow requires the timely configuration of all network devices along the route between a given source and a given destination. In the general case, such a route passes through multiple administrative domains and there is no single control center able to perform the configuration of all devices involved. The TeraPaths system has a fully distributed, layered architecture (see Figure 2) and interacts with the network with the perspective of end-sites of communities. The local network of each participating end-site is under the control of an End-Site Domain Controller module (ESDC). The site's network devices are under the control of one or more Network Device Controller modules (NDCs). NDCs play the role of a "virtual network engineer" in the sense that they securely expose a very specific set of device configuration commands to the ESDC module. The software is organized so that NDCs can be, if so required by tight security regulations, completely independently installed, configured, and maintained.

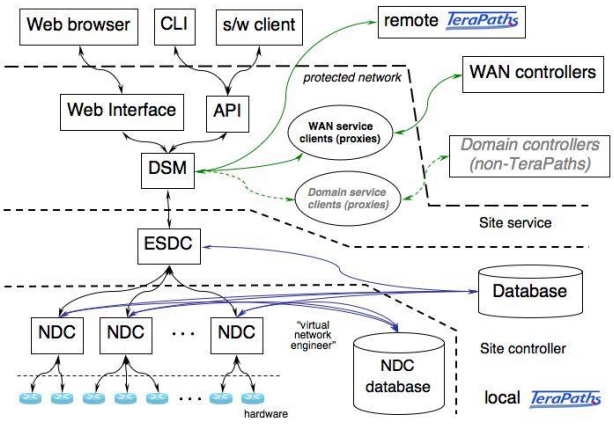


Figure 2. The software architecture of TeraPaths. Services of remote network domains are invoked through "proxy" server modules.

An NDC encapsulates specific functionality of a network device and abstracts this functionality through a uniform interface while hiding the complexity of the actual configuration of heterogeneous hardware from higher software layers. A site's ESDC and NDC(s) are complemented by a Distributed Services Module (DSM), which is the core of the TeraPaths service. The DSM has the role of coordinating all network domains along the route between two end hosts (each host belonging to a different end-site) to timely enable the necessary segments and establish an end-to-end path. The DSM interfaces with all ESDCs (local and remote) to configure the path, starting within the end-site LANs (direct control) and proceeding to arrange the necessary path segments through WAN domains (indirect control). To interface with non-TeraPaths domain controllers, primarily for WAN domains but also for end-sites that are using other controlling software (e.g., Lambda Station [11]), the DSM uses auxiliary modules that encapsulate the functionality of the targeted domain controller by invoking the required API but exposing a standardized abstract interface. As such, these auxiliary modules appear to a DSM as a set of "proxy" WAN or end-site services with a uniform interface. It should be noted that the responsibility of selecting and engineering the path within a WAN domain belongs to the controlling system of that domain. TeraPaths can only indirectly affect such a path by providing preferences to the WAN controlling system, if that system offers such a capability.

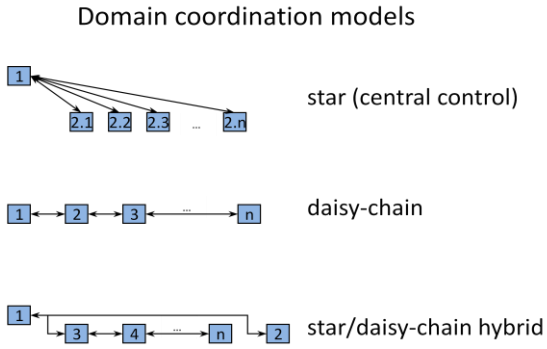


Figure 3. Coordination models. Each square represents a site's controller.

Currently, TeraPaths follows a hybrid star/daisy chain coordination model where the initiating end-site first coordinates with the target site and then indirectly sets up a WAN path by contacting its primary WAN provider and relying on that provider's domain to coordinate, if necessary, with other WAN domains along the desired route (see Figure 3). The hybrid coordination model was adopted as the most feasible since end-site and WAN systems need only to interface/coordinate. Thus, no unified communication protocol is required, as in the case of the daisy chain model, and there is no centralization of control, as in the case of the star model. The hybrid model essentially splits the network in two large segments: the end-sites and the WAN domains,

with each segment coordinating with the other to setup a path.

The result of the domain coordination process is the establishment of dynamic Service Level Agreements (SLAs) between all network domains along an end-to-end path. TeraPaths is responsible for the two end-sites and OSCARS for one or more peering WAN domains. The Message Sequence Chart (MSC) in Figure 4 shows the messaging sequence taking place in the current system implementation: initiating end-site A negotiates with the other end-site B to reach a consensus based on the resource availability of both sites. Then, site A send the negotiated request to the WAN domain manager, in this case, OSCARS, which responds with a success or failure message.

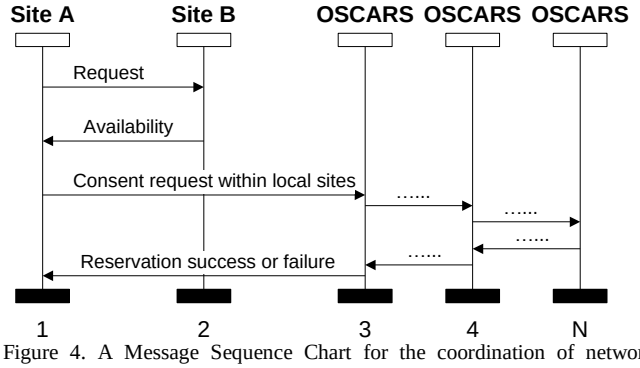


Figure 4. A Message Sequence Chart for the coordination of network domains controlled by TeraPaths and OSCARS.

B. OSCARS

The DOE-funded On-demand Secure Circuit Advance Reservation System (OSCARS) [13] is a project initiated by ESnet. Initially, OSCARS could dynamically provision secure layer-3 (L3) circuits with guaranteed bandwidth in the form of MPLS tunnels, only within the ESnet domain.

Through collaboration between ESnet and Internet2, OSCARS evolved into a more general Inter-Domain Controller (IDC), a WAN domain controller, enabling adjacent WAN domains to interoperate and establish secure circuits spanning multiple domains via the use of a special protocol specifically developed for domain interoperation. While still capable of providing MPLS tunnels within ESnet, OSCARS can additionally provide guaranteed bandwidth layer-2 (L2) circuits within and between ESnet's Science Data Network (SDN) and Internet2's Dynamic Circuit Network (DCN). SDN and DCN are interconnected at New York and Chicago and bring together DOE laboratories and Universities across the United States.

Access to OSCARS circuit reservations is offered via a web interface. Additionally, the system's functionality is exposed through a web services API for automatic invocation from programs. The API includes basic primitives for establishing and managing circuit reservations (create, cancel, query, list) and L2-specific primitives to signal and teardown dynamic circuits. TeraPaths utilizes a client module to automatically submit circuit reservation requests and further manage these reservations on behalf of end site users/applications. The selection of the actual WAN path is currently left at the discretion of OSCARS for simplicity and maximum flexibility in satisfying a request. The path provisioned by an OSCARS reservation is expected to satisfy the bandwidth requirements, however, the end-sites do not participate in routing decisions. The latest versions of OSCARS include support for obtaining topology information and specifying preferred path in reservation request. Selecting inter-domain paths is desirable from the end-site perspective for reserving, e.g., lower latency routes. However, it adds another dimension of complexity to reserving a path, as end-sites need to pull topology information and decide on which route they prefer based on certain criteria, while the chances of successfully reserving a path are probably decreasing as OSCARS is presented with a less flexible request. Nevertheless, we plan to explore such capabilities in our future work.

C. The TeraPaths Testbed

The TeraPaths project utilizes a multiple-site testbed for research, software development, and testing. Currently, the testbed encompasses subnets at three sites, BNL, University of Michigan (UMich) and Boston University (BU) (see Figure 5). Each site runs its own instance of the TeraPaths service.

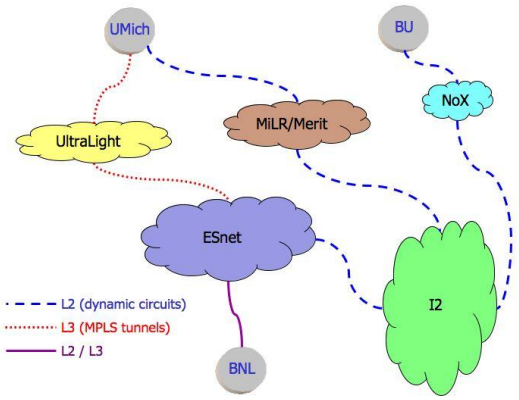


Figure 5. The TeraPaths testbed encompasses subnets at BNL, UMich, and BU. Only BNL is directly connected to ESnet.

All instances can interface with OSCARS interdomain controllers to setup MPLS tunnels through ESnet and dynamic circuits through ESnet and Internet2. Future end-sites will have similar interconnecting capabilities depending on which WAN they subscribe to (ESnet supports both L2 and L3 circuits, while Internet2 only L2). Figure 6 presents the results of traffic tests between BNL and UMich. The target host at UMich, the same for all traffic streams, has a maximum capacity of 10 Gbits/second. Priority traffic between BNL and UMich is competing against other inter-site traffic and traffic local to UMich. The desired rate of the priority traffic is 700 Mbits/second, achieved only when a TeraPaths reservation is active. The rate of the competing traffic drops by approximately 500 Mbits/second, which is gained by the priority traffic for the duration of the reservation.

TeraPaths instances can regulate and guarantee the bandwidth of multiple flows between the testbed sites. These flows may utilize individual WAN circuits or may be grouped together, based on source and destination, into the same WAN circuit (which accommodates the aggregate bandwidth). Figure 7 shows a demonstration of flow bandwidth regulation for multiple periodic data transfers as monitored by Internet2's perfSONAR system. The aggregate bandwidth passing through circuits between BNL, UMich, and BU is displayed. Two transfers take place during each period, with each transfer maintained at a guaranteed bandwidth level. The second transfer (2) starts later than the first (1) and continues after the latter finishes. Each flow is policed to its guaranteed bandwidth level preventing competition within the circuit. Use of DiffServ QoS in the end site LANs and dynamic WAN circuits ensures that presence of any other traffic does not affect the regulated flows. In the particular example, transfer (2) is being policed even after transfer (1) is over. In the general case, it is possible to alter the policing rules to allow the continuing transfer to use all the bandwidth of the circuit. The QoS guarantee provided by the TeraPaths and OSCARS systems is at the network device level, i.e., network devices are configured to recognize specific packet flows and offer them a different level of service as determined by the coordinated system reservations. The quality of the guarantee mainly depends on the implementation DiffServ, MPLS, and GMPLS technologies in the network devices along a path. During our experiments we have observed a bandwidth variance of less than 10%, depending also on the load conditions of the network. Specifically for the end sites where DiffServ is used, the highest level of guarantee is achieved when utilizing the Expedite Forward (EF) class of service, as traffic belonging to this class is typically serviced by strict priority queuing schemes.

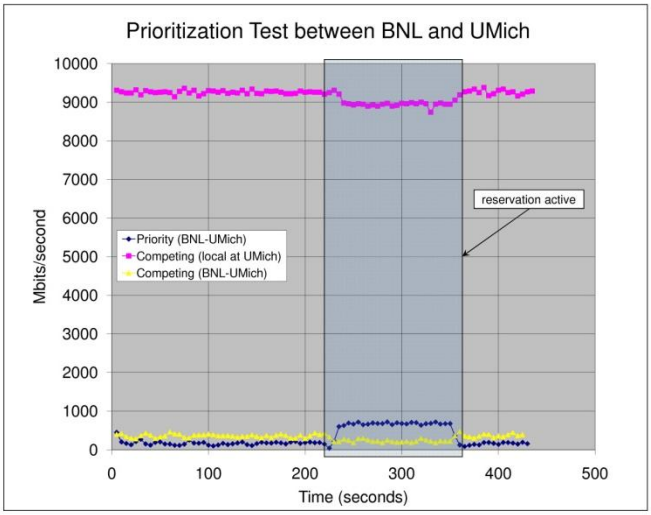


Figure 6. Traffic tests between BNL and UMich: priority inter-site traffic competing against (a) local and inter-site traffic (b) local traffic.

IV. LAYER-3 VS. LAYER-2

From the perspective of end sites, the requirements for utilizing a L2 or a L3 circuit are significantly different. In this section we discuss these requirements and related issues.

A. MPLS Tunnels (L3)

In the case the path through one or more WAN domains is established in the form of an MPLS tunnel (see Figure 8a), admission control into the tunnel is done at the ingress device of the MPLS tunnel on the WAN side. Packets that belong to an authorized flow or group of flows are recognized based on source and destination IP address and possibly additional selection criteria (e.g., port numbers). The source end site essentially hands over all packets to the WAN but only those that belong to authorized flows enter their corresponding tunnel. The MPLS tunnel maintains the packet DSCP markings so that flows emerging at the egress of the tunnel receive differential treatment within the destination end site LAN.

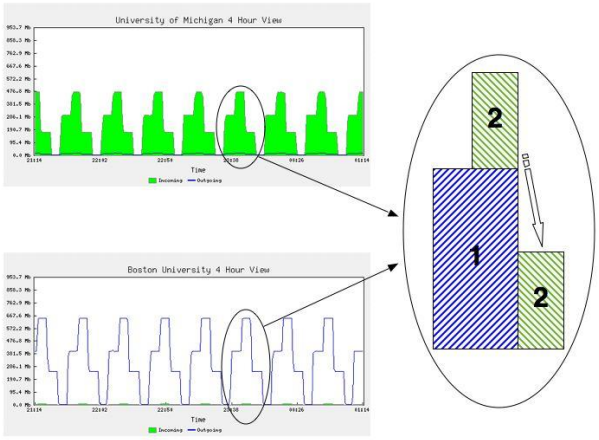


Figure 7. Demonstration of flow bandwidth regulation at SuperComputing 2007 and Joint Techs winter 2008.

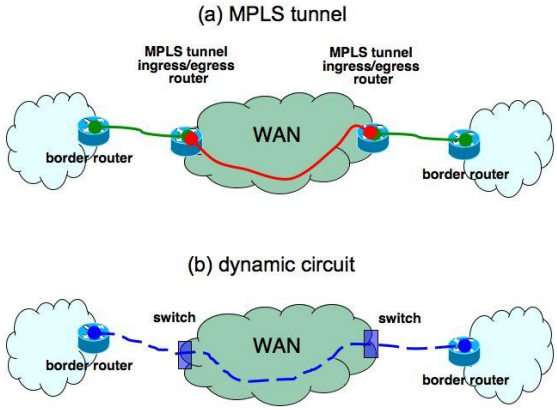


Figure 8. WAN circuits: (a) MPLS tunnels vs. (b) L2 dynamic circuits.

B. Dynamic Circuits (L2)

The infrastructure for the utilization of dynamic L2 circuits is quite different (see Figure 8b). In this case, the

WAN circuit established between two end sites makes those sites members of the same Virtual LAN (VLAN). The interfaces of the end site border routers participating in the connection appear as if connected directly with a patch cable, i.e., there is a single hop between them. Forwarding authorized traffic to the VLAN assigned to the circuit is the responsibility of each end site's border router. Each router uses Policy Based Routing (PBR) to selectively forward authorized flow packets (identified by source and destination IP addresses and possibly other criteria, e.g., ports) into this VLAN. For bidirectional traffic through a circuit, the border routers have to be configured in a mirrored configuration so that the destination site's border router appears as the next hop to the source site's border router and vice versa.

C. Related Issues

When an end site gains access to a WAN domain through a Regional Network (RN) that cannot be dynamically configured through a domain controller, it is necessary to statically configure the RN's devices so that (a) DSCP markings are not reset at the boundaries and (b) VLANs are extended through the RN. The same techniques need to be used within an end site LAN for network devices that are along routes used by end-to-end paths but are not under direct TeraPaths control. The static configuration is applied only to those specific device interfaces that interconnect TeraPaths-controlled devices with WAN devices. We call such statically configured network segments "pass-through" segments, in the sense that they honor DSCP markings and allow extension of VLANs through them. Figure 9 gives an example of a "pass-through" setup.

In both L2 and L3 circuit cases, scalability issues must be considered because both technologies require all involved network devices to be configured to recognize specific data flows. Both MPLS tunnels and dynamic circuits are technologies well suited to establish special connections between WAN endpoints and accommodate qualifying traffic between sites connected to these endpoints. However, dedicating an MPLS tunnel or a dynamic circuit to each individual flow between a pair of end sites may cause severe scalability problems, especially in the case of dynamic circuits. With MPLS tunnels, scalability depends on the limitations and efficiency of the WAN hardware, while reserved bandwidth is allocated only when qualifying flows are present. MPLS tunnels are unidirectional, so bidirectional flows require two separate WAN reservations, one for each direction. With L2 dynamic circuits, additional restrictions apply. Because a circuit behaves as an Ethernet-based VLAN, a fundamental requirement is the utilization of the same VLAN tag along the entire route covered by the circuit. All network devices along the path must use the same VLAN tag. This is a severe restriction as current devices support a total of roughly 4,000 tags with several tag ranges reserved for device use and for administrative reasons. Therefore, only a small fraction of the overall tag range is actually available for utilizing dynamic circuits, furthermore, each domain may have its own tag subset. The establishment and utilization of a circuit between two end sites requires all domains along the path to have a common subset of tags. In the current implementation of TeraPaths, this is required so that no tag conflicts exist when setting up a circuit. This requirement may be relaxed in the future by exploiting VLAN renaming capabilities.

In the TeraPaths testbed there is an agreement that 50 VLAN tags, 3550-3599, are reserved for dynamic circuit use. Ensuring that no tag conflicts exist within the testbed is relatively easy, because all testbed sites are serviced by ESnet and Internet2, which form a composite domain that can be configured by contacting a single OSCARS instance. Thus, it is possible to rely on OSCARS to select an available VLAN tag within a range suitable for the end sites involved.

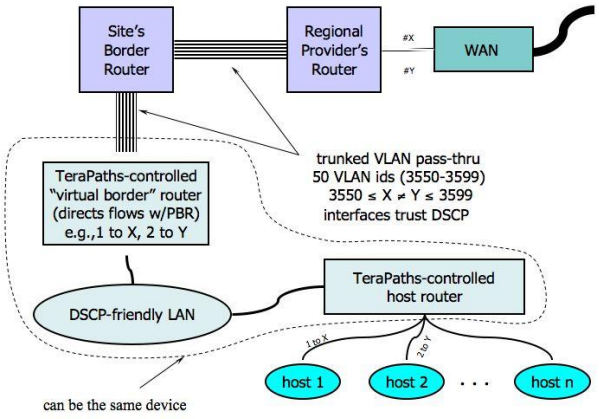


Figure 9. Example pass-through configuration for the end site's regional network and border router. The router where circuit VLANs terminate plays the role of a "virtual border" router. If only one router is controlled by TeraPaths, this router both conditions and forwards authorized traffic.

The limitation in the number of available VLAN tags and the additional properties of circuits to reserve bandwidth regardless of the presence of qualifying traffic and to be bidirectional make evident the need to treat L2 dynamic circuits as an "expensive" resource requiring sophisticated techniques to maximize utilization efficiency. Clearly, such circuits need to be viewed as "highways" between end sites. Flows with matching source and destination need to be grouped together and forwarded through common circuits, configured so that they accommodate the aggregate bandwidth of the grouped flows.

V. MANAGING WAN RESERVATIONS

Grouping together individual data flows or flow groups with common source and destination and forwarding them to a common WAN circuit with enough total bandwidth and duration to accommodate all flows can drastically reduce the number of circuits that are needed between a pair of end sites simultaneously and increase the availability of the dedicated paths. The first step of this approach is to decouple the end site reservations with the WAN reservations. End

sites still reserve resources for individual flows, however multiple end site reservations can be accommodated by a single WAN circuit reservation as long as the aggregate duration and bandwidth can be determined. The level of reservation consolidation (or unification) needs to be controlled by suitable criteria to minimize waste of resources. Figure 10 shows an example of such criteria. If all reservations #1 through #5 were to be associated with a single encompassing WAN reservation, the resource waste would be significant because of the short but high-bandwidth reservation #4 and the distance in time between #4 and #5. Therefore, limits in the maximum difference in bandwidth between reservations (Δbw) and the time period between the end of one reservation and the beginning of the next (Δt) have to be taken into account when selecting which reservations should be consolidated.

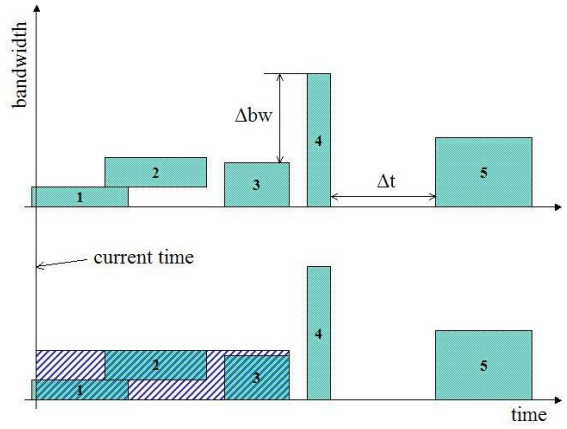


Figure 10. Example of reservation consolidation. Unifying reservations #1, #2, and #3 is feasible, #4 has too big Δbw, #5 is too distant in the future.

The initiating ESDC needs to handle the WAN reservations on the one hand, and the configuration of both end sites on the other. Although basic WAN reservation primitives can be used for consolidating reservations, additional primitives may be necessary to streamline the process and make it effective. Using basic primitives, the ESDC can create a new WAN reservation (for a dynamic L2 circuit this requires at least one VLAN tag to be available) to accommodate a newly arrived reservation that fulfills the criteria to use a specific circuit. If the circuit is pending, the consolidated WAN reservations can be immediately cancelled. However, if the circuit is already active, all relevant traffic must be switched to the new VLAN before the cancellation. With L3 circuits, this switching is not necessary. A problem with this technique is that the submission of the new WAN reservation may fail due to lack of available bandwidth occupied by reservations that will be cancelled. A new WAN primitive, allowing the submission of a reservation while taking into account the simultaneous cancellation of a set of existing ones would greatly increase the efficacy of the technique.

If the WAN domain controller allows modification of its reservations to a certain degree, it is possible to extend a reservation time-wise and/or to modify its bandwidth. While time-wise modifications are straightforward and are contingent on resource availability, bandwidth modifications need to be considered not only with regard to when they should take place within active or pending reservations, but also with regard to what the repercussions will be for existing connections through an active circuit which may be interrupted during reconfiguration.

We consider here two optimization and consolidation techniques for WAN reservations. We assume that initially WAN reservations correspond 1-to-1 to end site reservations. However, committing a reservation and deactivating a reservation are events triggering an optimization and consolidation phase for the WAN reservations. In both event cases, active or pending reservations within specific time "distance" before the beginning and/or after the end of a new reservation can be selected for consolidation. These techniques are roughly analogous to disk buffering or caching, i.e., "read ahead" and "write behind". The goal of disk caching is to maximize the utilization of the disk and speed up access by buffering as much data as possible with read operations and before write operations. In a similar sense, selecting WAN reservations based on optimization criteria (e.g. reduce waste of resources) and consolidating them maximizes the utilization of a circuit and reduces the number of expensive create and teardown operations. We thus call these two techniques "create ahead" and "teardown behind."

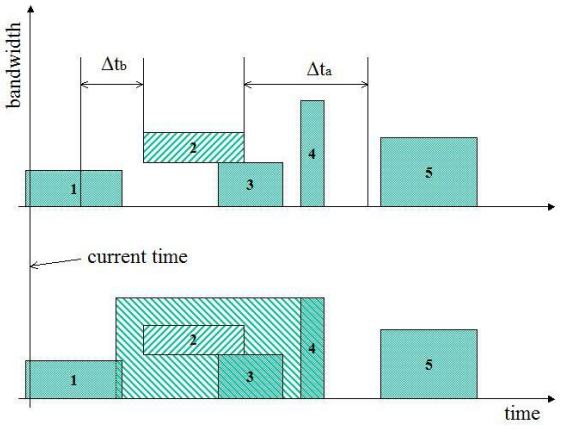


Figure 11. An example of "create ahead". #2 is a new reservation. Circuit corresponding to #1 is modified to accommodate #2, #3, and #4 with a single reservation. #5 is too distant.

"Create ahead" (see Figure 11) selects WAN reservations within Δt_b before the start of a new reservation and Δt_a after the end of a new reservation for consolidation, if additional limits in bandwidth differences and time distance are met. To reduce waste of resources, the second technique "teardown behind" (see Figure 12) modifies a unified reservation to conform to the bandwidth requirements at the time when the corresponding end site reservation expires by consolidating WAN reservations within Δt_a after the expiration of the end site reservation. The net result of the

combination of the two techniques is to reduce the number of required circuits and the frequency of circuit creation and teardown operations for circuits between the same end sites while also reducing the waste of WAN resources.

In the remainder of this section, we formulate the reservation consolidation problem and devise an algorithm to apply the above techniques to minimize the request blocking rate. We consider both the offline case, where a set of reservation requests are given in a batch, and the online case, where a new request is serviced with possible reconfiguration of existing reservations. Extensive simulation results show the tradeoff between bandwidth utilization and VLAN ID utilization.

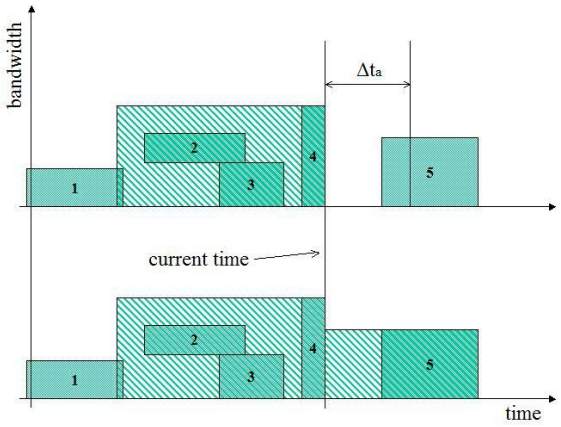


Figure 12. An example of "teardown behind". When #4 expires, the circuit servicing #2, #3, and #4 is not torn down, but instead modified to accommodate #5.

A. Models and Assumptions

An advance reservation request can be represented by a 3-tuple $r_i = (r_i^s, r_i^e, r_i^b)$, which asks for a reservation with bandwidth r_i^b within an active window (r_i^s, r_i^e), where r_i^s is a future starting time. The main challenging issue is, when given a request or a set of requests, to find the most cost-effective way to allocate bandwidth for each circuit and map each request to a circuit. In our model, one circuit has to be established with a constant bandwidth during its life since bandwidth-varying circuit reservations are not supported in the WAN. However, more than one reservation can be consolidated at the end site and then be carried on one circuit. This flexibility intuitively leads to two benefits: saving VLAN IDs and reducing the number of tear-down and setup operations. These two benefits are important because the number of VLAN IDs can be very limited in practice and the tear-down and setup operations can be costly. The downside of consolidating reservations with different bandwidth requests and active windows is that not all reserved bandwidths are used for the actual data transfer during certain intervals, which translates to lower resource utilization. In the following, we will study the tradeoffs between bandwidth utilization and circuit management efficiency.

B. Bandwidth Allocation and Circuit Assignment (BACA)

1) Offline case

We first study the problem of how, given a set R of requests $r_i, i \in \{1,2,...,m\}$, to allocate bandwidths and assign

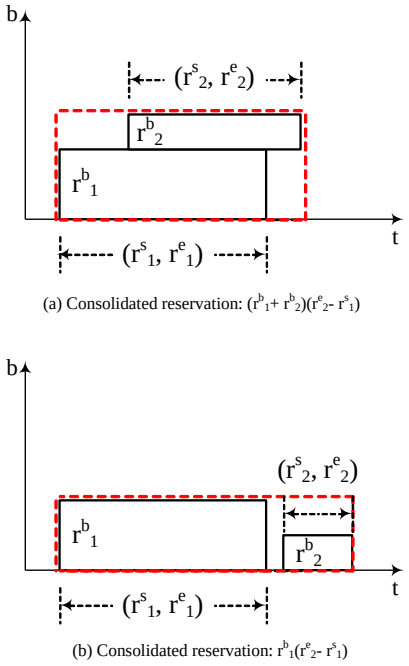


Figure 13. Illustration of reservation.

requests to circuits such that the maximum number of requests can be satisfied. In this way, the service provider can accommodate as many requests as possible or in other words, achieve high availability.

More specifically, we need to make decisions on 1) the bandwidth allocation c_j^b and active duration (c_j^s, c_j^e) for each circuit $c_j, j \in \{1,2,...,n\}$, and 2) the assignment of reservations to circuits $x_{ij}, i \in \{1,2,...,m\}, j \in \{1,2,...,n\}$. The objective is to satisfy as many requests as possible, while observing the following constraints:

- Each reservation is assigned to a circuit.
- The total bandwidth used at any time is bounded by a given capacity C.
- If a reservation is assigned to a circuit, its active window must be within the active window of that circuit.
- Within one circuit, the maximum simultaneous data transmission rate must be bounded by the bandwidth allocated for that circuit.
- The bandwidth utilization in each circuit must be higher than a given value β.
- The number of available circuit IDs are constrained by a given value.

2) Efficient Heuristics for the BACA Problem

First, we order requests by their start times such that $r_i^s < r_j^s, i < j$. Second, if two reservations are not overlapping but are close enough to justify consolidation against additional tear-down and setup operations, we also consider them "overlapping", which makes them subject to consolidation too. Last, we perform admission control. That is, if $r_i^b > C$, we reject (and remove) the request by setting $x_{ij} = 0, \forall j \in \{1,2,...,n\}$. Before we describe the heuristic, we define the following:

- One-to-one assignment: allocate a circuit c for a request r by setting $c^b = r^b, c^s = r^s, c^e = r^e$ and set $x_{rc} = 1$.
- Consolidated reservation: If two reservations are overlapping, $r^v = (r_1^b + r_2^b)(\max(r_1^e, r_2^e) - \min(r_1^s, r_1^s))$, which is illustrated in Figure 13(a), where the *x* axis is time *t* and *y* axis is bandwidth *b*. If two reservations are not overlapping but very close, $r^v = \max(r_1^b, r_2^b)(\max(r_1^e, r_2^e) - \min(r_1^s, r_1^s))$ as illustrated in Figure 13 (b).
- Minimum bandwidth utilization guarantee: If $r^v \le \frac{1}{\beta}[r_1^b(r_1^e - r_1^s) + r_2^b(r_2^e - r_2^s)]$ is satisfied.

Now, we describe the algorithm as shown in Figure 14.

1: initiation.
2: **while** *R* is not empty and $k \le n$ **do**
3: select request *r* with earliest start time from *R*. Set *consolidation = false*.
4: let *R′* be the set of reservations that will start (have not yet started) earlier than *r* and also overlaps with *r*
5: **for** each reservation $r' \in R'$ in increasing order of consolidated reservation volume. **do**
6: **if** the consolidated reservation (from *r* and *r′*) meets the minimum bandwidth utilization guarantee and does not violate the total capacity constraint **then**
7: consolidate r and *r′*.
8: set *consolidation = true*
11: **end if**
12: **end for**
13: **if** *consolidation = false* **then**
14: **if** all circuit IDs are used **then**
15: reject r and return failure;
16: **else**
17: find first available ID *k*.
18: **if** assigning *r* to circuit *k* in a one-to-one fashion violates the total capacity constraint **then**
19: reject *r* and return failure;
20: **else**
21: assign *r* to circuit *k* in a one-to-one fashion.
22: **end if**
23: **end while**

Figure 14. Proposed BACA algorithm.

3) Online case

The above algorithm can be easily adapted for use with an online case, where a new request is serviced without the information of future reservation requests. More specifically, given a new request, we retrieve its adjacent reservations within a predefined "optimization window" and form a set of reservations *R* (including the newly arrived one) for re-optimization. We then can use the above algorithm to reconfigure existing reservations in order to maximize the number of satisfied reservations. However, if the reconfiguration rejects existing reservations, we will reject *r* instead. In other words, only when the reconfiguration can reserve all the requests in *R* do we actually commit the new configurations in the reservation table. In addition, those reservations in *R* that have already been in effect will not be reconfigured. However, we need information about them in the re-optimization in order to obtain the current bandwidth and VLAN ID usage.

C. Qualitative Analysis

In general, if we require a higher bandwidth utilization β when we optimize bandwidth allocation and circuit assignment using reservation consolidation, more VLAN IDs will be used. In the extreme case when $\beta = 100\%$, each reservation uses a distinguished VLAN ID. In this way, we limit the bandwidth waste in each circuit (as shown in Figure 12) so that the total capacity consumption is lower. The above qualitative analysis or hypothesis is summarized in Table 1:

Bandwidth utilization β in one circuit	VLAN ID Consumption	Capacity Consumption
high	high	Low
low	low	High

Table 1. Qualitative analysis summary.

In the following, given the relative magnitude of available number of VLAN IDs and available capacity, we conduct simulation to obtain the bandwidth utilization β that leads to lowest (or desired) job blocking rate.

D. Numerical Study

In this section, we simulate a large number of come-and-go jobs (i.e., the online case) and evaluate the proposed BACA algorithm considering a variety of cases. To facilitate the presentation, we define a ratio r_{cb}, which is used to govern the magnitude of average bandwidth of requests compared to the total capacity and traffic intensity. The traffic intensity is defined to be the product of average request arrival rate and average reservation duration. In the simulation, we use r_{cb} to generate various jobs with different average bandwidth requests as follows:

$$\text{Average bandwidth} = \frac{\text{total capacity}}{(\text{traffic intensity} \times r_{cb})}$$

We now present results for the following cases:

1) Case 1: Sufficient VLAN IDs and varying bandwidth requests

As shown in Figure 15 (assuming 10 VLAN IDs and varying r_{cb}), higher bandwidth utilization leads to a lower blocking rate in all cases. Therefore we can verify that reservation consolidation wastes bandwidth and result in higher blocking rate when bandwidth resource is scarce. More than 10 VLAN IDs will not make any difference. Therefore, 10 IDs are considered sufficient.

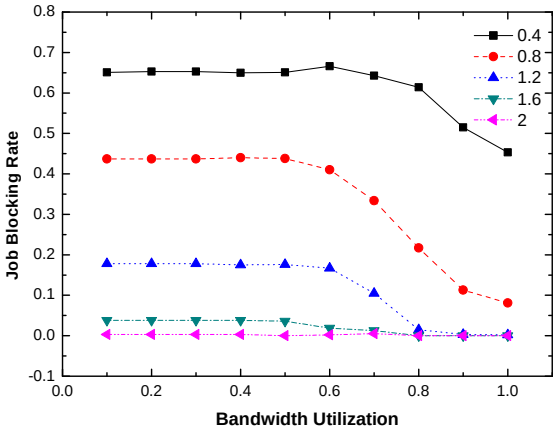


Figure 15. Sufficient (10) VLAN IDs, varying r_{cb}

2) Case 2: Sufficient capacity and varying number of available VLAN IDs

Figure 16 shows that reservation consolidation reduces the job blocking rate greatly when we have sufficient capacity (assuming r_{cb}=2) and varying number of available VLAN IDs in all cases. By "sufficient capacity", we mean r_{cb} is large enough so that a job will not be blocked due to the capacity constraint. Any value of r_{cb} larger than 2 will not make any difference.

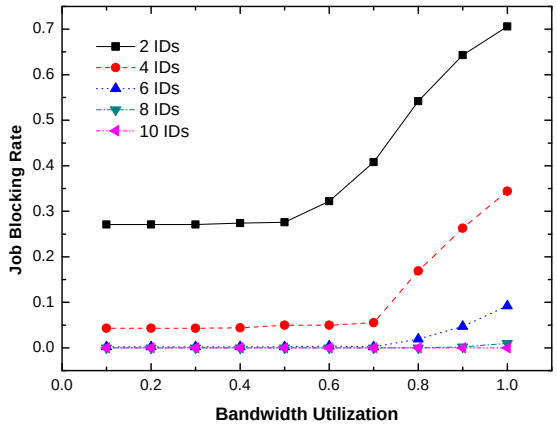
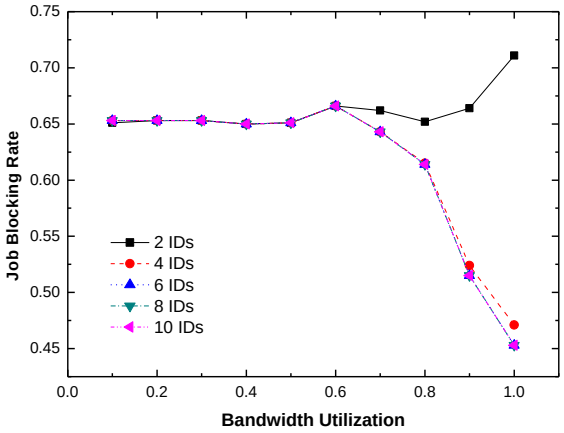


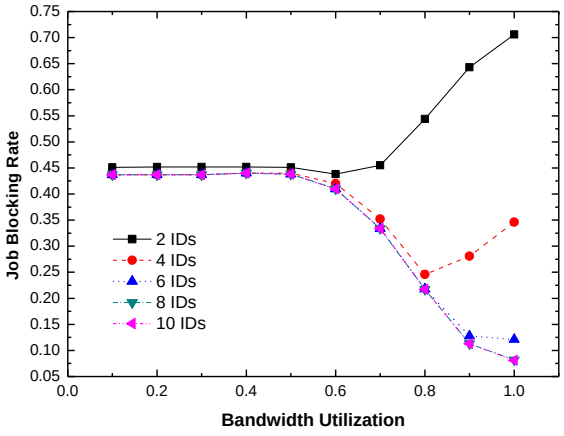
Figure 16. Sufficient capacity (r_{cb} =2), varying VLAN IDs

3) Case 3: Limited number of available VLAN IDs with different bandwidth requests

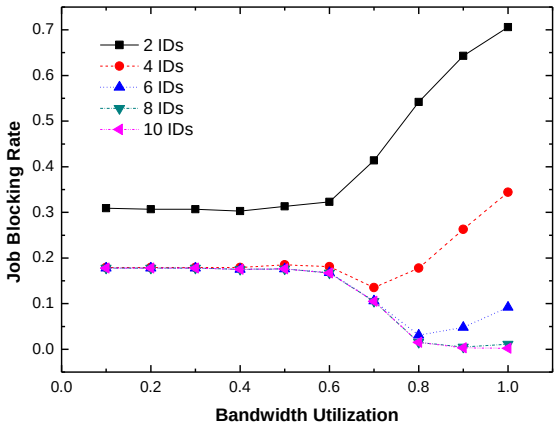
We further examine other cases. In each subfigure of Figure 17, we fix one value of r_{cb} and evaluate the job blocking performance with varying number of available VLAN IDs. For example, when r_{cb} = 1.2 and the bandwidth utilization is larger than 0.6, the blocking rate in the case of 2 available IDs begins to increase as in Case 1. However, we see a drop in blocking rate in other cases when we have more IDs. The uses of available IDs (by reducing circuit consolidation) can compensate for limited bandwidth.



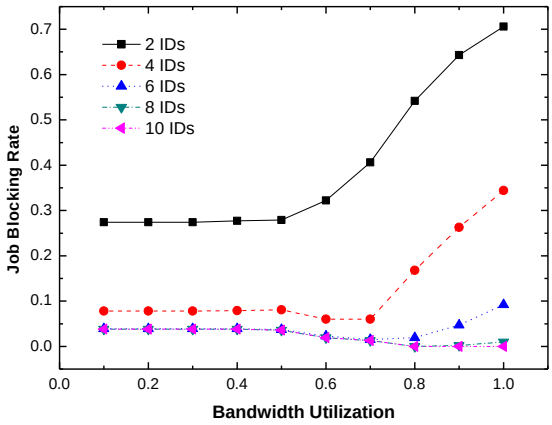
(a) r_{cb} = 0.4



(b) r_{cb} = 0.8



(c) r_{cb} = 1.2



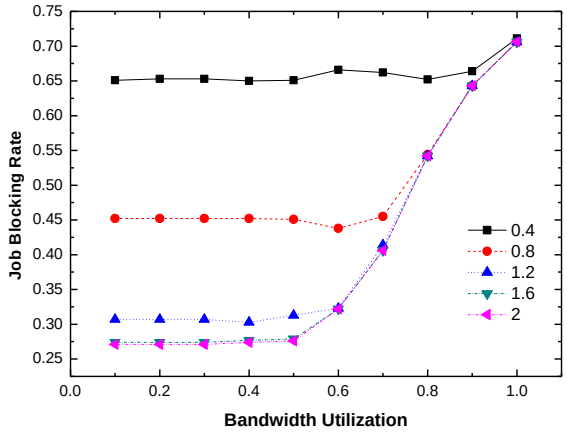
(d) r_{cb} = 1.6

Figure 17. Job blocking rate when r_{cb} = 0.4, 0.8, 1.2, 1.6

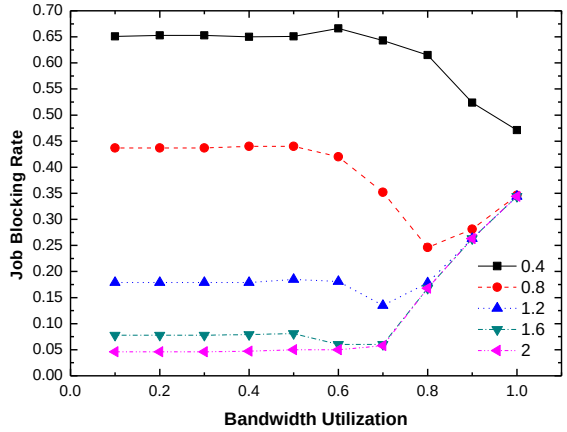
When bandwidth utilization increases further, we can see that all IDs are used up and then the blocking rate begins to increase again.

4) Case 4: Varying bandwidth requests under different number of available VLAN IDs

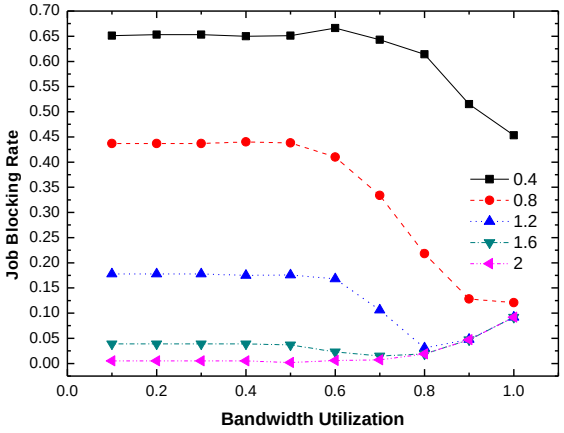
The graphs in Figure 18 also verify our hypothesis. In each subfigure below, we fix one value of available VLAN IDs and evaluate the job blocking performance with varying r_{cb}. These results can be explained by similar arguments as in Case 3.



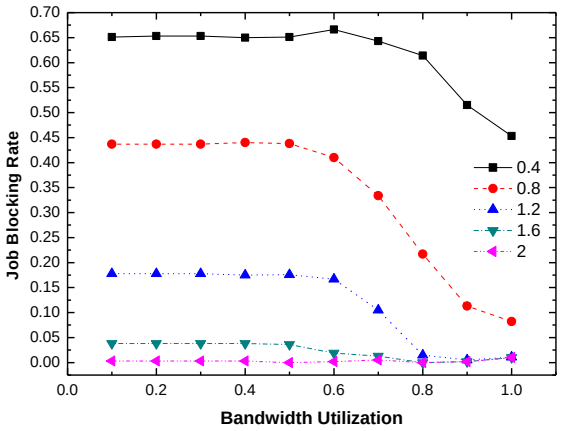
(a) 2 VLAN IDs



(b) 4 VLAN IDs



(c) 6 VLAN IDs



(d) 8 VLAN IDs

Figure 18. Job blocking rate when there are 2, 4, 6, 8 VLAN IDs

VI. FAULT TOLERANCE ISSUES

The survivability of a data transfer is crucial for data transfer applications. In TeraPaths, we view the survivability issue from a "do no harm" perspective. Because TeraPaths reserves an end-to-end path for better servicing the needs of an application, which may or may not be aware of the TeraPaths technology, our primary concern is to avoid situations where an application is disrupted because of a failure along the established end-to-end path. As such, we have started focusing on techniques to early detect and remedy configuration failures within end-sites network devices, and also handle WAN circuit failures.

In the event of a circuit failure, for any reason, flows that are being directed into that circuit will be interrupted, causing the corresponding applications to lose their connections. To prevent such situations, TeraPaths utilizes active circuit probing at the network device level. In this context, the end site network devices (border routers) that are the end points of a WAN circuit, periodically or on-demand exchange probes through that circuit for the duration of each related reservation. When a failure is detected, the immediate step is to stop forwarding traffic into the failed circuit and fall back to the standard IP network.

The next step is to attempt to acquire a new circuit and redirect traffic back into it (see Figure 19), while extending the reservations by the amount of time lost. The latter step is subject to WAN circuits becoming available again. Therefore, TeraPaths will keep trying for a pre-determined amount of time, after which the reservation will be considered failed.

With frequent periodic probes, it is possible to catch a circuit failure early and attempt to remedy the problem so that applications don't lose their connections. This approach is transparent to applications, however, it can impose significant load on the network hardware with increasing number of reservations. Thus, only highly critical reservations should be safeguarded with frequent periodic probing. A more scalable solution is to make applications aware of the probing/recovery capabilities (TeraPaths exposes these capabilities through its API) and enable them to trigger probing and recovery on-demand.

An alternative, albeit more resource-consuming, approach to recovery is to reserve in advance a backup circuit and, upon detection of failure, switch application traffic to it, instead of failing over to best effort and attempting to re-acquire the failed circuit. Steering traffic from one circuit to another is essentially instantaneous, once a failure is detected, therefore, the application should not notice anything more than a short-lived variation in bandwidth. We plan to explore this approach in our future work.

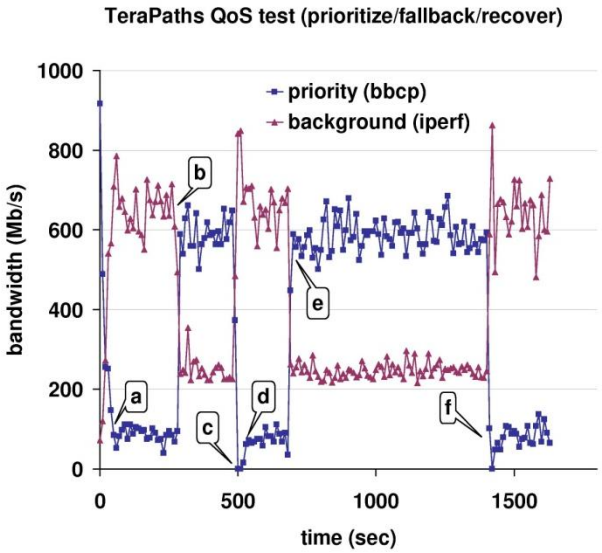


Figure 19. Demonstration of recovery: (a) competing traffic causes drop in bandwidth, (b) QoS/circuit reservation active, (c) circuit failure, (d) fall back to best effort, (e) recovery (acquired new circuit), (f) end of reservation.

VII. Related Work

The design parameters and goals of the TeraPaths project, i.e., provisioning of true end-to-end (host-to-host) virtual paths through direct configuration of end-site network devices and indirect configuration of WAN domains through tight interoperation with OSCARS, are, to the best of our knowledge, unique. In this section we compare our approach with several other systems, with which some similarities exist, in terms of design and implementation differences.

Lambda Station [11] is a Fermi National Accelerator Laboratory (FNAL) project with the goal to provide specific data intensive applications with alternate network paths between local production computing resources and advanced high performance networks. The Lambda Station service selectively forwards authorized data flows to alternate network paths, allowing such flows to utilize premium high bandwidth connections between end sites.

Phoebus [12], an Internet2 project, is a framework and protocol for high-performance dynamic circuit networks. The Phoebus approach is to split the end-to-end network path into distinct segments at "adaptation" points located at backbone ingress and egress points, then find and create an optimized network path for a specific application from each such point. Application-generated traffic between end sites is redirected to the circuit network via Phoebus Gateways.

While TeraPaths, Lambda Station, and Phoebus are all "consumers" of WAN circuits through OSCARS, TeraPaths is unique in that it uses DiffServ QoS and traffic conditioning at the edges to provide QoS guarantees to each individual flow within a group of flows going through the same WAN circuit and utilizes WAN circuit reservation consolidation techniques to practically address scalability issues.

Curti et al. [17] describes a system that can make advance reservations of lightpaths and MPLS-based layer-2 VPNs with QoS support in a large-scale network infrastructure. The authors mention possible approaches where users can scan the advertised resources of each domain and make a reservation by themselves in each administrative domain. However, synchronization problems may arise in the latter case if several reservation requests are processed at the same time.

Advance reservations have been studied in various scenarios and in different contexts. In the case of bulk data transfers, Rajah et al. [18] and Chen and Primet [19] have taken a centralized approach where resource reservation and allocation decisions are based on a global view of the network and on all job requests. As a result, it is possible to allocate network resources more efficiently. In order to improve the resource utilization, other approaches were considered in [20, 21]: a) transferring the data at time-varying bandwidth instead of constant bandwidth; b) using multiple paths for each job. For applications involving a large number of users and reserving resources from multiple domains, a distributed approach is expected to be more appropriate due to its better scalability and flexibility. In the case of distributed advance reservations, users and resource managers may need to negotiate on the reservation schedule in order to increase the success rate of submitted requests. For example, it was proposed in [20][21] that the resource manager should find another acceptable set of reservation

characteristics and attach it to the resource allocation acknowledgment that is being returned to the requestor when rejecting a resource allocation request. Furthermore, if users are willing to negotiate a flexible reservation schedule (which is likely to happen in practice), the chance of satisfying requests is increased. Yuan et al. [22] proposes a probing mechanism to deal with requests that may have certain flexibility in starting time, duration or bandwidth (but only on one dimension). However, none of the above deals with the issues of providing connectivity with a specific service guarantee across heterogeneous network domains. In particular, previous studies have not studied the benefit of reservation consolidation.

In [5][23], a prototype of General-purpose Architecture for Reservation and Allocation (GARA) was implemented to support end-to-end QoS for high-end applications. The goal of the GARA framework was to support high bandwidth flows with different QoS specifications, provide advance reservation mechanisms, and facilitate application-level monitoring. In GARA, a resource manager works as a broker to reserve and manage various types of resources, such as bandwidth, CPU and disk. A major difference between GARA and TeraPaths is that in GARA the resource manager is deployed at each domain to control resources and only deals with layer 3 flows, whereas in TeraPaths a major challenge comes from the need to reserve resources across different domains (end-site LANs and multiple WAN domains in between) controlled by heterogeneous systems and deal with traffic in layer 2 and layer 3. As a result, TeraPaths selectively conditions and forwards layer 3 traffic into layer 2 to utilize dynamic WAN circuits and also addresses issues such as reservation consolidation and reservation negotiation across different domains to improve resource utilization and availability.

VIII. Conclusion and Future Work

New network capabilities enable the establishment of end-to-end QoS-aware paths across multiple domains, paths that can be dedicated to individual data flows. Although the overall framework is in its first steps, the technology is promising as it coexists with standard best-effort networking and is accessible transparently to specific data flows. We discussed issues involved with the utilization of WAN circuits from the perspective of end sites and presented techniques that the TeraPaths system utilizes for addressing the problem of scalability with increasing number of flows. We specifically focused on the problem of maximizing system availability (minimizing job blocking rate) constrained by limited VLAN IDs and bandwidth. This is a new problem, specifically encountered when utilizing L2 dynamic circuits, which we needed to address with novel heuristics. The effective resolution of this problem will make the technology applicable to an ever-growing number of data flows between end sites and will enable effective network scheduling. Our main approach, reservation consolidation, was shown to be effective in utilizing resources through extensive simulation studies.

The TeraPaths team continues the research and development effort to improve the functionality and reliability of the TeraPaths framework, in close collaboration with the OSCARS developers. Our near future plans include study and evaluation of an efficient negotiation protocol across multiple administrative domains to complement our BACA algorithm in providing end-to-end bandwidth guaranteed connections. This negotiation protocol considers flexible/negotiable user requests, suggestions of alternative reservations from services providers, and time-varying bandwidth within the same reservation in order to push the resource utilization as high as possible. We also plan to expand and improve upon the fault tolerance capabilities of TeraPaths, not only by pursuing early failure detection and recovery in a scalable way, but also by exposing services that make applications aware of such capabilities and enable them to request status checks and/or failover actions whenever they deem it necessary. In the longer term, we intend to incorporate the framework into a more general, application-centric network virtualization system. This system will provide individual applications with on-demand guaranteed network resources dedicated and tuned to their needs while isolating them from interference from other applications and strengthening security.

References

[1] D. Katramatos, D. Yu, K. Shroff, S. McKee, and T. Robertazzi. (2009, Aug). Establishment and Management of Virtual End-to-End QoS Paths through Modern Hybrid WANs with TeraPaths. Proceedings of the First International Conference on Evolving Internet (IARIA/IEEE INTERNET 2009), Cannes/La Bocca, French Riviera, France, August 23-29, 2009.

[2] High-Performance Networks for High-Impact Science. Report of the High-Performance Network Planning Workshop. August 2002. [Online] Available: http://www.doecollaboratory.org/meetings/hpnpw/finalreport/high-performance_networks.pdf (most recent access: May 2009).

[3] DOE Science Networking Challenge: Roadmap to 2008. Report of the DOE Science Networking Workshop (June 2003) [Online] Available: http://www.es.net/hypertext/welcome/pr/Roadmap/Roadmap%20to%202008.pdf (most recent access: May 2009).

[4] GGF4: Network QoS applied to GRIDs BOF. [Online]. Available: http://server11.infn.it/netgrid/ggf/ggf4-qos-bof/index.html (most recent access: December 2009).

[5] I. Foster, M. Fidler, A. Roy, V, Sander, and L. Winkler. (2004). End-to-End Quality of Service for High-end Applications. *Computer Communications*, 27(14):1375-1388, 2004.

[6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. (1998, Dec.). An architecture for differentiated services. IETF RFC 2475. [Online]. Available: http://ietf.org/rfc/rfc2475.txt (most recent access: May 2009).

[7] R. Braden, D. Clark, and S. Shenker. (1994, June). Integrated Services in the Internet Architecture: an Overview. IETF RFC 1633. [Online]. Available: http://www.ietf.org/rfc/rfc1633.txt (most recent access: May 2009).

[8] E. Rosen, A. Viswanathan, and R. Callon. (2001, Jan.). Multiprotocol label switching architecture. IETF RFC 3031. [Online]. Available: http://www.ietf.org/rfc/rfc3031.txt (most recent access: May 2009).

[9] E. Mannie. (2004, Oct.). Generalized Multi-Protocol Label Switching (GMPLS) Architecture. IETF RFC 3945. [Online]. Available: http://www.ietf.org/rfc/rfc3945.txt (most recent access: May 2009).

[10] The TeraPaths End-to-End QoS Networking Project. [Online]. Available: http://www.terapaths.org (most recent access: May 2009).
[11] The Lambda Station project. [Online]. Available: http://www.lambdastation.org (most recent access: May 2009).
[12] The Phoebus project. [Online]. Available: http://e2epi.internet2.edu/phoebus.html (most recent access: May 2009).
[13] On-demand Secure Circuits and Advance Reservation System (OSCARS). [Online]. Available: http://www.es.net/oscars/ (most recent access: May 2009).
[14] T. Lehman, X. Yang, C. P. Guok, N. S. V. Rao, A. Lake, J. Vollbrecht, and N. Ghani. (2007, May). Control Plane Architecture and Design Considerations for Multi-Service Multi-Layer, Multi-Domain Hybrid Networks. INFOCOM 2007 IEEE. [Online]. Available: http://www.es.net/oscars/documents/papers/2007hsn-infocom-paper-lehman-etal.pdf (most recent access: May 2009).
[15] Energy Sciences Network (ESnet). [Online] Available: http://www.es.net/(most recent access: May 2009).
[16] Internet2. [Online] Available: http://www.internet2.edu/ (most recent access: May 2009).
[17] C. Curti, T. Ferrari, L. Gommans, S. van Oudenaarde, E. Ronchieri, F. Giacomini, and C. Vistoli. (2005). On advance reservation of heterogeneous network paths. *Future Generation Computer Systems*, vol. 21, no. 4, pp. 525 – 538, 2005. High-Speed Networks and Services for Data-Intensive Grids: the DataTAG Project.
[18] K. Rajah, S. Ranka, and Y. Xia. (2009, Nov.) Advance reservations and scheduling for bulk transfers in research networks. Parallel and Distributed Systems, IEEE Transactions on, vol. 20, pp. 1682–1697, Nov. 2009.
[19] B. B. Chen and P. V.-B. Primet. (2007). Scheduling deadline-constrained bulk data transfers to minimize network congestion. Cluster Computing and the Grid, IEEE International Symposium on, vol. 0, pp. 410–417, 2007.
[20] D. Ferrari, A. Gupta, and G. Ventre. (1997). Distributed advance reservation of real-time connections. *Multimedia Systems*, vol. 5, no. 3, pp. 187–198, 1997.
[21] A. Hafid, G. von Bochmann, and R. Dssouli. (1998). A quality of service negotiation approach with future reservations (nafur): a detailed study. *Comput. Netw. ISDN Syst.*, vol. 30, no. 8, pp. 777–794, 1998.
[22] L. Yuan, C.-K. Tham, and A. L. Ananda. (2003). A probing approach for effective distributed resource reservation," in *QoS-IP 2003: Proceedings of the Second International Workshop on Quality of Service in Multiservice IP Networks*, (London, UK), pp. 672–688, Springer-Verlag, 2003.
[23] V. Sander, I. Foster, A. Roy, and L. Winkler. (2000). A differentiated services implementation for high-performance TCP flows. Computer Networks, Vol. 34, No. 6, pp. 915-929, 2000.

State of the Art and Innovative Communications and Networking Solutions for a Reliable and Efficient Interplanetary Internet

Giuseppe Araniti
DIMET
University "Mediterranea" of Reggio Calabria
Reggio Calabria, ITALY
araniti@unirc.it

Igor Bisio
DIST
University of Genoa
Genoa, ITALY
igor@dist.unige.it

Mauro De Sanctis
Dept. of Electronics Engineering
University of Rome "Tor Vergata"
Rome, ITALY
mauro.de.sanctis@ uniroma2.it

***Abstract* — In the last few years deep space exploration missions are undergoing a significant transformation as are the expectations of their scientific investigators and the public who participate in these experiences. National Aeronautics and Space Administration (NASA) and European Space Agency (ESA), recently, decided pursuing a mission to study Jupiter and its moons, and another to visit the largest moons of Saturn. Those missions need new communication and networking infrastructures able to support space exploration, to connect scientists and their instruments, and also to involve the public via common web interfaces. A possible solution is represented by the so called InterPlaNetary (IPN) Internet that introduces new challenges in the field of deep space communications.**

In that framework, the paper proposes a description of the challenging scenario, surveys its technical problems and envisages possible advanced communications and networking solutions starting from the analysis of a specific IPN architecture. In more detail, we study the network performance changes due to the nodes' movements from the communications and the networking viewpoint. It represents the main contribution of the paper and opens the doors to future advanced solutions suited to be employed in the IPN Internet.

Index Terms* — *Interplanetary Networks Architecture,, Delay Tolerant Network, Advanced IPN Node, Multicast, Link Selection.

I. INTRODUCTION

Nowadays the early exploration missions, sponsored by both NASA and ESA, are giving way to a new data-intensive era of long duration observational outposts, landed vehicles, sample returns, and multi-spacecraft fleets and constellations. These missions and the future ones will require a robust, efficient and flexible communication infrastructure able to connect earth mission centres with space elements (*Mission Applications*), scientists with remote instruments (*Scientific Applications*) and engage the public by giving them traditional Internet visibility into the space missions (*Public Applications*). This new connection capability matches the vision of the InterPlaNetary (IPN) Internet. In this view, the IPN Internet means orders of magnitude increases in data rates and highly automated communications between remote planets and Earth.

In synthesis, the purposes of this paper are:

- to survey the Interplanetary communication scenario, its characteristics, problems and existing/innovative solutions;
- to analyze a possible IPN network infrastructure, proposed in the following Section;
- to test some networking solutions applied to the Interplanetary network.

Starting from a preliminary study reported in [1] in this paper we aim to highlight the problems that compromise communications among planets, such as huge delays, limited bandwidth availability and link blackouts. In that environment, a set of partially unexplored technical solutions, aimed at connecting the IPN network reliably and efficiently, represents the starting point of the research [2]. In more detail, in this work a study of a IPN network architecture has been proposed and the extremely complex situation in which communication and networking systems operate is shown. In particular the enormous delays and the scarce available resources may compromise a communication process over that network significantly.

Furthermore, together with the study introduced previously, the paper also analyzes the possible exploitation of networking solutions, which plays a crucial role: the Multicast Transmissions [3] and the Link Selection [4]. In the former case, the necessity is due to the scarce resource availability: in case of multiple destinations of the same information, the minimization of the number of traffic flows is strictly needed. For example, if a Mission Control Centre needs to upgrade the software onboard of several IPN nodes (e.g., rovers over planets or orbiting satellites) just one traffic flows will be sent from Earth. Analogously, if a planetary image, acquired by a rover, should simultaneously reach two different ground stations on the Earth, just one copy of that image will be sent from the rover through the IPN network.

The latter considered solution, the link selection, is a network control approach, based on the employment of the Delay Tolerant Networking (DTN) paradigm [5], which allows selecting the best available IPN channel to forward the traffic flows. It is aimed at maximizing the network performance because it permits the exploitation of the best link currently available for a given IPN node that sends data traffic.

All the previously mentioned purposes of the paper are pursued in this work starting from the ongoing research activity of the authors [1].

The remainder of this paper is structured as follows. Section II introduces the general IPN network architecture and describes the network analysed in this work. Section III illustrates the essentials research challenges of the IPN environment and surveys the state of the art in the field. The simulative IPN network study concerning bandwidth availabilities, delays and link blackouts is presented in Section IV. Section V proposes an overview of possible technological communications and networking solutions for the future IPN Internet. Section VI proposes a functional architecture suited to be employed in the IPN scenario and, in particular, analyses the Multicast Transmission and Link Selection necessities with an

introductive performance investigation carried out by *ns-2* simulation. Conclusions are drawn, finally, in Section VII.

II. SCENARIO

An interesting overview of IPN network architectures is reported in [2] and it is briefly synthesized in this section. As depicted in Figure 1, an interplanetary network can be split into three different sub-networks:

- IPN Backbone Network;
- IPN External Networks;
- PlaNetary (PN) Networks.

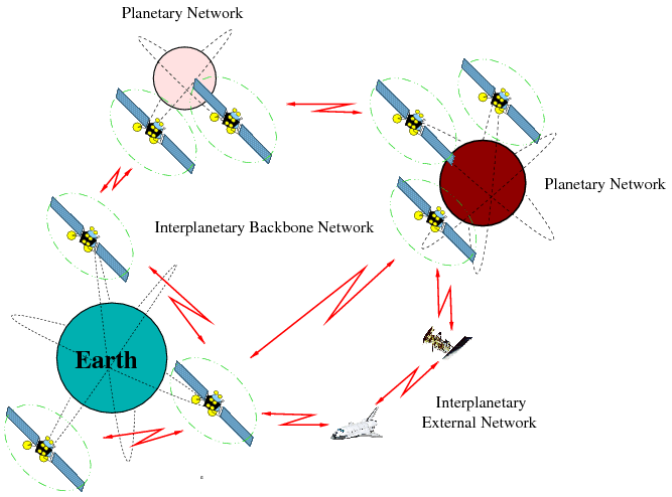


Figure 1. IPN Network.

The *IPN Backbone Network* provides a common infrastructure for communications among Earth, planets, moons, space probes and rovers through spacecrafts (e.g. satellites or orbiters), which operate as network nodes allowing transmissions over deep space channels.

The *IPN External Network* consists of nodes that are spacecrafts flying in deep space between planets, space probes, and orbiting space stations. Nodes of the IPN External Network have both long and short-haul communication capabilities. The former are employed if the nodes are at long distance from the other IPN nodes, the latter are employed at nodes flying in proximity of other ones.

The *PN Network*, depicted in Figure 2, is composed of the *PN Satellite Network* and the *PN Surface Network*. The former includes links among surface nodes, orbiting satellites and IPN Backbone Nodes, providing a relay service between surface network and backbone network and between two or more parts of the surface network. The latter provides the communication links among surface elements, such as rovers and sensor nodes which may have the communication capability towards satellites. It also provides a wireless backbone over the planet employed by surface elements that cannot communicate with satellites directly.

Concerning the *Mission Applications*, a first example is the reporting to the mission centre of astronauts' health and spacecrafts status telemetries. Another space mission application is the Command and Control of in-situ elements from Earth or from proximity spacecrafts.

Concerning *Scientific Applications*, a new approach is the so called "Virtual Presence". This type of application is intended to send great volumes of information about a monitored remote planet in order to allow scientists, or in-situ robots and astronauts to interact with high-fidelity models of the monitored area.

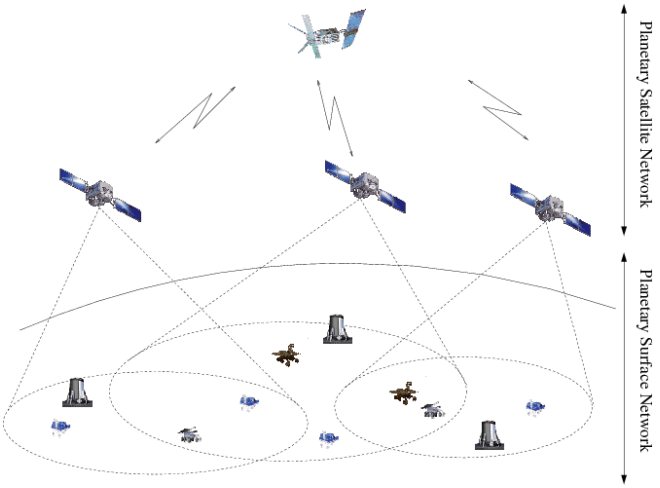


Figure 2. PN Network.

The prospective applications of a space mission, and its related communications network, can be extended beyond the mere space missions' management or the scientific applications. People surfing the Internet today access websites in extreme locations (e.g., Antarctica) and in the future they may be able to access servers on space to request data directly. This concerns the *Public Applications*. New technologies will enable communications to web servers on International Space Stations, space probes and crafts, on the Moon and other planets of the solar system.

In this paper, we utilize the previous mentioned network subdivision to define a specific IPN internet architecture able to provide connectivity between the Earth and two different PlaNetary networks. In more details, as shown in Figure 3, two PN networks are employed over a remote planet (e.g., Mars) and over the Moon. In both cases, the Surface PN network is composed of two landers (MS1 and MS2 over the remote planet and LS1 and LS2 over the Moon), able to transmit information such as images, sensed data (e.g., temperature, humidity etc.), towards the PN Satellite Network. PN satellite networks are structured with four orbiting satellites (MO1, MO2, MO3 and MO4) in the case of the remote planet and two orbiting satellites around the Moon (LO1 and LO2). Over Earth, the PN surface network is composed of six surface nodes. They are typically the destination of the information sent from remote planets and, simultaneously, the source of possible control messages transmitted towards the IPN nodes (e.g., from Mission Control Centres). In detail, Earth Surface nodes are the ones of the well-known DSN - Deep Space Network (ES1, ES2 and ES3) and other possible nodes, such as Space Science Research Centres, distributed over the planet (ET1, ET2 ET3 and ET4). Concerning the PN Satellite Network, three Geostationary satellites (GEO1, GEO2 and GEO3) have been included in the architecture. They are supposed spaced of 120° so allowing the maximum coverage of Earth surface.

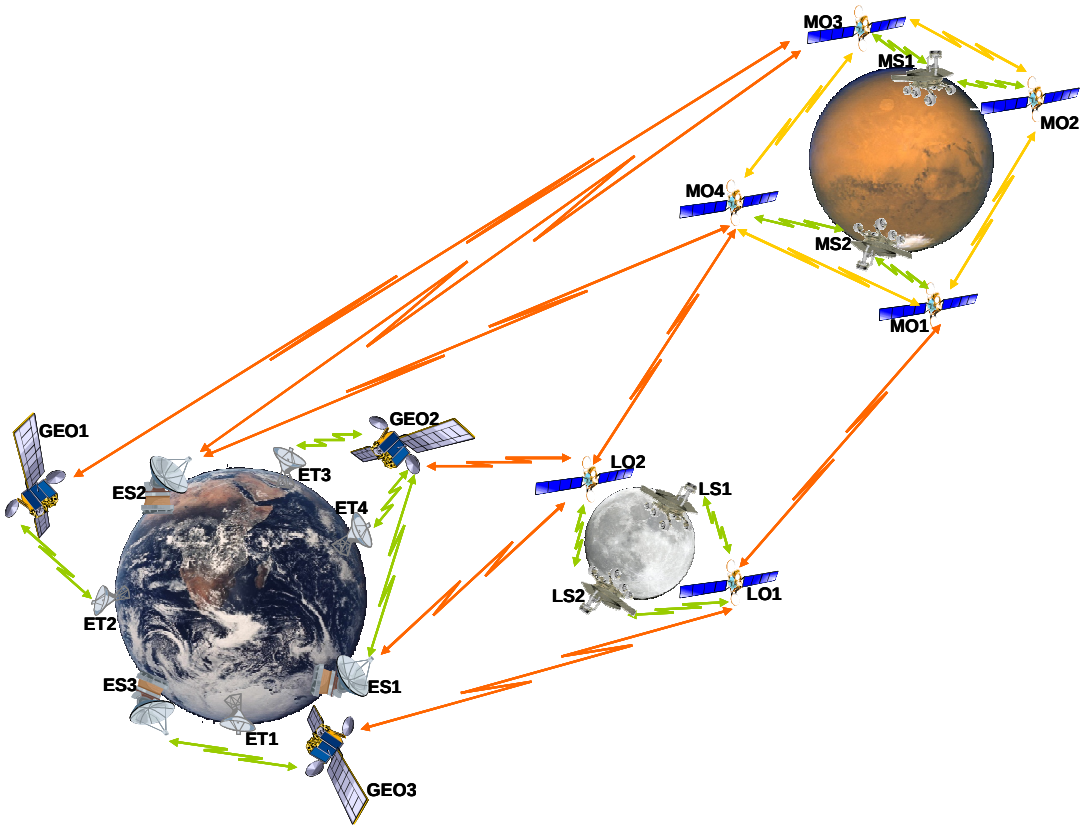


Figure 3. Example of IPN Network Architecture.

Each orbiting satellite of the IPN network has been also considered as a node of the Backbone Network. No External Networks have been considered in this architecture. All details concerning the link: available data rate, propagation delays, network movement and the consequent link blackouts, will be the object of investigation in Section IV where the simulative study of the IPN network architecture has been proposed.

III. CHALLENGES AND STATE OF THE ART

a) Challenges

From the communications viewpoint, the main problems of the IPN scenario concern: extremely long and variable propagation delays (e.g., 3-20 minutes for Mars to Earth); asymmetrical forward and reverse link capacities; high error probability; intermittent link connectivity (due to satellites, spacecraft and space probes eclipses and common link failures due to disturbances); absence of fixed communication infrastructure; attenuation of the transmitted signals due to distances; power, mass and size of communications hardware and costs, of both in terms of hardware and protocols' complexity; backward compatibility requirement due to high cost involved in deployment and launching procedures. These problems strongly compromise the reliability and the efficiency of a communications process over an IPN network and, as a consequence, the reduction of the impact of them on the communications represents the research challenge of the considered environment.

b) State of the Art

In this Sub-Section the State of the Art in the field has been surveyed and it constitutes the starting point for the solutions proposed and investigated in the following.

b1) Satellite Constellations

An important element of the IPN network architecture is the planetary satellite constellation. Generally, satellite constellations are needed instead of a single satellite, because the latter can cover only a limited portion of the Earth and only a limited number of planets [6]. On the other hand, satellite constellations can provide: simultaneous multiple coverage, continuous global coverage, continuous regional coverage or low revisit interval. Constellation design is generally a very difficult problem because each orbit has an infinite number of choices for the six orbital parameters, so for many satellites, the problem is of exceedingly high dimensionality. This is one of the primary reasons why the art of constellation design is presently suffering from a deep technology development delay. In order to solve this complex problem, satellite constellation designers adopt very limiting assumptions preventing the discovery and development of new, useful solutions. For instance, the assumption of circular orbits (e.g. Walker constellations), while simplifying the problem from one side, strongly limits the varieties of potential configurations.

b2) Physical Layers

Another hot topic in relation to IPN network concerns the physical layers. Novel physical layers are currently under investigation by considering power efficient modulation and coding schemes, the exploitation of Extremely High Frequency (EHF) bands and the employment of Ultra Wideband (UWB) communications technologies for satellite and IPN links.

In particular, the utilization of EHF bands allows obtaining a good trade-off between antenna size, bandwidth availability and path loss, [7]. They can be fully exploited in space communications and may represent an answer to the saturation of lower frequency bands, the growth of data-rate requests and the reduction of mass and size of equipment.

Ultra Wideband (UWB) communications for satellite and IPN links is another topic under investigation. The performance studies in the literature [8] concerning UWB signals over satellite links with the constraints on the received power on the Earth surface shows that, in the mentioned conditions, it is possible to achieve a data rate of 236 Mb/s.

b3) Networking Aspects

Concerning networking, traditional TCP/IP systems are not suitable for the IPN networks where transmissions are affected by very large delays and possible lack of connectivity. This does lead to exploit the DTN networking paradigm [5] in the IPN context.

As regards Transport Layer protocols, in [2] it emerges clearly that windows-based mechanisms used by the current TCP protocols, both for wired and wireless networks, achieve very poor performance in deep space communication networks, because of extremely high propagation delays. Also the TCP extensions for the space segment, such as the Space Communications Protocol Standards - Transport Protocol (SCPS-TP) developed by Consultative Committee for Space Data Systems (CCSDS), have not demonstrated to be satisfactory enough. The introduction of a bundling approach has allowed reliable transport over intermittent links. Nevertheless, a transport layer protocol, specifically tailored for IPN communication, is needed. In [9], the Licklider Transmission Protocol (LTP) is introduced for transmission of the bundles between bundling nodes. To achieve efficient routing, new mechanisms are needed. In more detail, the DTN architecture provides a framework for routing and forwarding at the bundle layer for unicast, anycast, and multicast messages that need to be exploited and enhanced. As explained in [2], also possible Data Link Layer solutions suited to be employed in the IPN Internet exist. Nevertheless, this area is vastly unexplored and open to extensive research efforts to develop innovative solutions coherent with the requirements of the IPN Internet.

b4) Network Controls

Another important topic concerning IPN networks is the reconfigurable protocol stack, which is aimed at using information from other protocol layers, and automating the communication process in extreme environments such as the IPN one. It is the principle of cross-layer design, which is envisaged again in [2] as a necessary solution for highly challenging environments. In this perspective, future extension of this work want to explicitly fill the control gap in the currently employed communications and networking solutions applied to IPN networks.

IV. IPN NETWORK ARCHITECTURE STUDY

The design of the system architecture of an IPN network concerns the definition of the number and type of required satellites and ground stations and their location/trajectory. The design of the system architecture is driven by:

- the minimization of number of events of link blackouts;
- the minimization of average duration of link blackouts;
- the minimization of point-to-point link propagation delay;
- the maximization of number of alternative routes .

The causes of blackouts can be: link obscuration between planets, situations of solar conjunctions or out of coverage. The analysis of the proposed architecture (described in Section II) has been carried out for a sample period of 24 hours and in this analysis only the former cause has occurred. It is evident that the complete study of all the events that occur in an IPN mission requires a prohibitively long sample period (many years) which should include all the possible geometric configurations of the architecture. However, in order to set up a significant and feasible simulation scenario, this 24 hours sample period has been selected so that the most important events regarding link blackouts are included.

The architecture considered in this paper is an example of a possible realization of IPN Internet with planetary networks on Earth, Moon and Mars. The proposed example represents an interesting benchmark for the comparison with more complex system architectures. Furthermore, this system architecture is here analyzed so that the performance evaluation of link parameters (i.e. availability, delay and path loss) can be used for the simulation of network protocols (as introduced in Section VI) employing *ns-2* simulator.

It is worth noting that no cable connections between ground stations have been considered and landers can only communicate with the relative planetary network of orbiters.

Furthermore, the lunar lander LS1 is positioned on the dark side of the Moon, and hence, it could not communicate directly with the Earth without using a Lunar relay orbiter. Lunar orbiters has the further task to relay the communications between Mars and Earth when direct communication is not possible.

The average blackout duration for a selected set of IPN links between external nodes is summarized in Table I.

TABLE I AVERAGE BLACKOUT DURATION FOR A SELECTION OF IPN LINKS.

Link	Average blackout duration	Link	Average blackout duration
LO1-GEO1	2286 s	LO1-MO1	2157 s
LO1-GEO2	2496 s	LO1-MO2	2258 s
LO1-GEO3	2170 s	LO1-MO3	1209 s
LO2-GEO1	2453 s	LO1-MO4	1209 s
LO2-GEO2	2410 s	LO2-MO1	2157 s
LO2-GEO3	2156 s	LO2-MO2	2258 s
ES1-LO1	10230 s	LO2-MO3	1209 s
ES1-LO2	8939 s	LO2-MO4	1209 s
ES1-MO1	6797 s	ES1-MO3	17198 s
ES1-MO2	6726 s	ES1-MO4	17199 s

From the values of the average blackout duration, it can be noticed that the DSN Earth station ES1 (Canberra), and hence similarly ES2 (Goldstone) and ES3 (Madrid), shows a long blackout duration of the links to the Lunar or Martian orbiters. However, this is overcome by using alternative links through three GEO satellites.

Another important aspect of the system architecture is the propagation delay. The mean value of the propagation delay is shown in Table II for a selection of IPN links. The propagation delay can be as long as 20 minutes in the case of Mars-Earth connection. However, since the shortest path from Mars to Earth (i.e. the MSx-MOy-ESz path) is not always available, in many cases the total end-to-end delay can be much higher.

TABLE II AVERAGE PROPAGATION DELAY FOR A SELECTION OF IPN LINKS.

Link	Average propagation delay	Link	Average propagation delay
LO1-GEO1	1.25 s	LO1-MO1	1210 s
LO1-GEO2	1.25 s	LO1-MO2	1210 s
LO1-GEO3	1.25 s	LO1-MO3	1210 s
LO2-GEO1	1.25 s	LO1-MO4	1210 s
LO2-GEO2	1.25 s	LO2-MO1	1210 s
LO2-GEO3	1.25 s	LO2-MO2	1210 s
ES1-LO1	1.3 s	LO2-MO3	1210 s
ES1-LO2	1.3 s	LO2-MO4	1210 s
ES1-MO1	1210 s	ES1-MO3	1210 s
ES1-MO2	1210 s	ES1-MO4	1210 s

The data rate of each link has been computed on the basis of the DVB-S2 standard and with realistic values of transmission power and antenna size [10]. The performance of the DVB-S2 standard in terms of Bit Error Rate (BER) versus Signal to Noise Ratio (SNR) E_s/N_0 follows a threshold behavior which is due to the adopted modulation and coding schemes. In fact when the SNR is lower than the required E_s/N_0 the BER is very large, while when the SNR is larger than the required E_s/N_0 the performance of the system is quasi error free (BER=10^{-10}) [11]. Therefore, a constant data rate has been considered. It has been computed in each link for the maximum distance (worst case) by using the lowest modulation index and code rate (i.e. QPSK 1/4) with a packet length of 64,800 bits. However, since the DVB-S2 standard foresees adaptive coding and modulation schemes and the propagation losses are highly variable, another possible approach is to consider variable data rates on the basis of the selected modulation and coding scheme for every set of propagation losses.

TABLE III DATA RATE FOR A SELECTION OF IPN LINKS.

Link	Forward link data rate	Reverse link data rate
LOx-GEOx	100 kbps	100 kbps
LOx-MOx	1 kbps	1 kbps
ESx-GEOx	10000 kbps	10000 kbps
ESx-LOx	1000 kbps	100 kbps
ESx-MOx	10 kbps	1 kbps

The parameters reported in Table I, II and III have been also employed in the performed simulations described in the following Sections.

V. COMMUNICATIONS AND NETWORKING SOLUTIONS: A GENERAL OVERVIEW

a) Topological Solutions

The first important topic to be addressed in an efficient and reliable IPN network, which suffers the problem previously described, is the design of a system of space systems such that the durations of the link unavailability and the propagation delay (i.e., the path length) are minimized. Therefore, the research, currently ongoing and object of future extension of this paper, concerns the optimization of the architecture, defining the number and type of the required satellites and ground stations and their location [6, 12]. The envisaged optimization will consider a combination of the average duration of the link unavailability and the average propagation delay and it will deal with the orbital parameters of each satellite included in the IPN architecture. In general it is possible to design a planetary satellite constellation network and a set of ground stations such that the availability of communication links is ensured, but this implies very high costs. As a consequence, a constraint on the maximum number of satellites will be fixed and the performance of the architecture in terms of link availability and propagation delay will be optimized. A new type of satellite constellations that can be used in the system optimization process is represented by the Flower Constellations set. The name Flower Constellation (FC) has been chosen because of the compatible orbit relative trajectories in the Earth-Centered Earth-Fixed (ECEF) reference frame, resemble flower petals. A FC is a set of spacecrafts characterized by the same repeating space track, a property obtained through a suitable phasing scheme [12]. The FC approach provides great flexibility and interesting dynamics. In particular, the FCs can be designed to offer dual compatibility, hence providing synchronization with both the Earth and the target planet (e.g. Mars). This synchronization can be exploited in such a way that the link availability is maximized and the propagation delay is minimized.

b) Communications and Networking.

The second scientific topic to address in the considered IPN environment, concerns advanced physical layers, networking layers and control procedures suited for application in the IPN scenario. It is worth noting that the IPN nodes will not be based on traditional Internet Protocols, but on innovative optimized protocols, though compatible with the former ones. This point has been highlighted since the origin of the IPN Internet when, at the beginning of this decade, the first short-lived IRTF “Interplanetary Internet” group was founded by Vint Cerf and a couple of internet-drafts were proposed. In particular, at the turn of 2002 and 2003 the IPN problem scope widens to “Delay Tolerant Networking” (proposed by Kevin Fall, mainly) and the concept of bundle, briefly described below in Section VI.a, was created.

The IPN nodes will include adaptive functions that will allow employing them in each part of the considered network whatever channel conditions are experienced.

TCP/IP systems are poorly suited for adoption in networks where links operate intermittently and over extremely long

propagation delay. This analysis leads to propose a network architecture based on an independent middleware, the Bundle Layer, which is the key element of the Delay/Disrupt Tolerant Network (DTN) paradigm [5, 9]. This architecture uses an overlay protocol, which allows storing packets, between the application and the locally optimized protocol stacks. The overlay protocol serves to bridge different stacks at the boundaries between environments (e.g., PN Network and IPN Backbone) providing a general-purpose application-level gateway. It is the networking paradigm considered. However, it is not sufficient to offer reliable and efficient transmission over the IPN Internet because of the dynamics of the considered environment. A more insightful approach is needed for the join optimization of the bundle overlay layer and the other layers.

VI. THE IPN NODE FUNCTIONAL ARCHITECTURE

Starting from the general overview proposed previously, in this paper a functional architecture suited to be employed in IPN networks has been proposed. In this Section, moreover, the introductive performance investigation of some features of the proposed node (Multicast Transmission and Link Selection) have been included.

The envisaged IPN Node architecture is reported in Figure 4. It includes the Bundle Layer and a Higher Convergence Layer that act as bridge between two different portions: a standard stack (e.g., the TCP/IP one) used to connect common network devices to the IPN Node and the space protocol stack suited to be employed in the IPN environment. The Higher Convergence Layer will allow managing traffic flows both sent by standard and DTN-compatible hosts. It acts as adaptation layer and realizes the backward compatibility with common protocol stacks. After the adaptation phase all packets become bundles (the transmission unit of DTNs) and they are sent though specific transport and network layers designed for the space portion of the IPN network. The IPN Node transport and network protocols parameters will be adaptively optimized starting from the employed channel conditions. Data Link and Physical Layers have been again differentiated into two families: Long and Short-haul. In the former case, the lower layers solutions will be specialized for very long distance channels (e.g., between satellites of the IPN backbone). In the latter case, solutions are suited to be used in short distance channels (e.g., between spacecrafts and proximity satellites of the IPN network or between PN satellites and planet surfaces). The Lower Convergence Layer acts as selector between the Long or Short-haul layers in dependence on the position of the IPN network elements. Long and Short-haul protocols, opportunely designed for the IPN environment, allow implementing possible adaptive functionalities of the lower layers.

In the following, each layer of the IPN node has been briefly described and some considerations concerning the related open research issues have been included.

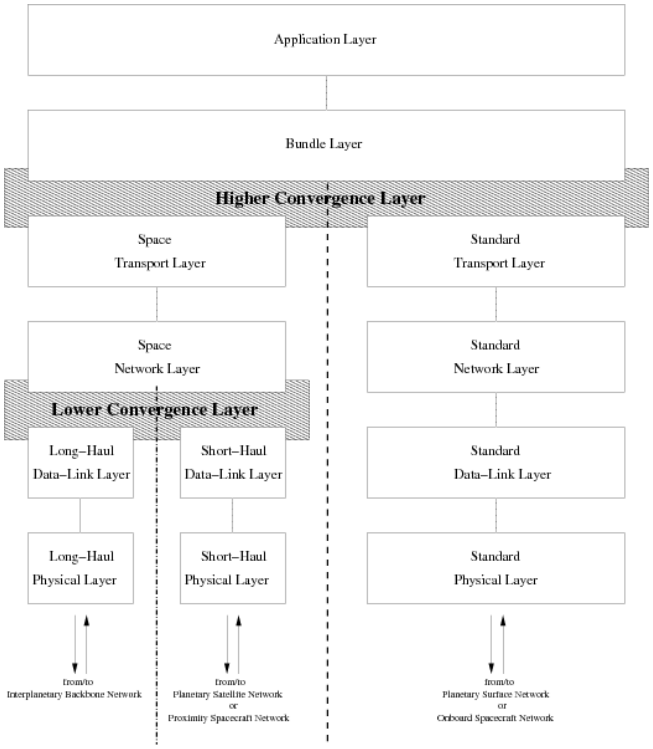


Figure 4. IPN Node Protocol Stack

a) Bundle Layer

To match the IPN environment requirements, the Bundle Layer needs to be extended. In more detail, its current specification does not include error detection mechanisms of bundles. It opens the doors to the employment of application layer coding, both in terms of source coding and error detection and recovery approaches. Other important open issues related to the Bundle Layer will be taken into account: the bundle size optimization and the related problem of fragmentation; the study and the design of common bundle layer routing approaches for the IPN environment; the Quality of Service (QoS) concept, whose meaning in the IPN network differs from the common one, together with new QoS mechanisms suited to be exploited in the considered environment.

b) Transport and Network Layers

The performance issues of the space transport and network layers represent another important research topic of the IPN node design [2]. In terms of recovery procedures and congestion control schemes, new transport protocol will be developed. For example, Additive Increase/Multiplicative Decrease concepts, able to cope with blackout events by taking advantage of probing packets will be taken into account to realize the transport layer. In turn, in the case of unavailable or strongly asymmetric return links, the transport protocol's reliability will be ensured by using appropriate strategies based on erasure codes. The problem of congestion events occurring at deep space IPN Node will be also solved by considering call admission and flow control schemes together with effective storage routing strategies.

The IPN Node protocol stack will also support the point-multipoint applications. Multicast/broadcast transmissions will allow reaching several IPN nodes, so optimizing the resource utilization. This requires the introduction of Multicast

Transmission approaches whose possible enhancements will be object of future and extensive research.

In this sub-section some preliminary simulation results, carried out by means of *ns-2* simulator, have been provided. Performance analysis has been conducted by implementing in *ns-2* the network topology and its evolution respectively described in Figure 3 an Section IV. In particular, two different files have been used as input for *ns-2* simulator: the former providing bandwidth capacities and propagation delays of different links, while the latter giving information regarding to all the link blackouts. These files are obtained from IPN network simulative studies described in Section IV. It is worth noting that results shown in this sub-section (VI.b) are obtained utilizing the standard version of *ns-2*, where packets delivery is based on the standard internet protocols, while in the remaining sub-sections, the *ns-2* simulator has been up-grated with *ns-DTN* module introducing an *DTN-Agent* with Bundle Protocols functionalities. In more details, in the *network simulator-2* the DTN module is implemented like a transport layer protocol, defining for each DTN node an *Agent* able to manage efficiently the routing and the reliable packets delivery. Moreover, the *DTN-Agent* supports the custody transfer procedure and allows to exchange bundle protocol signaling among the different DTN nodes.

They show the impact of multicast data delivery in deep space exploration missions. In particular, it has been highlighted the advantages that could be obtained utilizing groups oriented applications respect to point-to-point transmissions in the IPN scenario.

It is worth noting that, in the depicted IPN topology, (reported in Figure 3) two different kind of Multicast Connections could be thought: (i) Multicast Forward Connections (MFC), where sources are, for instance, Earth Mission Centers and receivers are the deep space nodes; Multicast Reverse Connections (MRC), for communications from remote planets to Earth. As mentioned in the introduction, the MFC could be used for Mission Applications to provide control information and to upgrade the software implemented in the IPN nodes. While, MRC could be utilized for Scientist and Public Applications to receive planetary images, videos and experimental results acquired by space stations.

The results highlight how a multicast approach could lead to a most efficient resource management compared with Unicast techniques. For instance, considering a scenario where a terrestrial node (i.e. ES1) sends data to receivers of a multicast group located on two different planets and supposing that four receivers belong to such a multicast group (two scattered on the Moon and the other ones located on Mars) the situation is as follows. Unicast approach foresees four connections between sender and receivers; this means that the same information is sent on the channel four times. Therefore, in this case a Unicast approach increases the accesses to the links needed to forward the same packet. Clearly, such a issue is more manifest when the number of receivers increases. While, a multicast approach always foresees the same number of accesses (i.e. one each planet) regardless of the number of receivers belonging to the same multicast group. These result are depicted in Figure 5 varying the number of multicast receiver for region/planet. The obtained result demonstrates that a multicast approach in IPN networks gives the following advantages: *(i)* it reduces the links utilization, saving radio resources that could be utilized to supply transmission of further services; *(ii)* it optimizes the memorization units size (buffer size) and reduces the signalization due to acknowledgment procedures and routing.

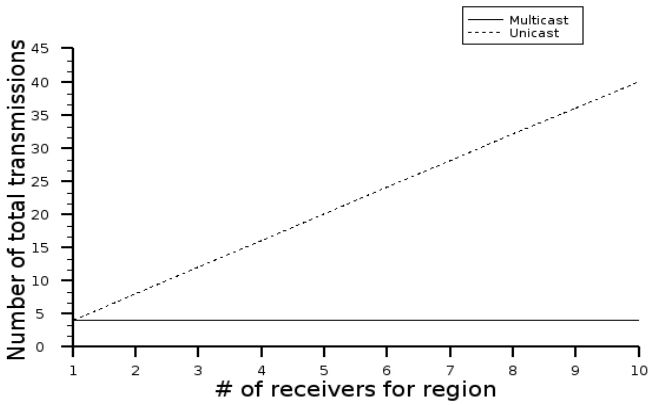


Figure 5. Number of accesses to the links varying the number of receivers per region.

The next results concern how Unicast and Multicast transmissions affect the buffer size (in terms of maximum number of packets that can be memorized) of IPN nodes. We assumed that the buffer size is equal for each IPN node. Figure 6 depicts the obtained results for a MFC connection in term of Packet Delivery Ratio (PDR).

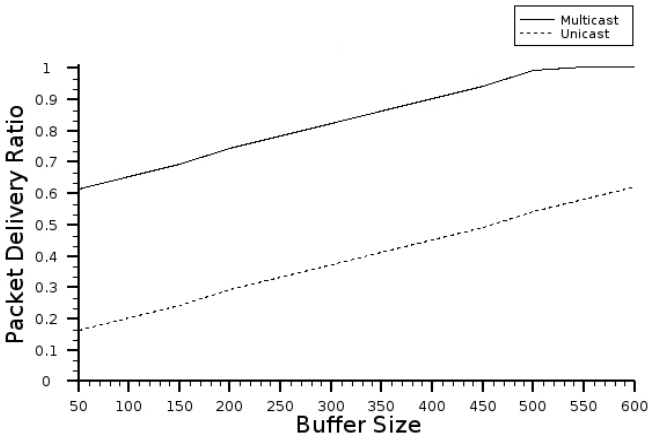


Figure 6. MCF: PDR varying buffer size.

How mentioned above Unicast transmissions foresee that the same information is sent on the channel for all the receivers. From Figure 6 such an issue affects the limits in terms of buffer size more in Unicast approach than in Multicast ones, clearly. On the other hands whether the buffer size is increased then the PDR also increases. From a buffer size equal to 500 packets (i.e. 83,4% of the overall forwarded traffic) there are not loss due to congestion of the buffers, considering multicast transmission.

In this case, the bottlenecks in the nodes MO1 and LO1 are the main reason of packets loss. Therefore, also in this case a Multicast approach improves the performances in IPN network with respect to Unicast transmission. Moreover, it is worth noting that these results have been obtained by considering unreliable traffic only. This means that a packet is removed from the buffer as soon as it is sent on the radio link. In case of reliable Multicast transmissions the propagation

delays on the links have to be taken into account; they clearly can get worse the performance showed in Figure 6 in both Unicast and Multicast approach, but affecting Unicast Transmissions significantly.

Future activities are aimed to improve the performance of Multicast Transmission in IPN network implementing DTN paradigm. In particular, the research activity will deal with the following issues: *(i)* definition of procedures for notifications and registration/de-registration of multicast groups; *(ii)* definition of multicast routing protocols that utilize models based on both tree or mesh topologies, in order to minimize the path length between source and destinations and to increase the probability that the bundle is delivered to as many destination nodes as possible; *(iii)* definition of transport and bundle layers suitable to provide end-to-end reliable connections, defining efficient transmission and retransmission procedures. In this context, the storing functionalities for the store-and-forward policies have to be design in order to guarantee data persistence in DTN nodes also for relatively large time slots. For dealing with that, DTN aggregates data into bundles and stores them in persistent storage of different IPN nodes so that in case of loss of connectivity, the bundles could be retransmitted from the closest storage points rather than from the source node. A key Bundle Protocol innovation is known as Custodial Delivery. The memorization functionality in DTN nodes will be considered as a new network resource that has to be administered and protected. Fundamental open issues in the definition of a new protocol stack are related to these topics. At the moment, the Bundle Protocol specifies the procedures for supporting custodial delivery of bundles destined to unicast applications. However, it does not discuss how Custodial Delivery should be provided for bundles destined to multicast groups (multicast bundle). There is a strong motivation for using custodial multicast in IPN to preserve the already-scarce resource of bandwidth during transmission and retransmission procedures [3].

c) Data Link and Physical Layers

Data Link Layers protocols of the IPN node include functionalities concerning the medium access control (MAC) and error control functions. Also in this case, advanced network control features need to be considered and they are aimed at optimizing the utilization of IPN channels. For both Long and Short-haul physical layers, specific solutions will be studied in terms of bandwidth/power efficient modulations and low complexity channel codes with high coding gain. Waveforms design and the exploitation of Ultra WideBand (UWB) systems needs to be considered with the goal to reduce the complexity of the system and the sensitivity to IPN channels' non-linearity [8].

Also space physical layer solutions that exploit Extremely High Frequency (EHF) bands can be taken into account. EHF employment, in particular the W-band [7], represents an answer to the needs of IPN links: the saturation of lower frequency bands, the growth of data-rate request and the reduction of mass and size of equipment. Considering that the main disadvantage of the use of W-band frequencies is the atmospheric attenuation, the benefits of its employment could be fully exploited in deep space channels where the atmosphere is absent. The reduced antenna size due to the use of higher frequencies represents a further advantage of this choice.

Two important factors that should be considered when dealing with the physical layer design are: antenna pointing and energy consumption. The reduction of antenna size has a positive impact on the pointing subsystem. On the other hand, the design of the physical and link layers should constrained in terms of QoS metrics and should be optimized in terms of energy efficiency.

d) Convergence Layers

Convergence Layers, both Higher and Lower, and IPN Network Control approaches concern another group of innovative solutions, envisaged in this work, which needs to be developed. As previously said, the action of the Higher Convergence Layer is to offer a common interface to the transport layers (space and standard). The Lower Convergence Layer will offer a common interface towards data link and physical layers and vice versa and it will offer innovative control functions in terms of selection of the opportune lower layer stack (e.g., vertical handover) by considering the situation in which the IPN Node operates (long- or short-haul network segment).

e) Network Controls

In order to smooth the effect of the intrinsic heterogeneity of the IPN network, adaptive mechanisms [13], based on the cross-layer principle [2], are needed. It means that appropriate solutions are necessary to harmonize each single layer solution and jointly optimize the capabilities of IPN Node layers. For example, the transport and network protocol parameters need to be dynamically tuned in dependence on the channels and network status, which is unpredictable, and on the basis of blackouts, which are predictable due to the knowledge of the IPN Network Elements' orbits (for satellites and orbiters) and trajectories (in the case of spacecrafts and probes).

The same concept holds true for all protocol layers, also with respect to the position of the IPN Node within the IPN topology. Figure 7 reports the envisaged main blocks concerning both data and control planes and synthetically indicates the main envisaged functions of each control component.

The chosen DTN paradigm, and the developed protocol solutions jointly used with novel network control procedures will allow the optimization of the networking and communication mechanisms of the IPN Node so guaranteeing a reliable and efficient communication process over the IPN Internet.

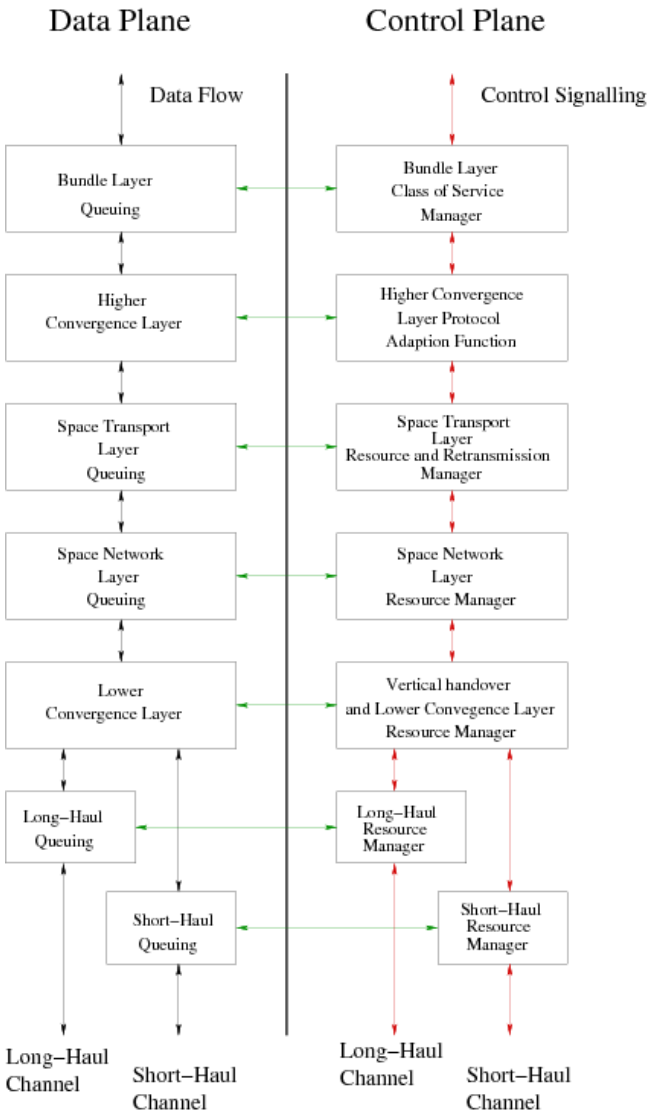


Figure 7. IPN Node Data and Control Planes.

In the set of possible Network Controls, a partially unexplored solution concerns the Link Selection strategies based on the exploitation of the Bundle Layer of the DTN paradigm. In more detail, Link Selection techniques, also called Congestion Aware Routing, have been proposed in [4] where the mathematical framework has been formalised. It has been taken as example in this paper.

In synthesis, the approaches proposed allow selecting a forwarding link, among the available ones, by optimising one or different metrics, simultaneously. In fact, in this paper, the optimization of one metric has been considered: the Bundle Buffer Occupancy (BBO). The Bundle Buffer Occupancy is the ratio between the number of bundles stored in the bundle layer buffer and the maximum size of the buffer itself. The evaluated Link Selection technique is based on its minimization.

As previously introduced, performance analysis has been conducted by taking network topology depicted in Figure 3 as reference and by considering the bandwidth capacities and propagation delays reported in the analysis of Section IV. All the link blackouts, due to IPN node movements, have been also included in the simulations whose results have been reported in the following. Moreover, each node implements a bundle layer buffer size equal to 400 bundles. Constant Bit Rate (CBR) traffic sources are considered: they are kept active for 50 s each hour of simulation and generate data bundles of 64 Kbytes at rate of 1 bundles/s, yielding 512 Kbit/s. Furthermore, in this case, the traffic sources have been set on the planetary regions, and in particular the traffic sources are the nodes MS1 and MS2 from the remote planet, LS1 and LS2 from the Moon. They send data over Earth to ET1, ET2, ET3 and ET4, respectively, which are set as receivers. The simulation duration was of 7200 s (2 hours out of 24, which is the duration of the analysis proposed in Section IV) for each test carried out by *ns-2* simulations.

The proposed results concern a macroscopic analysis of the Link Selection method's performance. It looks into performance provided by the whole network and, in this view, two metrics have been considered: Bundle Loss Rate (BLR) and Data Delivery Time (DDT) coherently with [4]. The first is defined as ratio between the number of received and of transmitted bundles. The second accounts for the time interval required to complete the data delivery to destinations. It is possible to observe, in Figure 8, the Bundle Loss Rate (BLR %) performance for each Flow where Flow1 is the data flow between LS1 to ET1, Flow2 is the data flow between LS2 to ET2, Flow3 is the data flow between MS1 to ET3 and Flow4 is the data flow between MS2 to ET4. The BLR measured highlights quite effective results. This means that a Link Selection Control (or Congestion Aware Routing) allows reaching good network performance also in challenging network as the IPN ones. In more detail, Flow1 is privileged with respect to the others. Actually, it is mainly due to the simulated period: in the first 2 hours, out of 24, the link blackouts have penalized Flow4 and, partially, Flow2 and Flow3. Moreover, Flow3 and Flow4 experience very low link capacities over the IPN network due to the very high distance between Mars and Earth.

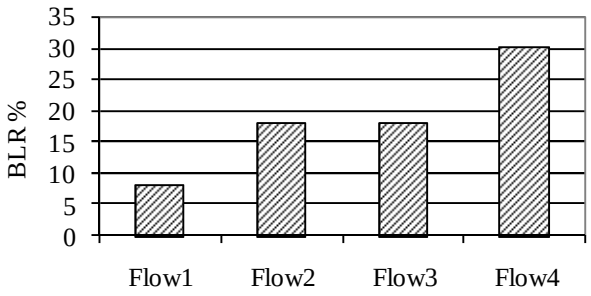


Figure 8. Bundle Loss Rate [%]

On the other hand, as far as Data Delivery Time (DDT) is concerned, it can be observed from Figure 9 that the Link Selection solution offer satisfactory performance. The shown DDT can appear very high but the enormous propagation delays and the very small available link capacities do not allow better performance. It is obvious in particular in case of transmissions from the remote planet (Mars in Figure 3): they require almost the overall time that has been simulates (about two hours). Transmission from the Moon requires about 260 [s] in average.

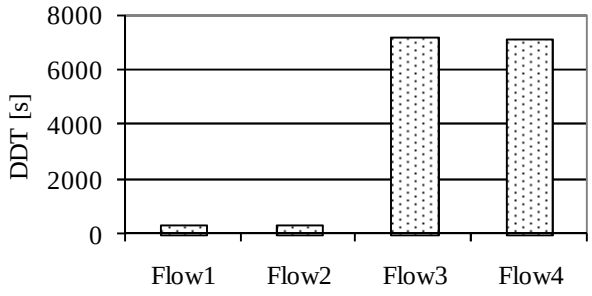


Figure 9. Data Delivery Time [s]

However, from the introductive evaluation proposed, it is worth noting that the proposed control technique have promising performance. This opens the doors to future extensions and investigations that will analyse in-depth the performance of the Link Selection over the considered network architecture.

VII. CONCLUSIONS

From the proposed IPN network analysis and the envisaged IPN node architecture, it appears clear that the technological challenges described in this paper are of great interest on a basic research perspective, and, simultaneously, let the space communications sector be strategically capable to provide future competitive services and solutions. The presented work opens the doors to new communications and networking challenges, which are the ones of the so called InterPlaNetary (IPN) Internet. In particular, the described innovative protocol solutions jointly used with novel network control procedures will allow the optimization of the networking and communication mechanisms of the network's nodes so guaranteeing a reliable and efficient communication process over the IPN Internet.

These solutions will be the object of ongoing and future research that will be developed as extensions of this work, which represents an introductive overview of them.

REFERENCES

[1] G. Araniti, I. Bisio, and M. De Sanctis, "Towards the Reliable and Efficient Interplanetary Internet: a Survey of Possible Advanced Networking and Communications Solutions", First International Conference on Advances in Satellite and Space Communications, SPACOMM 2009, Colmar, France, July 20-25, 2009.

[2] I. Akyildiz, O. E. Akan, C. Chen, J. Fang, and W. Su, "The State of the Art in InterPlaNetary Internet," IEEE Communications Magazine, July 2004 pp. 108–118.

[3] S. Symington, R. C. Durst, and K. Scott, "Custodial Multicast in Delay Tolerant Networks" Consumer Communications and Networking Conference, 2007. CCNC 2007. 4th IEEE Jan. 2007 Page(s):207 - 211.

[4] I. Bisio, T. de Cola, and M. Marchese, "Congestion Aware Routing Strategies for DTN-based Interplanetary Networks," IEEE GLOBECOM 2008, New Orleans, LA, USA, Nov.-Dec. 2008.

[5] V. Cerf, et. al., "Protocol Specification", IETF RFC 4838, experimental, April 2007.

[6] M. Lo, "Satellite-Constellation Design", IEEE Computing in Science and Engineering, vol. 1, no. 1, pp. 58-67, January 1999.

[7] E. Re, M. Ruggeri, V. Dainelli, M. Ferri, , "Millimeter Wave Technology for Moon and Mars Exploration", IEEE Aerospace Conference 2008, Big Sky (MO), 1-8 March 2008.

[8] Y. Kunisawa, H. Ishikawa, H. Iwai, and H. Shinonaga, "Satellite Communications using Ultra Wideband (UWB) Signals," Proceedings of the International Symposium on Advanced Radio Technologies (ISART 2004), March 2-4, 2004.

[9] S. K. Burleigh, M. Ramadas, and S. Farrell, "Licklider Transmission Protocol - Motivation," RFC 5325, Sept. 2008.

[10] G.K. Noreen, et al., "Integrated network architecture for sustained human and robotic exploration," IEEE Aerospace Conference 2005 (Big Sky, Montana), 5-12 March 2005.

[11] Digital Video Broadcasting (DVB), Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, ETSI EN 302 307, v1.1.2, 2006.

[12] M. De Sanctis, et al., "Flower Constellation of Orbiters for Martian Communication", IEEE Aerospace Conference 2007, Big Sky (MT, USA), March 3-10, 2007.

[13] C. Peoples, G. Parr, B. Scotney, and A. Moore, "A Reconfigurable Context-Aware Protocol Stack for Interplanetary Communication," In proc. IWSSC 2007, Salzburg, Austria, September 2007, pp. 163-167.

Circuit Analysis and Simulations through Internet

Jiří Hospodka
Department of Circuit Theory
Czech Technical University in Prague
Prague, Czech Republic
Email: hospodka@fel.cvut.cz

Jan Bičák
ASICentrum
a Company of the SWATCH GROUP
Prague, Czech Republic
Email: Jan.Bicak@asicentrum.cz

***Abstract*—This paper presents an application suitable for analysis of electric and electronic circuits through internet. The application is based on PHP scripts and uses SpiceOpus and Maple program with special library PraCAn as a computation engine. Continuous-time linear and nonlinear circuits as well as periodically switched linear circuits can be analyzed. Results can be obtained in symbolic or semisymbolic form for the case of linear circuits analyzed by Maple with PraCAn. Description of the circuit can be entered through a graphical schematic editor. It is a Java applet for scheme drawing and netlist creation. The possibilities of the application are demonstrated on a number of circuit analyses. The whole system was developed at the Department of Circuit Theory, for research and teaching support.**

***Keywords*-web-based application; simulations; analysis.**

I. Introduction

Nowadays, many systems for circuit analysis are available. Conventional programs like PSpice®, Micro-Cap, WinSpice [3], [4], [5], etc. are single-purpose programs. Evaluation versions of these programs are frequently used for teaching support. Most of them solve the task only numerically, hardly any program makes symbolic or semisymbolic computing possible. Its use is connected with installation of mentioned software to the user computer. In cotrast to this a web-based system offers the advantages of open and remote system for circuit analysis. These systems combine rich client technology, and circuit simulation, and provide convenient user interface for simulation capability. Web-based simulation environments, combining distance education, group training and real-time interaction, can serve as a good approach. The web-based virtual laboratory system for electronic circuit simulation (ECVlab) [7], trainer for electrical circuit analysis [8] or application [11] can serve as examples. Pages [9] based on C++ CGI Toolkit in Ch [10], provide interactive web-based calculation (computation of mathematical formulas in C expression, complex, matrix computation, and grade point average calculation), 2D/3D plotting, numerical analysis (analysis of linear systems, differential equation solving, integration, non-linear equations, Fourier analysis), OpenGL graphics and control system design and analysis (analysis of continuous-time or discrete-time linear time-invarient (LTI) control systems described by particular transfer function). Special application called "Remote Wiring and Measurement Laboratory" (RwmLAB) is described in [14]. RwmLAB is intended to address real-time remote wiring of electrical circuits and real data acquisition over the Internet instead of using simulated data. Simulations through internet using Spice (Spice internet Packages – SIP) were introduced in [11], [12], [13], where internet serves as a graphical user interface for the program simulator – Spice. It is based on CGI (common gateway interface) scripts, PERL and PHP scripting language. Similar application is also the system for electric filter design. Web-based application [15] uses Maple™ [6] as a computational engine with special package Syntfil [16] developed by authors. However synthesis of electric circuit is performed instead of circuit simulations.

The application presented in this article is based on client-server concept [17], which uses special simulation program on the server side. However the described system offers better user interface and greater scope of simulations including symbolic analysis and analysis of periodically switched linear (PSL) circuits, i.e., circuit with switched capacitors (SC) or switched currents (SI). Representation of circuit description is graphical using schematic editor like Java applet [1].

Our motivation was to create application for circuit analysis with a great potential. This goal can be realized by an easy-to-use web-based application, which is available for wide range of users. It can combine several simulation engines, can be acomlished using any computer with internet connection and can be easily administrated and updated, because it runs on a server.

The paper is divided into the three main parts. Programs used for analysis are described first. Introduction of application description including schematic editor follows. Examples of circuit analyses are shown to demonstrate usability of the application and its facilities at the end of the paper.

II. Programs for Analysis

A forementioned application uses client-server conception where programs are installed and run on the server side. It is necessary that programs on the server side have to enable batch processing for easy operation through control scripts. Circuit analysis is provided by two main programs. The numeric analyses are performed by SpiceOpus, while semisymbolic and symbolic analyses including switched circuit are powered by Maple program with special package named PraCAn.

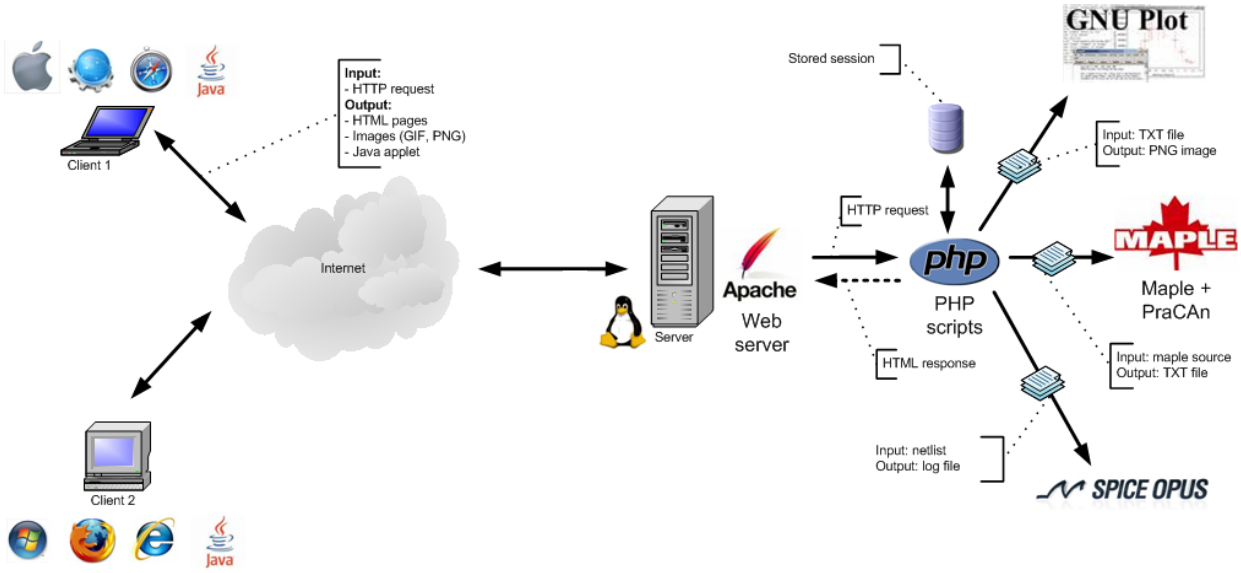


Fig. 1. Principle of the client-server concept used for the interface realization.

A. SpiceOpus

SpiceOpus is a circuit simulator with optimization utilities [18]. It is a recompilation of the original Berkeley's source code [2] for Windows and Linux operating systems. Georgia Tech Research Institute's XSpice mixed-mode simulator was added to the Berkeley code. The simulator includes an interpreted programming language called Nutmeg, which allows interactive Spice sessions. The program compilation is powerful enough to support the mentioned application.

B. PraCAn Package

PraCAn package is a library of functions for Maple, which facilitates the symbolic and semisymbolic analysis of continuous and discrete-time linearized circuits. PraCAn is the acronym for Prague Circuits Analyzer [19]. The input syntax for circuit description is almost the same as in Spice program. The package contains functions for parsing circuit description (netlist), which enable easy identification of syntax errors. Therefore the package is used for parsing netlist in the application. Next functions are designed for the analysis of continuous-time linear circuits as well as the periodically switched linear (PSL) circuits.

The package originates from SCSyrup package [20], [21] for frequency analysis of idealized SC and SI circuits. The functions used in SCSyrup package are based on the analysis of SC circuits using nodal charge equations. The analysis of SI circuits is also based on nodal equations but the currents are used instead of charges [26]. The package was modified and algorithms for PSL circuits were added to make it possible to analyze SC or SI circuits with real qualities such as switch resistances r_{on} and r_{off}. The package was renamed to PraSCAn (Prague Switched Circuits Analyzer), which uses algorithms based on modeling of periodically switched networks using mixed s-z description [25]. SCSyrup was a table-based package whereas PraSCAn is a module-based package.

PraCAn has been created from PraSCAn package [22], [23], which was completely rewritten and new functions were added for continuous-time circuit analysis. The package preserves all functions of SCSyrup and it can be used as its complete replacement with the same results. The circuit is processed by modified nodal voltage method [26]. Unlike the methods in [24], the presented method for multiphase PSL circuits provides symbolic analysis in frequency domain and closed form solution in time domain. The linear system with the time-varying parameters is modeled by nonstationary transfer functions $K(s,t)$. If the parameters vary periodically (e.g., in SC and SI circuits with externally controlled switches), then the system response contains both continuous and discrete time parts and it can be described by a generalized transfer function GTF [25]. The frequency response of the system is obtained by substitution of both s and z. PraCAn package also contains function for direct calculation of time response. The response can be calculated with respect to real input signal character, i.e., including also the so-called leakage effect if Sample&Hold circuit is not used in the input of the analyzed circuit. This way actual spectrum of the signals can be calculated including frequency response of the circuit with undersampling.

III. Web-based Application

The application is based on www interface, which utilizes www client-server concept. The computation and the interface programs run on the server and a user uses standard www browser (Internet Explorer, Mozilla Firefox, Opera, etc.) as a grafical user interface. This principle is illustrated on the following flow-chart, see Figure 1.

The server runs under operating system Linux. The analysis of the required circuit is solved using SpiceOpus program or

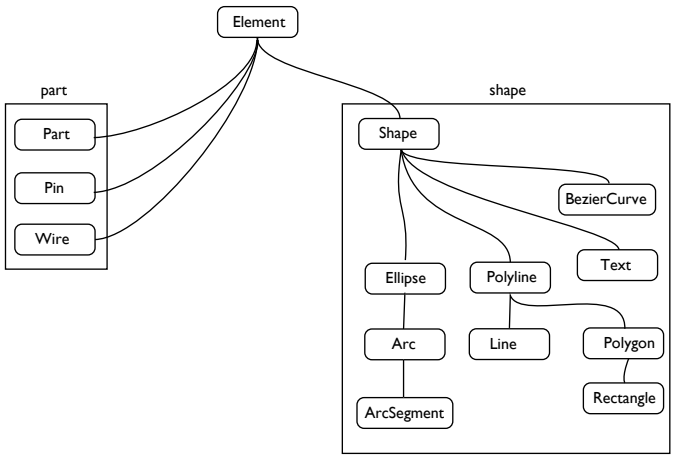


Fig. 2. Elements hierarchy of shematic editor.

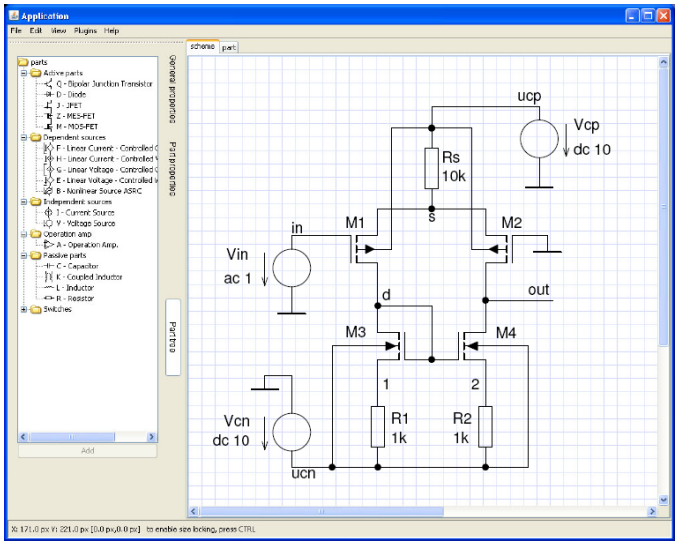


Fig. 3. Graphical interface of shematic editor.

PraCAan package in Maple. It runs using batch-processing, which is necessary for the interface where the programs are called by the PHP scripts [17]. According to client requests the results may be presented by the dynamically created www pages. These pages are provided to the client by means of HTTP server Apache. Described application of circuit analysis was realized according to this model.

Input requests are inserted into the forms of the application using www browser. The program in JavaScript tests the validity of these requests before sending them to the server, where they are tested too. Input files for Maple are generated from the input requests by scripts in PHP and results are saved in separate files. The PHP scripts procese these files and create the structure of dynamic www pages, which are sent to the client. It is necessary to solve many other problems, for example, to distinguish simultaneously connected users, deleting temporary files and directories, etc. These tasks are solved using cookies and session variables (PHP). Other programs are used for additional functions – GnuPlot program is used for graph drawing and typographical system LaTeX for graphical representation of terms. Procedure of calling GnuPlot and LaTeX is analogous to a procedure of calling Maple or OpusSpice.

In contrast to conventional programs it is possible to create nonstandard analyses and different types of parametric analyses, using batch processing managed by PHP scripts using this application. These analyses can be easily defined by user without studying any programm routines as is demonstrated in examples in Section III-B.

A. Schematic Editor – Application for Visual Scheme Editing

Schematic editor is being developed to help users with the creation of electronic circuit schemes by providing means of visual editing. It also allows an export of created schemes using variety of formats from SVG to netlist. The application supports plugins, which are loaded dynamically during the application start. Plugins can receive notification in case of scheme modification, so they can update properly.

The application of the editor is written with regards to object-oriented programming (OOP) techniques using Java 5. Architecture is modular, each module has specialized scope. The modules are divided into:

- *Configuration module* – implements classes, which are responsible for configuration,
- *Elements module* – implements all graphical elements used in application,
- *Gui* – contains classes responsible for graphical user interface (GUI) of application,
- *Manipulation* – contains classes, which implement all supported manipulations,
- *Units* – implements variety of units,
- *VectorEditorEngine* – core module, implements the most basic functionality of the vector editor.

Elements module for example, is represented by data structures, which can be devided into shapes and parts, see Figure 2.

Figure 3 shows the GUI of the editor. It consists of three parts, graphical editor, schematic parts and schematic editor. It is possible to draw basic vector objects (line, rectangle, polygon, Bezier curve, ellipse, etc.) in basic grafical editor. On this basis elements the editor was created where the schematic parts can be formed – to define shape, number of connecting pins and their location, part properties, . . . The final part of the editor is the schematic editor, which consits of a menu, several toolbars and drawing pane used for drawing. All basic parts are defined and prepared in properties toolbar. The grapcical elements can be used together with defined parts to create the circuit scheme in the main (schematic) editor. It is possible to choose one of the required parts and use it for circuit diagram creation in the schematic editor. After placing of the selected part a dialog appears for value or model definitions. The application is connected with an interactive catalogue, where the models are defined. After wiring the scheme the netlist is created and transmitted to the web server for analysis.

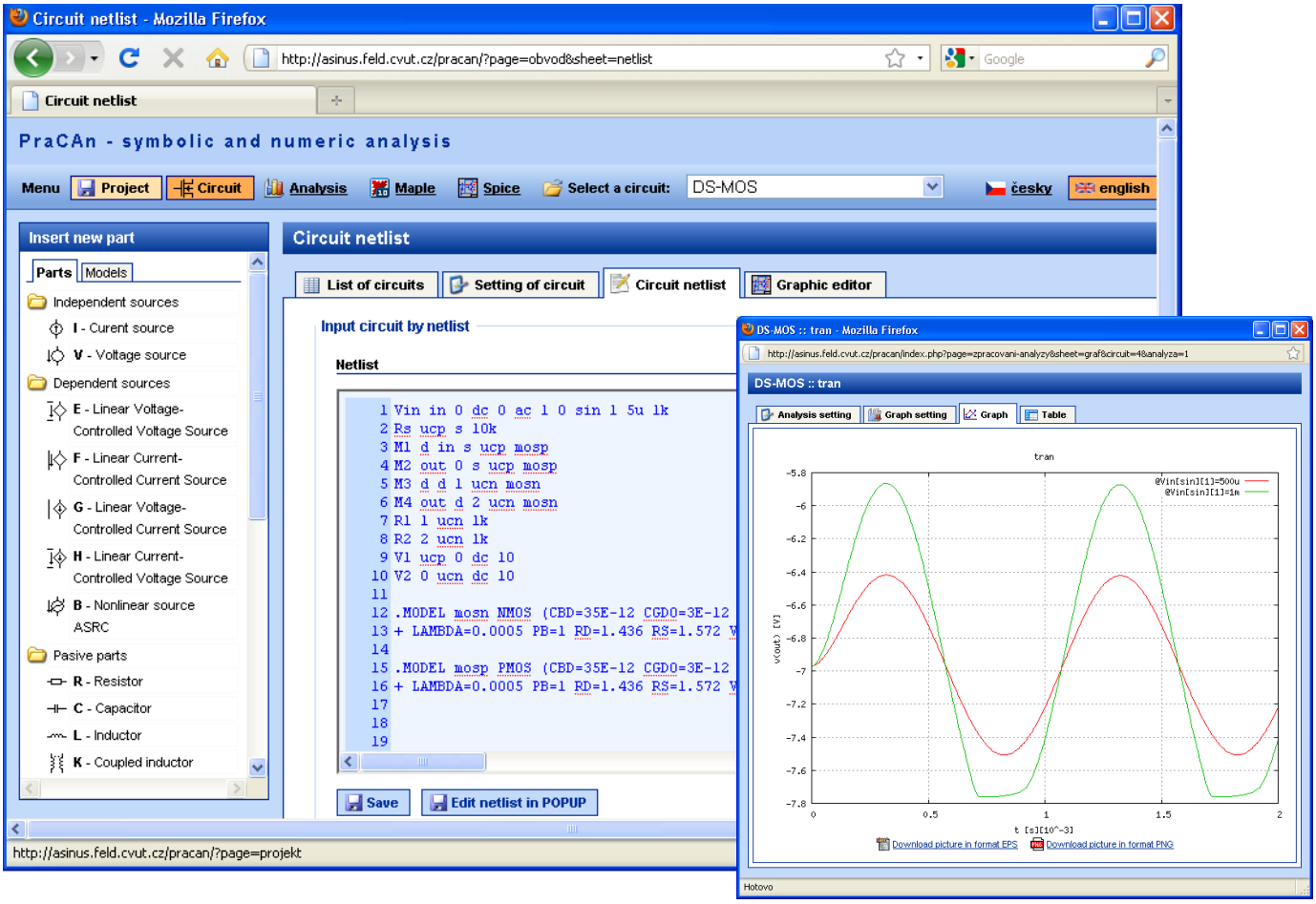


Fig. 4. Application page with inserted netlist of a circuit and window with the time response plot.

Besides the circuit netlist the scheme can be saved in a native format of the editor or exported into graphical format SVG or EPS.

User starts the editor as an applet included in the web application. It is necessary that the Java runtime environment has been installed on the client computer. The second way of starting the editor is by calling a special application instaled on the user computer.

Connecting of schematic editor as an applet with www pages of application is made according [28], [29].

The applet is included at the level of class `cz.cvut.fel.schematicEditor.launcher.Applet` in Java. This class implements methods `void init()`, `void start()` and `void stop()` for applet initialization, start and stop. The methods come from JAplet class: `String getSession()` for getting of session from editor, `void setSession(String session)` for restoring of saved session and `String getNetlist()` for getting of netlist.

B. Examples of Circuits Analysis

The following figures are screent shots taken during the application running circuit analysis. As examples three circuit were chosen – elementary twin-T-network with parametric part values (Figure 7), a third-order elliptic SC low pass ladder realized by the technique of switched capacitors filter [27], see Figure 12 and current mode circuit as an implementation of a simple serial resonant circuit (see Fig 15) by the technique of the switched transconductances, which is analyzed by PraCAn package in Maple program and difference amplifier with active load in CMOS technology, which is analyzed by SpiceOpus.

Figure 4 shows one of the main pages of the application with imported netlist of circuit form Figure 3. The overlapping window shows the time response of the circuit for two levels of excition as a result of numeric analysis powered by SpiceOpus. The analyses are defined on the main page. The user can define number of analyses, which are displayed into the separated windows. Figure 5 shows a window for basic setting of AC analysis. The result is displayed in “Graph” bookmark, see Figure 6. It is not classical example of frequency response, where arbitrary variable (x-axis) is frequency. It is a result of parametric analysis of AC analysis in one frequency (1 kHz, see Figure 5), where module of output voltage (gain) is calculated versus resistor R2 value. Huge possibilities of parametric analysis are demonstrated. Spectral analysis including total harmonic distortion can be calculated similarly as a special

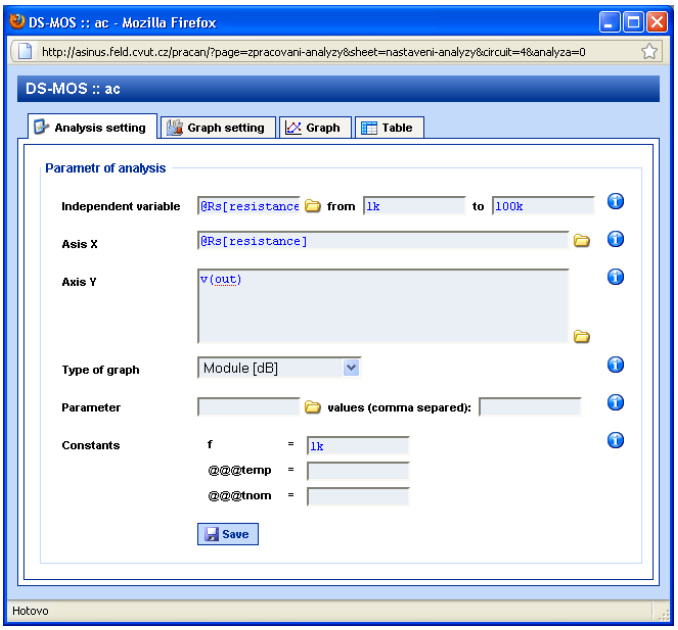


Fig. 5. Definition and settings of AC analysis – frequency response.

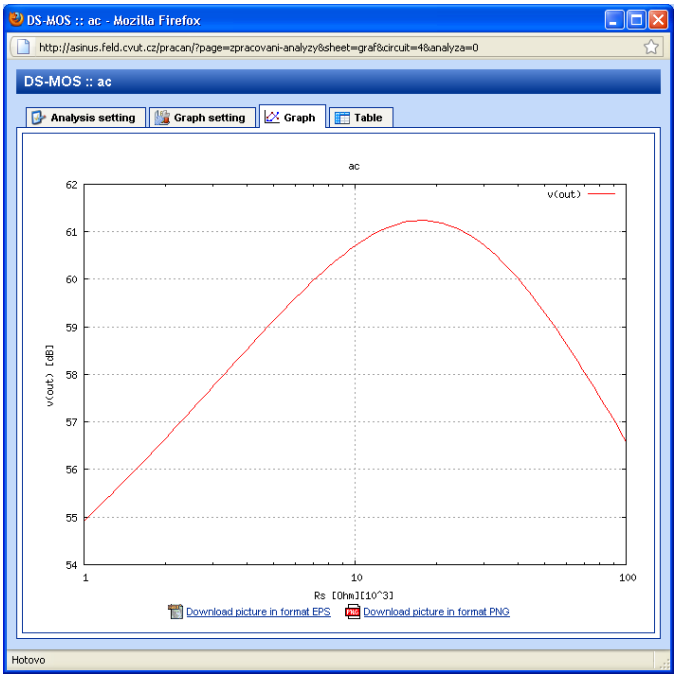


Fig. 6. Result of parametric AC analyses – modulus of output voltage vs. resistor R2 value.

case of transient analysis. A special page exist for ”quick” summary analysis, where operational point can be calculated and displayed together with main circuit parameters (ac gain in one specified frequency or range of frequencies and time response also in one specified time or a time range).

Next examples demonstrate possibilities of semisymbolic and symbolic analyses based on PraCAn package powered by Maple program. Elementary circuit – twin-T-network with parametric part values form Figure 7 is analyzed first.

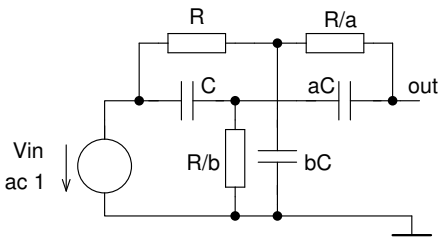


Fig. 7. Twin-T-network.

Figure 8 shows a symbolic frequency analysis of the circuit from figure 7, where transfer function is calculated. The result is displayed in basic text form (for clipboard copying) and also in graphic form (using LaTeX). Next Figure 9 shows the benefit of Maple program implementation – all results can be mathematically treated in bookmark ”Maple calculation”. Resonant frequency (`w0`), transfer function in resonance (`P0`), slope of phase characteristic in resonance and its evaluation for given parameter values are calculated in this case. Maple commands and all circuit variables (voltages and currents) can be used for direct calculation in Maple. Syntax of these commands is necessary to know of course. Circuit variables can be easily chosen in the right part of the window (see Figure 9).

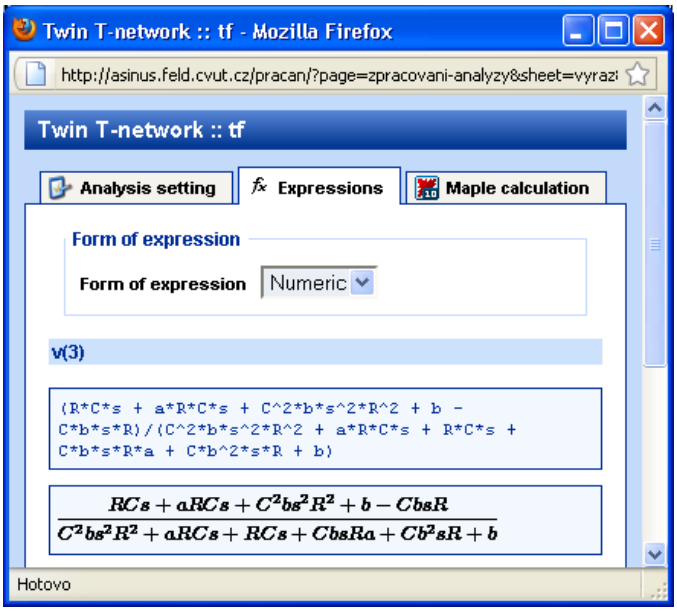


Fig. 8. Result of symbolic analyses of circuit from Figure 7.

The following two figures show frequency analysis of the circuit. The setting of the analysis (Figure 10) is nearly the same as in the previous case. Only the number of circuit constants is higher. In addition one constant (`b`) is chosen as a parameter with number of values (1, 1.9 and 2.1) for parametric analysis. The result of frequency response (modulus together with phase) is shown in Figure 11.

The last two examples demonstrate possibilities of discrete-time circuit analyses. The first circuit presents a third-order

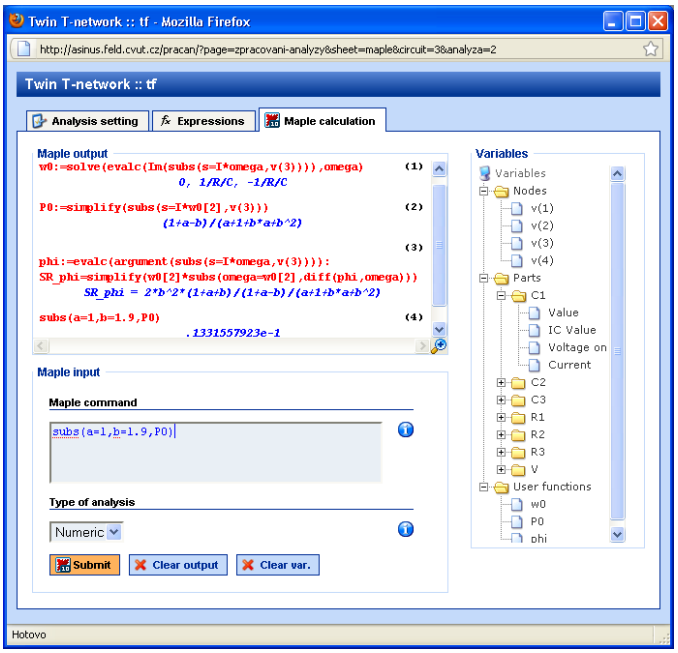


Fig. 9. Page for direct computation in Maple.

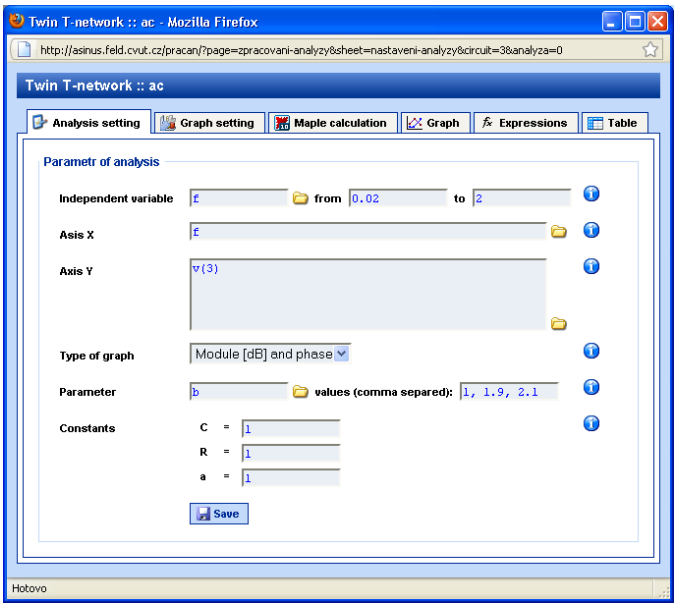


Fig. 10. Setting of frequency analysis of the circuit from Figure 7.

elliptic lowpass filter realized by switched capacitor technique (Fifure 12) and the second one is a resonant circuit realized by the technique of switched currents, in this example the so-called technique of switched transconductances (Fifure 15). The results of analysis of circuit from Figure 12 is displayed in the following figures. Figure 13 represents the result of parametric AC analysis for two values of switching frequency f_c. This way the filter can be tuned as is seen from the figure. The overlapping window shows the poles and zeros of computed filter transfer function in $\mathcal{Z}$-domain. Figure 14 shows results of transient analysis of this SC filter excited by

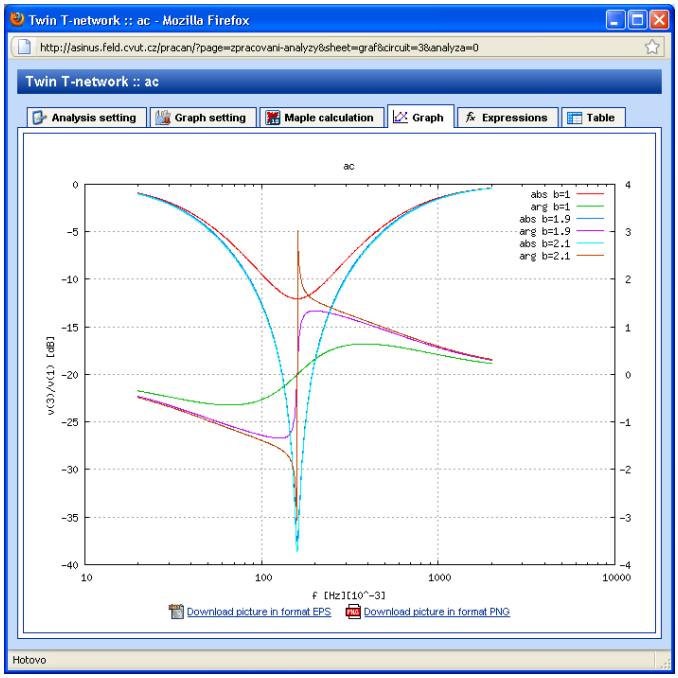


Fig. 11. Parametric frequency response (modulus and phase characteristic).

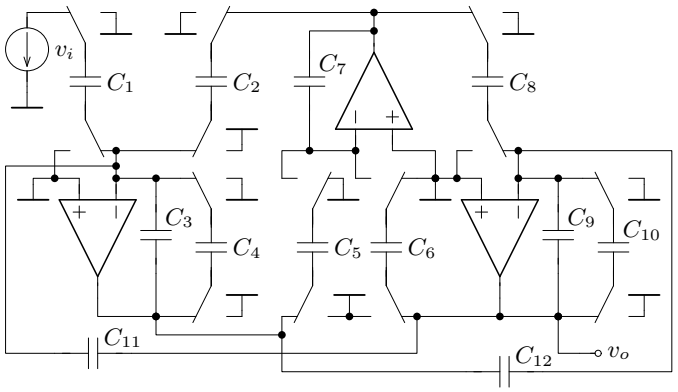


Fig. 12. A third-order elliptic SC lowpass filter.

sinusoidal signal. One can see that input signal leaks directly to output signal because Sample&and Hold circuit is not used in the filter input and direct signal way exists from input to output.

A current mode switched circuit is analyzed in the last example. It is an implementation of a simple serial resonant circuit (see Fig 15) by switched transconductances, as was noted above. The symbolic analysis is provided first, see Figure 16. Using Maple direct computation interface, the input ”impedance” in node 2 in the first phase is calculated as is shown in Figure 17. Then transformation of the impedance from z to s domain is done by means of inverse backward Euler transformation (BD). The input impedance corresponds to impedance of the RLC series resonant circuit, which is evident from expanded result.

The application enables analysis of number of circuits and their management. Different analysis can be created for each

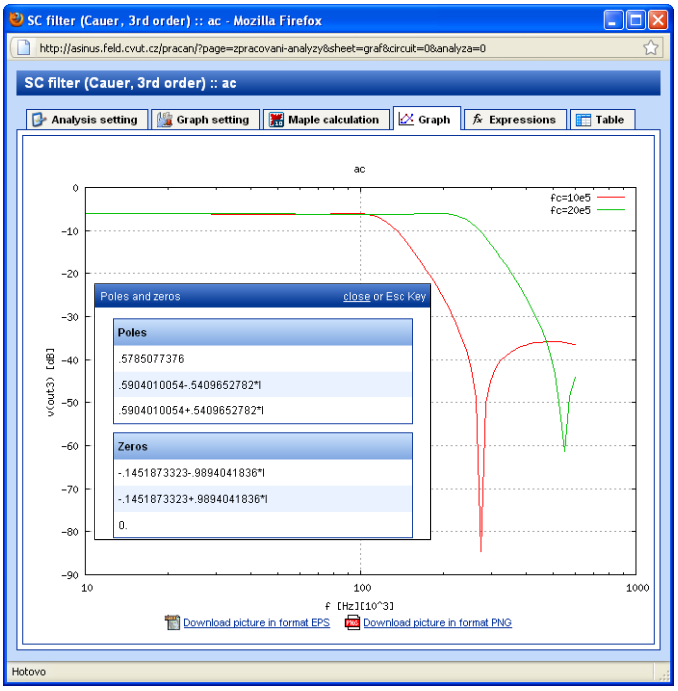


Fig. 13. Frequency response of SC filter and window with pole and zero calculation of transfer function in $\mathcal{Z}$-domain.

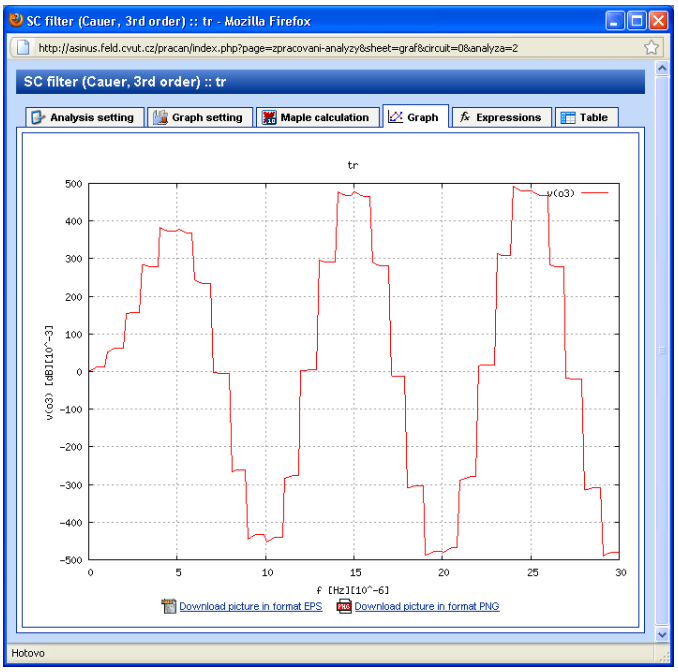


Fig. 14. Result of transient analyses of SC filter.

circuit. The circuit (scheme, netlist, setting, . . .), analysis definitions and results can be saved to the project, see Figure 18. User can save complete work and present it this way.

IV. Conclusion

The web-based application has been created to enable analysis of electric and electronic circuits for a wide range of

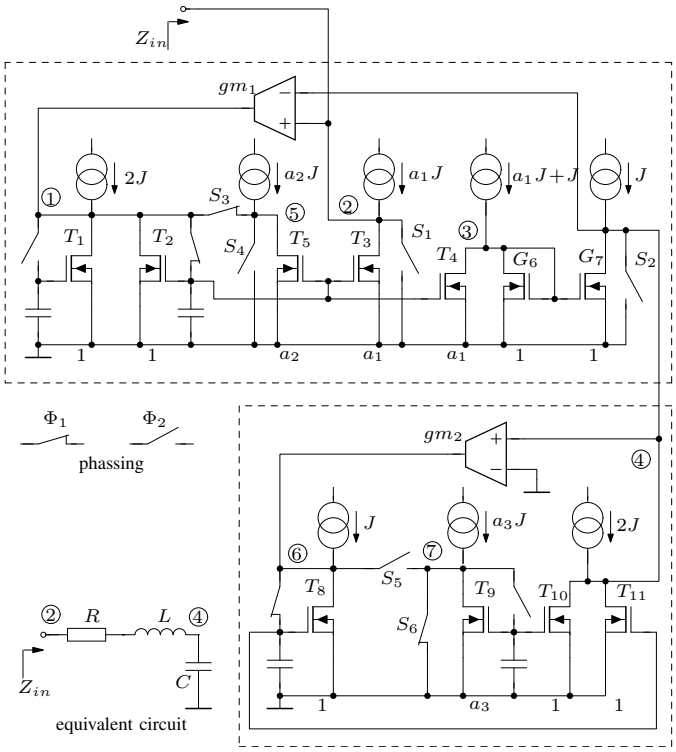


Fig. 15. Realization of RLC circuit by switched transconductances.

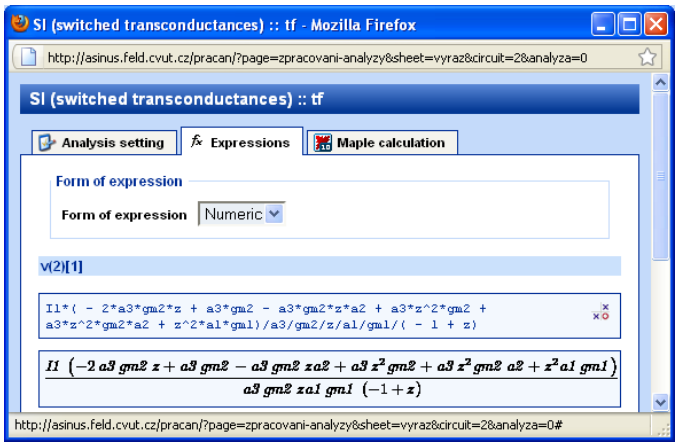


Fig. 16. Result of symbolic AC analysis of the circuit from Figure 15 – voltage in node 2 in the first phase.

users. Designed system combines technologies as rich as client technology, PHP, Java, and circuit simulation, and provides the user with vivid interface, convenient operation and powerful simulation capability. The application uses facilities of SpiceOpus program for numeric analysis and PraCAn package in Maple program for symbolic and semisymbolic analysis of continuous-time as well as periodically switched linear circuits. No known web-based system offers such range of capabilities.

Operating of the interface is very easy. Circuit description can be entered using graphical interface of schematic editor. All pages of application are supplemented by interactive help.

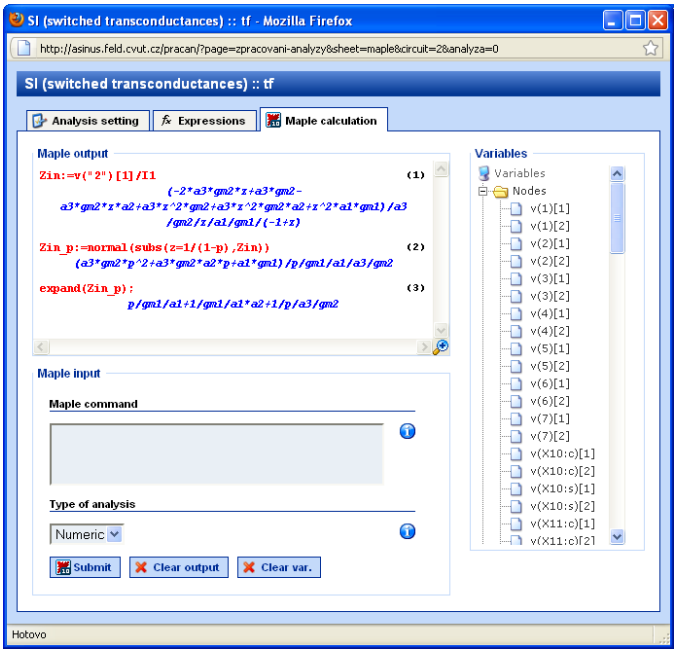


Fig. 17. Page with direct Maple interface for computation and result treatment.

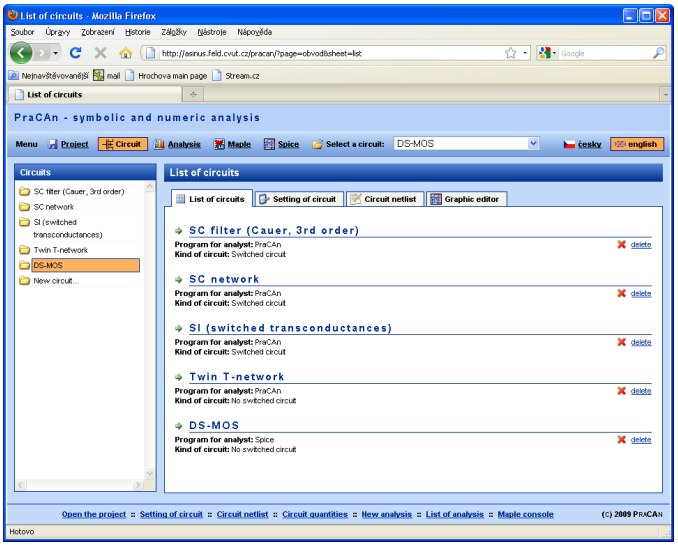


Fig. 18. Page for circuit management.

User can use the interface without any manual or study of syntax. The analysis can be very simply created and modified. It is possible to create number of different analysis of one or more circuits in one project. Results of the analyses can be displayed together and they are simultaneously recompiled (redisplayed) if the netlist of the corresponding circuit is changed. It represents with the symbolic analysis very good tool for engineers and students dealing with electronics.

The application was created especially for teaching support on the Faculty of Electrical Engineering, Czech Technical University (CTU) in Prague. The easy-to-use web based interface will aid analysis of electric and electronic circuits without any program installation and without learning of any command syntax. Students make significant learning gain as a result of using this system. Their interest for electric circuits rises thanks to operation capability and potential of the application. Learning process can be well facilitated if tools are widely available, not just in the dedicated laboratories. Web-based simulation environment, combining distance education, group training and real-time interaction, can serve as a good approach.

The analyses are powered by Maple program whose utilization is restricted by license. This is the reason why the application is not free for all users. Nowadays the system is open from CTU domain; nevertheless it can be open for others who meet the license requirements.

Acknowledgment

The work has been supported by the grant of Ministry of Education, Youth and Sports No. 2388F1d, grant GA102/07/1186 of the Grant Agency of the Czech Republic and by the research program No. MSM6840770014 of the CTU in Prague.

References

[1] J. Hospodka and J. Bičák, *Web-based Application for Electric Circuit Analysis*, Proceedings of The Fourth International Multi-Conference on Computing in the Global Information Technology [CD-ROM]. Los Alamitos: IEEE Computer Society, 2009, p. 157-160, ISBN 978-0-7695-3751-1.

[2] *SPICE – general-purpose circuit simulation*, URL: http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE/, 2009.

[3] Cadence Design Systems, Inc., *Pspice – analog and mixed-signal circuit simulator*, URL: http://www.cadence.com/, 2009.

[4] Spectrum Software, *Micro-Cap – schematic editor and mixed analog/digital SPICE circuit simulator*, URL: http://www.spectrum-soft.com/, 2009.

[5] M. Smith, *WinSpice User's Manual*, http://www.winspice.com, 2007.

[6] Maplesoft, a division of Waterloo Maple Inc., *Maple – essential technical computing software*, URL: http://www.maplesoft.com/, 2009.

[7] Y. Ouyang, Y. Dong, M. Zhu, Y. Huang, S. Mao, and Y. Mao, *ECVlab: A web-based virtual laboratory system for electronic circuit simulation*, ICCS – International Conference on Computational Science 2005, Atlanta, Ga, USA, pp.1027-34, ISBN-10: 3540260323.

[8] L. Weyten, P. Rombouts, and J. De Maeyer, *Web-Based Trainer for Electrical Circuit Analysis*, IEEE Transactions on Education, 2009 Volume: 52, Issue: 1, pp: 185-189, ISSN: 0018-9359.

[9] SoftIntegration, Inc., *Web-Based Control System Design and Analysis*, URL: http://www.softintegration.com/webservices/control/, 2009.

[10] SoftIntegration, Inc., *Ch CGI toolkit for CGI programming*, URL: http://www.softintegration.com/products/toolkit/cgi/, 2009.

[11] B. M. Wilamowski, A. Malinowski, and J. Regnier, *Internet as a New Graphical User Interface for the SPICE Circuit Simulator*, IEEE transaction on industrial electronics, Vol. 48, No. 6, 0278-0046/01$ 10.00, December 2001.

[12] B. M. Wilamowski, A. Malinowski, and J. Regnier, *SPICE based Circuit Analysis using Web Pages*, ASEE 2000 Annual Conference, St. Louis, MO, June 18 to 2, 2000, CD-ROM session 2520.

[13] J. Regnier and B. M. Wilamowski, *SPICE simulation and analysis through Internet and Intranetnetworks*, IEEE Circuits and Devices Magazine, Vol. 14, Issue 3, pp 9-12, ISSN 8755-3996, May 1998.

[14] J.A. Asumadu, R. Tanner, J. Fitzmaurice, M. Kelly, H. Ogunleye, J. Belter, and Song Chin Koh, *A Web-based hands-on real-time electrical and electronics remote wiring and measurement laboratory (RwmLAB) instrument*, Proceedings of the 20th IEEE IMTC – Instrumentation and Measurement Technology Conference, Volume 2, 2003, pp: 1032-1035.

[15] J. Bičák, J. Hospodka, J. Vrbata, and P. Martinek, *Design of Electric Filters in Maple and through WWW Interface*, Proceedings of ICECS – The 8th IEEE International Conference on Circuits and Systems, Vol. 3, pp. 1619–1622, ISBN: 0-7803-7058-9, Malta 2001.

[16] J. Hospodka and J. Bičák: *Syntfil - Synthesis of Electric Filters in Maple*, MSW 2004 [CD-ROM]. Waterloo, ON: Maplesoft, a division of Waterloo Maple Inc., 2004.

[17] J. Hospodka and O. Kobliha, *Internet Pages as an Interface between a User and Computing Program*, Digital Communications'03, EDIS Žilina University Publisher, Žilina, 2003, pp. 45-48.

[18] SpiceOpus – SPICE with integrated OPtimization UtilitieS, URL: http://www.fe.uni-lj.si/spice/ 28.3. 2009.

[19] J. Bicak and J. Hospodka, *PraCAn - Maple Package for Symbolic Circuit Analysis*, Digital Technologies 2008, EDIS Žilina University Publisher, Žilina, 2008, ISBN 978-80-8070-953-2.

[20] J. Bicak, J. Hospodka, and P. Martinek, *Symbolic Analysis of SC Circuits in Maple*, ECCTD '99, pp. 1079-1082, Torino: Politecnico di Torino.

[21] J. Bicak, J. Hospodka, and P. Martinek, *Analysis of SI Circuits in MAPLE Program*, Proceedings of the 15th European Conference on Circuit Theory and Design ECCTD'01, Helsinki: Helsinki University of Technology, 2001, vol. 3, pp. 121-124, ISBN 951-22-5572-3.

[22] J. Bicak and J. Hospodka, *PraSCAn – Maple Package for Analysis of Real Periodically Switched Circuits*, Maple Conference 2005 Proceedings. Waterloo, ON: Maplesoft, a division of Waterloo Maple Inc., 2005, vol. 1, s. 8-18. ISBN 1-894511-85-9.

[23] J. Bicak and J. Hospodka, *Symbolic Analysis of Periodically Switched Linear Circuits*, SMACD'06 – Proceedings of the IX. International Workshop on Symbolic Methods and Applications to Circuit Design, Firenze, Universita degli Studi, 2006, ISBN 88-8453-509-3.

[24] F. Yuan and A. Opal, *Computer Methods for Switched Circuits*, IEEE Transactions on CAS I, Vol. 50, pp. 1013-1024, Aug. 2003.

[25] D. Biolek, *Modeling of Periodically Switched Networks by Mixed s-z Description*, IEEE Transactions on CAS I, Vol. 44, pp. 750-758, 1997.

[26] J. Vlach and K. Singhal, *Computer Methods for Circuit Analysis and Design*, Van Nostrand Reinhold Company Inc., New York 1994, 2nd Edition, ISBN 0-13-879818-4.

[27] P. V. Ananda Mohan, V. Ramachandran, and M. N. S. Swamy, *SWITCHED CAPACITOR FILTERS Theory, Analysis and Design*, Prentice Hall 1995, ISBN 0-13-879818-4.

[28] E. A. Musayev, *How to connect Java applet, Javascript and HTML form*, URL: http://www.galiel.net/el/howto/jvjvs.html, 2009

[29] *Call Javascript from a Java applet*, URL: http://www.rgagnon.com/javadetails/java-0172.html, 2009

Performance Analysis of Scheduling and Dropping Policies in Vehicular Delay-Tolerant Networks

Vasco N. G. J. Soares
Instituto de Telecomunicações /
U. Beira Interior, Covilhã, Portugal
Polytechnic Institute of C. Branco,
Castelo Branco, Portugal
vasco.g.soares@ieee.org

Farid Farahmand
Department of Engineering Science,
Sonoma State University
California, USA
farid.farahmand@sonoma.edu

Joel J. P. C. Rodrigues
Instituto de Telecomunicações /
Department of Informatics
University of Beira Interior,
Covilhã, Portugal
joeljr@ieee.org

***Abstract*—Vehicular Delay-Tolerant Networking (VDTN) was proposed as a new variant of a delay/disruptive-tolerant network, designed for vehicular networks. These networks are subject to several limitations including short contact durations, connectivity disruptions, network partitions, intermittent connectivity, and long delays. To address these connectivity issues, an asynchronous, store-carry-and-forward paradigm is combined with opportunistic bundle replication, to achieve multi-hop data delivery. Since VDTN networks are resource-constrained, for example in terms of communication bandwidth and storage capacity, a key challenge is to provide scheduling and dropping policies that can improve the overall performance of the network. This paper investigates the efficiency and tradeoffs of several scheduling and dropping policies enforced in a Spray and Wait routing scheme. It has been observed that these policies should give preferential treatment to less replicated bundles for a better network performance in terms of delivery ratio and average delivery delay.**

Keywords- Vehicular Delay-Tolerant Networks; Delay-Tolerant Networks; Scheduling Policies; Dropping Policies; Performance Analysis

I. INTRODUCTION

The delay-tolerant network (DTN) architecture [1] was conceived to support data communication in highly challenged environments characterized by any combination of the following aspects: sparse connectivity, network partitioning, intermittent connectivity, long propagation delays, asymmetric data rates, high error rates, and even the potential non-existence of a contemporaneous end-to-end path. To handle these issues, the DTN architecture introduces a bundle layer, which builds a store-and-forward overlay network above the transport layers of underlying networks [2].

Although DTN architectural concepts were initially proposed to deal with interplanetary connectivity [3], over the last years they have been applied in terrestrial environments over a wide range of application scenarios including underwater networks [4], wildlife tracking networks [5], sparse wireless sensor networks [6], transient networks [7-9], disaster recovery networks [10], people networks [11], and military tactical networks [12].

Vehicular networks are another example of networks that can benefit from the application of the DTN paradigm [13-15]. It is important to note that these networks are characterized by a highly dynamic network topology, and short contact durations, which are caused by the high velocity of vehicles [16, 17]. In addition, limited transmission ranges, physical obstacles, and interferences, lead to connectivity disruption and intermittent connectivity issues [18]. Furthermore, these networks may be partitioned, because of the large distances usually involved and to low node density. Hence, a complete path from source to destination may not exist for most of the time.

The vehicular delay-tolerant network (VDTN) architecture has been proposed to deal with these challenging connectivity issues. VDTN architecture is based on the principle of asynchronous, bundle-oriented communication from the DTN architecture. However, the design of the VDTN network architecture, and its protocol layering, considers an Internet protocol (IP) over VDTN approach, and features an out-of-band signaling approach, with the separation between the control plane and the data plane [15].

The effective operation of a VDTN relies on the cooperation of network nodes to store-carry-and-forward data bundles. In addition, routing protocols may perform bundle replication to discover more possible paths, and thus increase the bundle delivery rate and decrease the delivery delay. However, the problem is that the combination of bundle storage during large periods of time and their replication leads to high storage and bandwidth overhead. As network nodes have limited resources, this may degrade the overall network performance. To tackle this problem, scheduling and dropping policies are used, determining bundle replication order at the contact opportunities, and taking bundle drop decisions when buffer space is exhausted.

This work studies the influence of scheduling and dropping policies on the performance of VDTN networks in terms of the delivery ratio and the delivery delay. This paper extends a preliminary contribution about the impact of scheduling and dropping policies for improving the bundles delivery time on VDTNs [19]. The paper has been extended with the introduction of new scheduling and dropping policies, based on distinct criteria, and the performance evaluation of the proposals through extensive simulation in different scenarios.

The remainder of this paper is organized as follows. Section II presents a brief overview of the VDTNs background, while Section III describes the problem statement, and related work. Section IV proposes the scheduling and dropping policies studied in this work. Section V focuses on the comparative analysis of the proposed approaches and Section VI concludes the paper and points further research directions.

II. VEHICULAR DELAY-TOLERANT NETWORKS

Vehicular delay-tolerant networking has been proposed to address challenged vehicular communications [15]. Although VDTN architecture is based on the DTN store-carry-and-forward model of routing, it presents unique characteristics such as *i)* IP over DTN approach; *ii*) control plane and data plane decoupling; and *iii*) out-of-band signaling.

Figure 1 shows a comparison between DTN and VDTN network architecture layers. As may be seen, DTN architecture introduces a bundle layer that creates a store-and-forward overlay network, allowing the interconnection of highly heterogeneous networks [20]. On the contrary, the VDTN architecture follows another approach based on IP over DTN, placing the DTN-based layer over the data link layer. Like on DTN architecture, bundles are also defined as the protocol data unit at the VDTN bundle layer. However, in a VDTN, a bundle aggregates several IP packets with common characteristics, such as the same destination node or generated with data from the same application.

The DTN store-carry-and-forward paradigm is used to recover from network partitions, and to cope with node sparsity. According to this paradigm, the network nodes keep the bundles on their buffers, while waiting for contact opportunities to forward them to intermediate nodes, or to the final destination. This long-term storage paradigm exploits opportunistic contacts between network nodes that arise with the vehicles mobility, to bring bundles closer and closer to destination.

Another distinctive feature of the VDTN architecture is the separation between control and data planes, as illustrated in Figure 1. The VDTN bundle layer is divided into the following sublayers: bundle signaling control (BSC) and bundle aggregation and de-aggregation (BAD). BSC is responsible for executing the control plane functions, such as signaling messages exchange, node localization, resources reservation (at the data plane) and routing, among others. The signaling messages include information such as, but not limited to, node type, geographical location, route, velocity, data plane link range, power status, storage status, bundle format and size, delivery options, and security requirements, among others. BAD controls the data plane functions that deal with data bundles. These functions include, among others, buffer management and scheduling, traffic classification, data aggregation/de-aggregation, and forwarding.

VDTN uses out-of-band signaling, meaning that the control plane uses a separate, dedicated, low-power, low bandwidth, and long-range link to exchange signaling information. On the contrary, the data plane uses a high-power, high bandwidth, and short-range link to exchange data bundles. While the control plane link connection is always active to allow node discovery, the data plane link connection is active only during the estimated contact duration time, and if there are data bundles to be exchanged between the network nodes. Otherwise, it is not activated. The use of out-of-band signaling procedures is described in [15, 21], and offers considerable benefits because it not only ensures the optimization of the available data plane resources [21], but also allows saving power, which is very important for power-limited network nodes [15, 22].

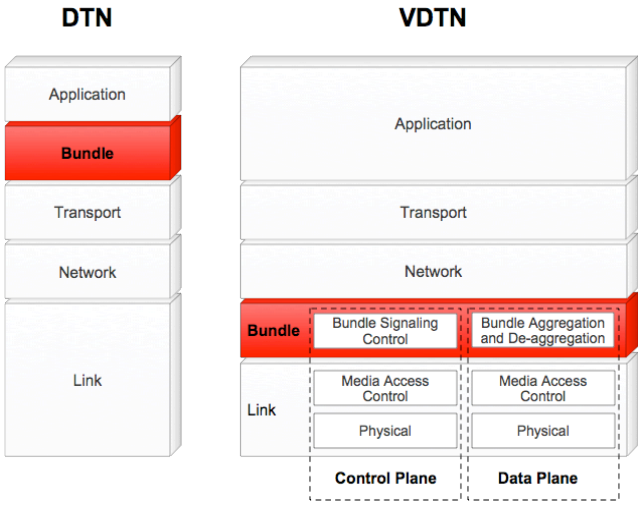


Figure 1. DTN and VDTN network architecture layers.

Figure 2 shows data exchange between two types of network nodes in a VDTN: mobile nodes (e.g., vehicles), and stationary relay nodes. Mobile nodes opportunistically collect and disseminate data bundles. Stationary relay nodes are power limited, fixed devices, which are located at road intersections. These nodes increase contact opportunities in scenarios with low node density. They allow passing by mobile nodes to pickup and deposit data on them. Thus, they contribute to increase the bundles delivery ratio and to decrease their delivery delay [23].

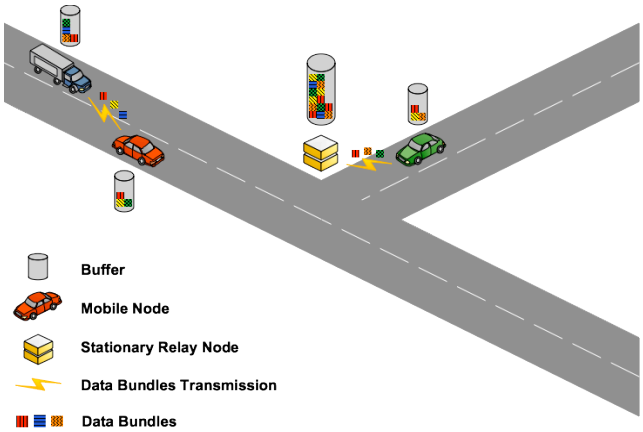


Figure 2. Illustration of VDTN network nodes exchanging data bundles.

III. PROBLEM STATEMENT

VDTNs are characterized by very high node mobility, which results in frequent topological changes and network partition. In these networks, the node density and the mobile nodes mobility pattern have a direct effect over the observed transmission opportunities, contact durations, and inter-contact times.

Figures 3 and 4 illustrate the effect of the mobile nodes (e.g., vehicles) density and mobile nodes velocity in a simulation scenario detailed in Section V. As expected, Figure 3 shows that the number of contact opportunities is directly related to the number of mobile nodes available on a network scenario. Moreover, it allows concluding that the number of contact opportunities grows as the mobile nodes velocity increases. As may be seen, this effect is more pronounced for the scenario with 50 mobile nodes.

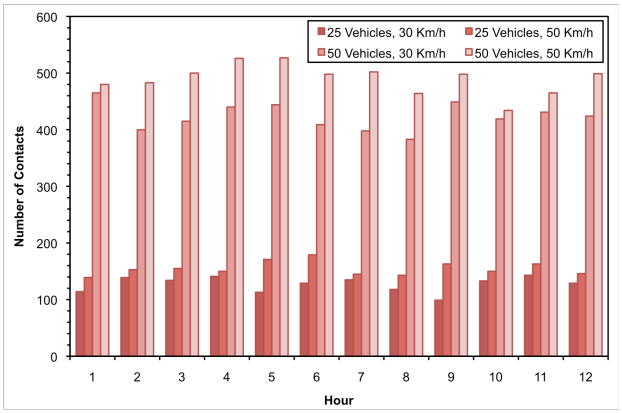


Figure 3. Number of contacts per hour between network nodes, with 25 or 50 vehicles moving at a speed of 30 Km/h or 50 Km/h.

Figure 4 shows that VDTN network nodes register short contact durations, due to the velocity of mobile nodes. As mobile nodes move at a faster speed, more contacts are registered, but the contact duration decreases even more. This impacts bundles transmission, since the available bandwidth is further restricted, which may turnout to be insufficient to transmit all intended bundles.

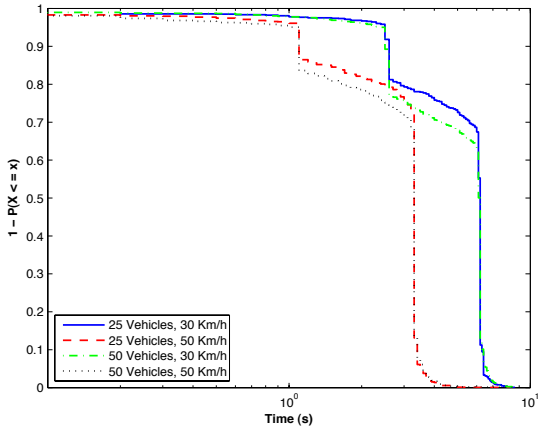


Figure 4. Contact durations with 25 or 50 vehicles moving at a speed of 30 Km/h or 50 Km/h.

Moreover, in such challenged scenarios, long-term storage is often combined with replication-based routing schemes [24]. Spreading multiple copies of bundles to several network nodes improves the delivery rate and/or reduces the delivery latency. However, in a resource-constrained network, these techniques can cause contention for network resources (e.g., bandwidth and storage), and can greatly influence the performance of routing protocols [25-27].

This emphasizes the need for efficient scheduling policies to decide the order by which bundles are transmitted at the brief contact opportunities, and efficient drop policies to decide which bundles are discarded when a node's buffer is full. Although scheduling policies and dropping policies play an important role in improving the overall performance of any DTN-based network, to the best our knowledge, little research has been done in this field.

In [28], Lindgren and Phanse compare the performance of Epidemic [29] and PRoPHET [30] routing protocols when different combinations of queuing and forwarding policies are used. They show that these policies can optimize the limited system resources utilization, leading to performance improvement of the routing protocols, in terms of message delivery, overhead, and end-to-end delay. They also conclude that when bandwidth is limited, it is not enough only to decide what messages should be forwarded, but also the order in which they must be forwarded.

In [31], Zhang *et al.* present an analysis of buffer-constrained Epidemic routing. Simple buffer management policies are evaluated. The authors conclude that with adequate buffer management schemes, smaller buffers can be used without negative impact on the delivery ratio observed with this routing protocol.

In [32], Krifa *et al.* also base their study on Epidemic routing. The authors consider the theory of encounter-based message dissemination to propose an optimal buffer management policy based on global knowledge about the network. This policy can either maximize the average delivery probability or minimize the average delivery delay.

In [33], Erramilli and Crovella observe that it is important to study forwarding and dropping policies independently of each other. The focus of their work is on comparing message prioritization schemes (for transmission or dropping) that do not take into account network information with schemes based on delegation forwarding algorithms [34]. The authors conclude that the latter schemes perform better in terms of delivery rate, delay and cost. Based on their results, the authors also state that forwarding policies have less impact in the network performance than dropping policies.

In [35], Li *et al.* study the impact of buffer management strategies under Epidemic routing. They propose a congestion control mechanism called N-Drop. This policy takes into account the number of times a message has been forwarded, and a threshold related to the size of the buffer, to decide which messages should be dropped when buffer overflow occurs.

IV. SCHEDULING AND DROPPING POLICIES

As part of the resource allocation mechanisms, each VDTN network node must implement some queuing discipline that governs how data bundles are buffered while being carried and waiting to be transmitted. This queuing discipline, illustrated in Figure 5, consists of both a scheduling and a dropping policy. The scheduling policy determines the order in which bundles are transmitted at a contact opportunity. The dropping policy selects bundles to be dropped upon buffer overflow.

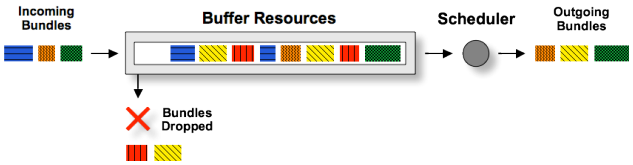


Figure 5. Illustration of a queueing discipline composed by a scheduling and a dropping policies.

This section describes several scheduling and dropping policies, and identifies some variations that can be applied to them. Their performance is evaluated and compared through a simulation study in Section V.

A. Scheduling Policies

The following scheduling policies are considered and studied in this work.

FIFO: FIFO orders bundles to be transmitted at a contact opportunity, based on their arrival time at the node's buffer (based on a first-come, first-served approach).

Random: Due to short contact duration and finite bandwidth, FIFO approach may only serve bundles that arrived first to the node's buffer. To avoid this situation, random scheduling policy selects bundles randomly within a queue.

Remaining Lifetime (RL): Both FIFO and Random scheduling policies do not take into account the time-to-live (TTL) of bundles. TTL is a timeout value that expresses the amount of time that bundles should be stored before being discarded, since they are no longer meaningful. Remaining Lifetime scheduling policy orders bundles based on their remaining TTL. Two variations of this scheduling policy are considered, either (*i*) bundles with smaller remaining TTLs are scheduled to be sent first (RL Ascending Order) or (*ii*) bundles with longer remaining TTLs are scheduled to be sent first (RL Descending Order).

Replicated Copies (RC): This scheduling policy assumes that nodes keep track of the number of times each bundle has been replicated. Hence, two variations of RC policy may be considered. On the first, bundles that have been less replicated are scheduled to be sent first (RC Ascending Order) or, in the second, bundles that have been more replicated are scheduled to be sent first (RC Descending Order).

Both Remaining Lifetime and Replicated Copies scheduling policies need a tiebreaking rule. In the case of RL scheduling policy, a node may store in its buffer more than one bundle with the same remaining lifetime. The same happens with RC scheduling policy, where a node may store bundles that have been replicated the same number of times. FIFO or Random scheduling policies can be applied to tiebreak these cases.

B. Dropping Policies

The following dropping policies are considered in this work.

Drop Head: This dropping policy discards the bundle that has been stored for the longest period of time in the node's buffer, to create available apace for the next incoming bundle.

Random: When a receiving buffer is congested, this dropping policy randomly selects one of the bundles within a queue to be dropped.

Remaining Lifetime (RL): Remaining Lifetime dropping policy selects bundles that get discarded based on their remaining TTL. Two variations of this policy are considered, either (*i*) the bundle with the smallest remaining TTL is discarded first (RL Ascending Order) or (*ii*) the bundle with the longest remaining TTL is discarded first (RL Descending Order).

Replicated Copies (RC): When buffer overflow occurs, the number of times each bundle has been replicated can be used to decide which bundle should be dropped. The following two variations of RC dropping policy may be used: (*i*) the bundle that has been less replicated is dropped first (RC Ascending Order) or (*ii*) the bundle that has been more replicated is dropped first (RC Descending Order).

For the same above-mentioned reasons, a tiebreaking rule must be used for both Remaining Lifetime and Replicated Copies dropping policies. FIFO or Random scheduling policies can be applied to tiebreak.

V. PERFORMANCE ANALYSIS

This section investigates the effect of the above described scheduling and dropping policies on the performance of a vehicular delay-tolerant network. The study was conducted by simulation using a modified version of the Opportunistic Network Environment (ONE) simulator [36]. ONE was modified to support the VDTN layered architecture model proposed in [15]. Additional modules were developed to implement the scheduling and dropping policies. Next subsections describe the simulation scenario and the corresponding performance analysis.

A. Simulation Scenario Parameters

The simulation scenario is based on a map-based model of a part of the city of Helsinki presented in Figure 6. During

a 12 hours period of time (e.g., from 8:00 to 20:00), mobile nodes (e.g. vehicles) move on the map roads between random locations, with random pause times between 5 and 15 minutes. To obtain scenarios with different numbers of contact opportunities we change the number of mobile nodes between 25 and 50 across the simulations. We also vary the mobile nodes average velocity between 30 Km/h and 50 Km/h, to obtain scenarios with different contact durations. Each of the mobile nodes has a 25 Megabytes buffer.

To increase the number of contact opportunities, five stationary relay nodes were placed at the road as may be seen in Figure 6. Each stationary relay node has a 500 Megabytes buffer.

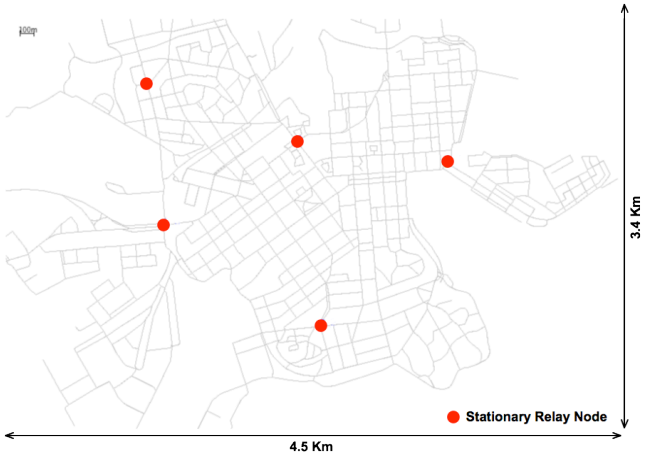


Figure 6. Helsinki simulation scenario (area of 4500×3400 meters), with the locations of the stationary relay nodes.

Data bundles are generated using an inter-bundle creation interval that is uniformly distributed in the range of [15, 30] (seconds), and have random source and destination vehicles. Data bundles size is uniformly distributed in the range of [250 KB, 2 MB] (bytes). Bundles have a time-to-live (TTL) that changes between 30, 60, 90, 120, 150, and 180 minutes, across the simulations, and are discarded when the TTL expires. Increasing TTL leads to having more bundles stored at the network nodes' buffers, and during larger periods of time. Therefore, more bundles will be exchanged between network nodes, and this will also potentially increase buffer overflows. All network nodes use a data plane link connection with a transmission data rate of 4.5 Mbps and an omni-directional transmission range of 30 meters, as proposed in [37].

Spray and Wait [38] is used as the underlying DTN routing scheme. Spray and Wait routing limits the number of bundle replicas (copies), and it assumes two main phases. In the "spray phase", for each original bundle, L bundle copies are spread to L distinct relay nodes. At the "wait phase", using direct transmission, it waits until any of the L relays finds the destination node. This work considers a binary spraying method, where the source node starts with a number of copies N (assuming 12, in this study) to be transmitted ("sprayed") per bundle. Then, at any node A that has more than 1 bundle copies and encounters any other node B that does not have a copy, forwards to B $N/2$ bundle copies and keeps the rest of the copies. When a node carries only 1 copy left, it only forwards it to the final destination.

Performance metrics considered in this study are the bundle delivery probability (measured as the relation of the number of unique delivered bundles to the number of bundles sent), as well as the bundle delivery delay (measured as the time between bundles creation and delivery). We measure the different performance results for the combination of the above-described scheduling and dropping policies, presented at Table I. A designation for each scheduling/dropping policy pair was created in order to improve chart readability in the next subsection. FIFO policy is used as a base-case of comparison with the other proposed policies. The results presented in the next subsection are averages from 12 simulation runs.

TABLE I. COMBINED SCHEDULING AND DROPPING POLICIES

Designation	Scheduling Policy	Dropping Policy	Tie-break
FIFO	FIFO	Head Drop	-
Random	Random	Random	-
RL ASC	Remaining Lifetime *Ascending Order*	Remaining Lifetime *Descending Order*	FIFO
RL DESC	Remaining Lifetime *Descending Order*	Remaining Lifetime *Ascending Order*	FIFO
RC ASC	Replicated Copies *Ascending Order*	Replicated Copies *Descending Order*	RL DESC
RC DESC	Replicated Copies *Descending Order*	Replicated Copies *Ascending Order*	RL DESC

B. *Performance Analysis for a Scenario with 25 Vehicles*

The evaluation study starts with a comparison of the delivery probability registered when 25 vehicles move with a speed of 30 Km/h. Figure 7 shows that when the initial bundles' TTL is lower than 120 minutes, FIFO, Random, RL ASC, and RC DESC policies register similar delivery probabilities. However, when the TTL is great than 120 minutes, RC DESC policy performs much worse than the other policies. This means that scheduling bundles that have been more replicated to be sent first, is not a good option.

Enforcing a RL DESC policy and, therefore, giving preferential treatment to bundles with larger remaining lifetimes, leads to increase the delivery ratio when compared to those policies. Since bundles exchanged between network nodes will have longer remaining lifetimes, this increases their probability to be relayed more times between network nodes, until eventually reaching the destination. This figure shows that when the initial bundles' TTL is equal or lower than 150 minutes, RL DESC increases the delivery probability about 3% for TTL=60min., 5% for TTL=90min., 5% for TTL=120min., 4% for TTL=150min., and 2% for TTL=180min., when compared to the traditional FIFO policy. For a TTL of 180 minutes, RC DESC and FIFO register a similar delivery probability.

RC ASC policy improves these results further. This policy gives preferential treatment to bundles that have been less replicated, using as a tiebreak criterion for bundles that have been replicated the same amount of times, a second scheduling policy - RL DESC. As may be observed, when RC ASC policy is compared to FIFO, it provides up to 3%, 6%, 7%, 6%, 6% and 5% of gain in delivery ratio, respectively, across all simulations.

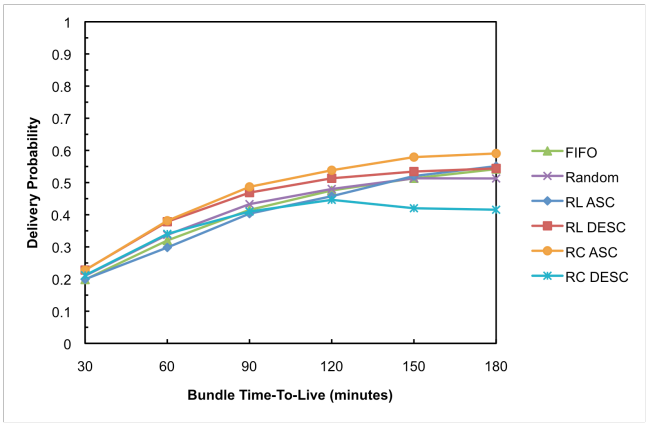


Figure 7. Bundle delivery probability as function of TTL in a scenario with 25 vehicles moving at a speed of 30 km/h.

Figure 8 shows a comparison between average delay and (initial) bundles TTL, for the combinations of scheduling and dropping policies considered. Average delay is an interesting metric, since minimizing the delivery delay reduces the time that bundles spend in the network and, thus, reduces the contention for resources.

As expected, the observed results show that deploying a RL DESC based policy and, therefore, giving preferential treatment to bundles with larger remaining TTLs, decreases the bundle average delay considerably. This policy performs better than the others in this performance metric. On the contrary, the other variant of RL policy - RL ASC - gives preferential treatment to bundles with lower remaining lifetimes, trying to deliver them before expiring. This results in the worst average delays across all simulations.

FIFO criterion based on the order of bundle arrival to the buffer, also leads to longer average delays. When RL DESC policy is compared to FIFO, bundles arrive at the destination nodes approximately 1, 3, 8, 14, 16, and 18 minutes sooner, in average. As a final note to Figures 7 and 8, it can be seen that RC ASC outperforms all the other policies in terms of delivery probability, presenting the second best results in terms of delivery delay, due to its tiebreak criterion.

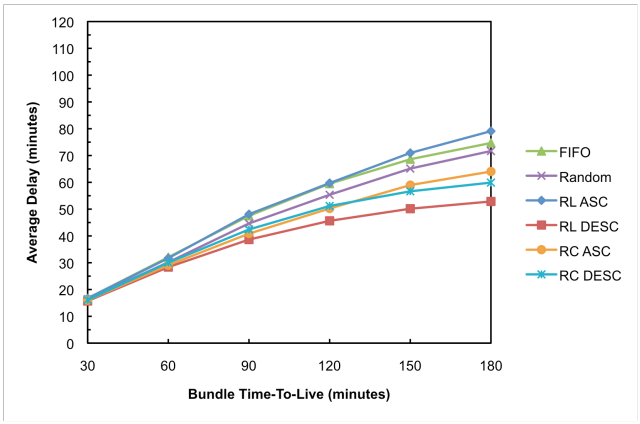


Figure 8. Bundle average delay as function of TTL in a scenario with 25 vehicles moving at a speed of 30 km/h.

Although the increase of vehicles average speed from 30 Km/h to 50 Km/h increases the number of contact opportunities (Figure 3), it decreases the contact duration (Figure 4). Hence, the number of bundles exchanged during a contact opportunity also decreases. As may be seen through the comparison between results shown in Figures 7 and 9, this resulted in lower delivery ratios for all combinations of scheduling and dropping policies, across all simulations. In this scenario, the difference in terms of performance between the policies decreases. The values of delivery ratio for RL DESC are close to the ones observed with RC ASC. Nevertheless, Figure 9 shows that RC ASC still performs better than the other policies, increasing about 3%, 4%, 8%, 8%, 7%, and 5% for each of the considered values of TTL, the bundle delivery probability, when compared to FIFO. Figure 10 shows that FIFO and RL ASC present the higher average delivery delay values. The difference, in terms of the average delay observed for RL DESC and RC ASC, increased for TTLs higher than 90 minutes, when compared to the results shown in Figure 7.

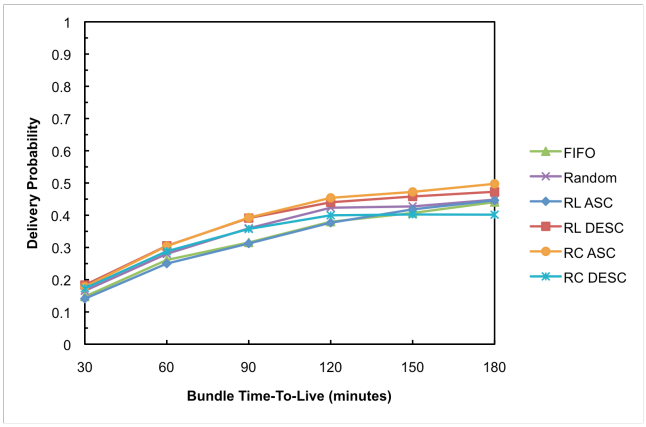


Figure 9. Bundle delivery probability as function of TTL in a scenario with 25 vehicles moving at a speed of 50 km/h.

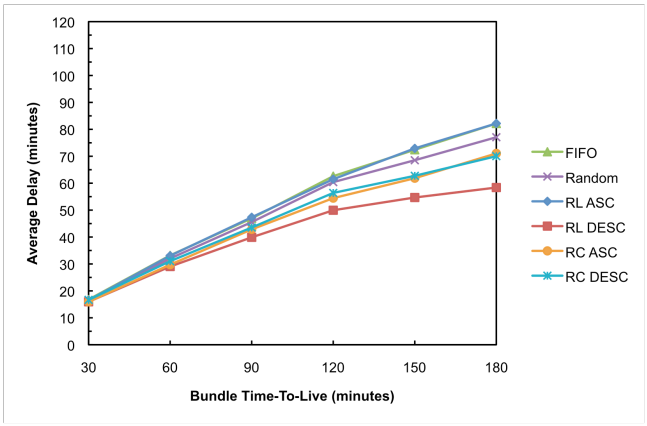


Figure 10. Bundle average delay as function of TTL in a scenario with 25 vehicles moving at a speed of 50 km/h.

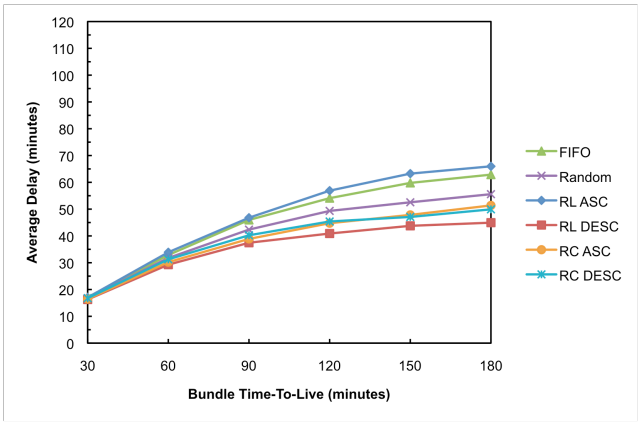


Figure 12. Bundle average delay as function of TTL in a scenario with 50 vehicles moving at a speed of 30 km/h.

C. *Performance Analysis for the Scenario with 50 Vehicles*

The second scenario considers 50 vehicles moving across the map (Figure 6). As shown in Figure 3, increasing node density the number of contact opportunities also increases. Then, it increases the number of relayed bundles and, potentially, causes more contention for network resources. Recall that the traffic generated is equal on both scenarios.

A comparison between results depicted in Figures 7 and 11 shows that policies have the same behavior and the delivery probability increases for all policies (Figure 11), when compared with the first scenario (Figure 7). Furthermore, this analysis also reveals that RC ASC registers the best results in terms of the delivery ratio, irrespective of the number of mobile nodes. In this second scenario, when vehicles move with an average speed of 30 Km/h, it presents gains of 4%, 9%, 6%, 6%, 4% and 5% for each of the considered values of TTL, respectively, when compared to the FIFO policy (Figure 11).

Figure 12 confirms the conclusions obtained in the first scenario. Although RC ASC policy requires slightly more time to deliver bundles than RL DESC, it achieves a higher delivery ratio (Figure 11).

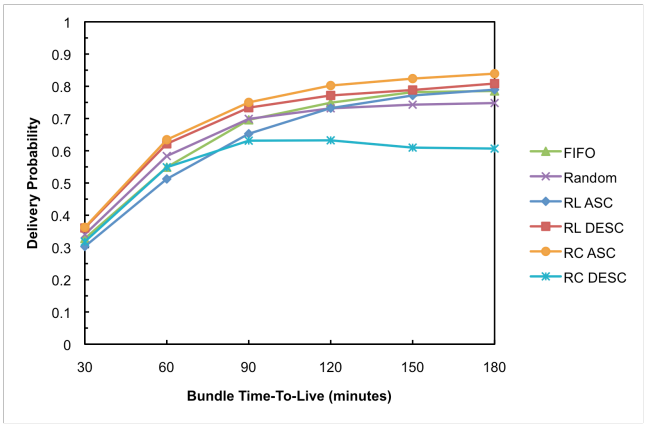


Figure 11. Bundle delivery probability as function of TTL in a scenario with 50 vehicles moving at a speed of 30 km/h.

Following the results shown in the previous scenario, increasing the vehicles average speed to 50 Km/h, decrease the number of successfully delivered bundles (Figures 11 and 13). However, it is interesting to observe that RC ASC policy is less affected by this change than the remaining policies. Due to this fact, RC ASC shows greatest improvements. Compared to FIFO, RC ASC increases the delivery probability in 6%, 11%, 11%, 10%, 6%, and 7% for each of the considered values of TTL, respectively.

As previously observed, the gains in the delivery ratio performance metric are attenuated when bundles have a large TTL. This is due to the fact that network nodes have large buffers and can carry and exchange these bundles during longer periods of time before expiring. However, increasing the TTL reinforces the improvement on average delay. When comparing with FIFO, bundles arrive at the destination nodes approximately 1, 4, 8, 12, 17, and 20 minutes sooner in average, if RC ASC policy is used.

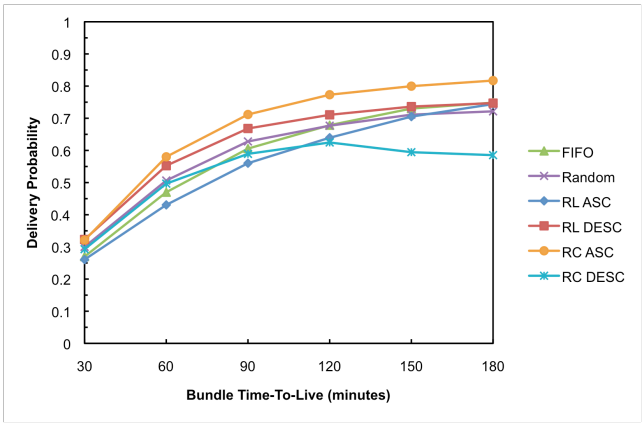


Figure 13. Bundle delivery probability as function of TTL in a scenario with 50 vehicles moving at a speed of 50 km/h.

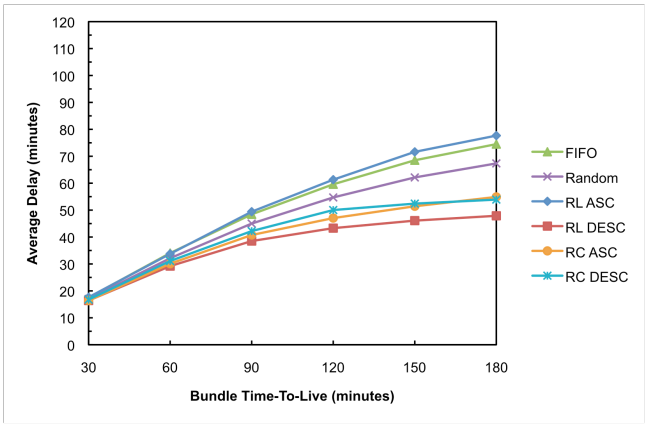


Figure 14. Bundle average delay as function of TTL in a scenario with 50 vehicles moving at a speed of 50 km/h.

VI. Conclusions and Future Work

This paper focused on the impact of scheduling and dropping policies on the performance of vehicular delay-tolerant networks. This work tried to find a good alternative to the traditional FIFO scheduling with "drop head" dropping policy, which would improve the VDTN network performance. In this context, several combinations of scheduling and dropping policies were proposed, and their relative performance was analyzed in terms of bundle delivery probability and average delivery delay. These policies were enforced on a Spray and Wait routing scheme.

This study considered two urban scenarios with different node densities and contact durations. The simulation results reveal a good performance obtained by a combination of a scheduling policy and a dropping policy that gives preferential treatment to less replicated bundles. It has been shown that such an approach outperforms the commonly used FIFO scheduling and "drop head" buffer management, in both performance metrics. This result was obtained and confirmed for all simulation scenarios.

For future work, we plan to investigate the use of scheduling and routing strategies based on geographical information for VDTNs.

Acknowledgments

Part of this work has been supported by *Instituto de Telecomunicações*, Next Generation Networks and Applications Group (NetGNA), Portugal, in the framework of the Project VDTN@Lab, and by the Euro-NF Network of Excellence of the Seventh Framework Programme of EU, in the framework of the Project VDTN.

References

[1] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, "Delay-Tolerant Networking Architecture," RFC 4838, April 2007, [Online]. Available: ftp://ftp.rfc-editor.org/in-notes/rfc4838.txt.

[2] K. Fall and S. Farrell, "DTN: An Architectural Retrospective," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, June 2008.

[3] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, "Delay-Tolerant Networking: An Approach to Interplanetary Internet," in *IEEE Communications Magazine*, vol. 41, 2003, pp. 128-136.

[4] J. Partan, J. Kurose, and B. N. Levine, "A Survey of Practical Issues in Underwater Networks," in *1st ACM International Workshop on Underwater Networks, in conjunction with ACM MobiCom 2006*, Los Angeles, California, USA, Sep. 25, 2006, pp. 17 - 24.

[5] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-Efficient Computing for Wildlife Tracking: Design Tradeoffs and Early Experiences with ZebraNet," *ACM SIGOPS Operating Systems Review*, vol. 36, no. 5, pp. 96-107, 2002.

[6] S. Jain, R. Shah, W. Brunette, G. Borriello, and S. Roy, "Exploiting Mobility for Energy Efficient Data Collection in Wireless Sensor Networks," *ACM/Kluwer Mobile Networks and Applications (MONET)*, vol. 11, no. 3, pp. 327-339, June 2006.

[7] N4C and eINCLUSION, "Networking for Communications Challenged Communities: Architecture, Test Beds and Innovative Alliances," [Online]. Available: http://www.n4c.eu/ [Accessed: January, 2009].

[8] A. Doria, M. Uden, and D. P. Pandey, "Providing Connectivity to the Saami Nomadic Community," in *2nd International Conference on Open Collaborative Design for Sustainable Innovation*, Bangalore, India, December, 2002.

[9] A. Pentland, R. Fletcher, and A. Hasson, "DakNet: Rethinking Connectivity in Developing Nations," in *IEEE Computer*, vol. 37, 2004, pp. 78-83.

[10] M. Asplund, S. Nadjm-Tehrani, and J. Sigholm, "Emerging Information Infrastructures: Cooperation in Disasters," in *Lecture Notes in Computer Science, Critical Information Infrastructure Security*, vol. 5508/2009: Springer Berlin / Heidelberg, 2009, pp. 258-270.

[11] N. Glance, D. Snowdon, and J.-L. Meunier, "Pollen: Using People as a Communication Medium," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 35, no. 4, pp. 429-442, March 2001.

[12] T. Olajide and A. N. Washington, "Epidemic Modeling of Military Networks using Group and Entity Mobility Models," in *5th International Conference on Information Technology : New Generations (ITNG 2008)*, Las Vegas, Nevada, USA, April 7-9, 2008, pp. 1303-1304.

[13] Y. Shao, C. Liu, and J. Wu, "Delay-Tolerant Networks in VANETs," in *Vehicular Networks: From Theory to Practice*, S. Olariu and M. C. Weigle, Eds.: Chapman & Hall, 2009.

[14] L. Franck and F. Gil-Castineira, "Using Delay Tolerant Networks for Car2Car Communications," in *IEEE International Symposium on Industrial Electronics 2007 (ISIE 2007)*, Vigo, Spain, 4-7 June, 2007, pp. 2573-2578.

[15] V. N. G. J. Soares, F. Farahmand, and J. J. P. C. Rodrigues, "A Layered Architecture for Vehicular Delay-Tolerant Networks," in *The Fourteenth IEEE Symposium on Computers and Communications (ISCC ' 09)*, Sousse, Tunisia, July 5 - 8, 2009, pp. 122-127.

[16] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant

Networks," in *INFOCOM 2006 - The 25th IEEE International Conference on Computer Communications*, Barcelona, Catalunya, Spain, April 23-29, 2006, pp. 1-11.

[17] V. Bychkovsky, K. Chen, M. Goraczko, H. Hu, B. Hull, A. Miu, E. Shih, Y. Zhang, H. Balakrishnan, and S. Madden, "The CarTel Mobile Sensor Computing System," in *The 4th ACM Conference on Embedded Networked Sensor Systems (ACM SenSys 2006)*, Boulder, Colorado, USA, October 31 - November 3, 2006, pp. 383-384.

[18] J. Morillo-Pozo, J. M. Barcelo-Ordinas, O. Trullos-Cruces, and J. Garcia-Vidal, "Applying Cooperation for Delay Tolerant Vehicular Networks," in *Fourth EuroFGI Workshop on Wireless and Mobility*, Barcelona, Spain, January 16-18, 2008.

[19] V. N. G. J. Soares, J. J. P. C. Rodrigues, P. S. Ferreira, and A. Nogueira, "Improvement of Messages Delivery Time on Vehicular Delay-Tolerant Networks," in *The 38th International Conference on Parallel Processing (ICPP-2009) Workshops - The Second International Workshop on Next Generation of Wireless and Mobile Networks (NGWMN-09)*, Vienna, Austria, September 22-25, 2009.

[20] K. Scott and S. Burleigh, "Bundle Protocol Specification," RFC 5050, November 2007, [Online]. Available: http://www.rfc-editor.org/rfc/rfc5050.txt.

[21] V. N. G. J. Soares, J. J. P. C. Rodrigues, F. Farahmand, and M. Denko, "Exploiting Node Localization for Performance Improvement of Vehicular Delay-Tolerant Networks," in *2010 IEEE International Conference on Communications (IEEE ICC 2010) - General Symposium on Selected Areas in Communications (ICC'10 SAS)*, Cape Town, South Africa, May 23-27, 2010.

[22] N. Banerjee, M. D. Corner, and B. N. Levine, "An Energy-Efficient Architecture for DTN Throwboxes," in *26th IEEE International Conference on Computer Communications (INFOCOM 2007)*, Anchorage, Alaska, USA, May 6-12, 2007, pp. 776-784.

[23] J. J. P. C. Rodrigues, V. N. G. J. Soares, and F. Farahmand, "Stationary Relay Nodes Deployment on Vehicular Opportunistic Networks," in *Mobile Opportunistic Networks: Architectures, Protocols and Applications*, M. K. Denko, Ed. USA: Auerbach Publications, CRC Press, May 2010.

[24] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-copy Case," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 1, pp. 77-90, February 2008.

[25] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem," in *ACM SIGCOMM 2007*, Kyoto, Japan, August 27-31, 2007, pp. 373-384.

[26] Z. Zhang, "Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 8, no. 1, pp. 24-37, 2006.

[27] V. N. G. J. Soares, F. Farahmand, and J. J. P. C. Rodrigues, "Evaluating the Impact of Storage Capacity Constraints on Vehicular Delay-Tolerant Networks," in *The Second International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2009)*, Colmar, France, July 20-25, 2009, pp. 75-80.

[28] A. Lindgren and K. S. Phanse, "Evaluation of Queueing Policies and Forwarding Strategies for Routing in Intermittently Connected Networks," in *First International Conference on Communication System Software and Middleware (COMSWARE 2006)*, Delhi, India, January 8-12, 2006, pp. 1-10.

[29] A. Vahdat and D. Becker, "Epidemic Routing for Partially-Connected Ad Hoc Networks," Duke University, Technical Report CS-2000-06, April, 2000.

[30] A. Lindgren, A. Doria, E. Davies, and S. Grasic, "Probabilistic Routing Protocol for Intermittently Connected Networks," draft-irtf-dtnrg-prophet-02, March 9, 2009, [Online]. Available: http://tools.ietf.org/html/draft-irtf-dtnrg-prophet-02.

[31] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance Modeling of Epidemic Routing," *Computer Networks, Elsevier*, vol. 51, no. 10, pp. 2867-2891, July 2007.

[32] A. Krifa, C. Barakat, and T. Spyropoulos, "Optimal Buffer Management Policies for Delay Tolerant Networks," in *Fifth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks - SECON 2008*, San Francisco, California, USA June 16-20, 2008, pp. 260-268.

[33] V. Erramilli and M. Crovella, "Forwarding in Opportunistic Networks with Resource Constraints," in *ACM MobiCom Workshop on Challenged Networks (CHANTS 2008)*, San Francisco, California, USA, September 15, 2008, pp. 41-48.

[34] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation Forwarding," in *9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2008)*, Hong Kong, China, May 27-30, 2008, pp. 251-260.

[35] Y. Li, L. Zhao, Z. Liu, and Q. Liu, "N-Drop: Congestion Control Strategy under Epidemic Routing in DTN," in *5th International Wireless Communications and Mobile Computing Conference (IWCMC 2009)*, Leipzig, Germany, June 21-24, 2009, pp. 457-460.

[36] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation," in *Second International Conference on Simulation Tools and Techniques (SIMUTools 2009)*, Rome, March 2-6, 2009.

[37] A. Keränen and J. Ott, "Increasing Reality for DTN Protocol Simulations," Helsinki University of Technology, Networking Laboratory, Technical Report, July, 2007.

[38] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," in *ACM SIGCOMM 2005 - Workshop on Delay Tolerant Networking and Related Networks (WDTN-05)*, Philadelphia, PA, USA, August 22-26, 2005, pp. 252-259.

Cost-Optimal and Cost-Aware Tree-Based Explicit Multicast Routing

Miklós Molnár
University of Montpellier 2,IUT, Dep. of Computer Science, LIRMM
161 rue Ada, 34095 Montpellier Cedex 5 France
Email: Miklos.Molnar@lirmm.fr

***Abstract*— This paper aims to introduce the hard optimization problem of determining tree-based explicit multicast routes with minimum cost. Explicit multicast routing has been proposed as a technique to solve the problem of multicast scalability in IP-based networks. Tree-based explicit routing is a special routing technique, in which the multicast tree is computed at the source and encoded explicitly in the datagram headers. These enlarged headers may result in significant overhead traffic, so the cost minimization of this kind of routing is a relevant topic. In this particular multicast routing, the well known minimum cost spanning trees (Steiner trees) do not corresponds to the optimal solution: the overhead induced by the large header corresponding to a Steiner tree can be excessive. This paper proposes the optimization of the routing minimizing the communication cost per bit in tree-based explicit multicasting. If the multicast group is large and the header size is limited, several trees are needed to provide routing for the entire group. In this case, the optimization can be seen as a particular constrained partial spanning problem. It is demonstrated that the computation of the minimum cost tree and the set of trees with minimum cost are NP-difficult problems. The presented theoretical analysis is indispensable to find cost efficient routes for these kinds of multicast routing protocol. Some algorithmic issues of the tree set construction are also discussed in the paper: exact and heuristic algorithms are presented. In real routing protocols, expensive exact algorithms cannot be applied. So, the paper also aims with the presentation of some tree-based explicit multicast routing algorithms using polynomial execution time.**

***Keywords*-Communication theory; multicast routing; combinatorial optimization; minimum cost routing; Steiner problem; hierarchy; QoS-based routing;**

I. INTRODUCTION

Multicasting was proposed to minimize bandwidth and network resource usage (for instance in IP based networks) by Deering in [2]. This kind of communication allows messages to be sent to a set of destinations in a special way: at most one copy of each message is forwarded on each link of a multicast tree. There is a large variety of distributed applications including television, video on demand, games and video-conferences, which benefit from multicast communication. In IP based networks the deployment of multicasting has been delayed by the well known problem of scalability. Because IP multicast addresses do not contain any specific information (for example: localization of the destination), address based aggregation of multicast communications is not possible and thus multicasting does not scale with the number of multicast groups. Indeed, IP routers store an entry for each multicast group using the given router. The large number of multicast entries in the forwarding tables retard the forwarding process. Another problem for the deployment of multicasting is that currently not all routers in the Internet are multicast capable. To introduce multicast communication progressively, it is important to design protocols, which allow multicast via unicast forwarding in certain domains. For this reason, protocols such as REUNITE (cf. [3]) and HBH (cf. [4]) have been proposed. In these protocols forwarding is done in the traditional unicast way and the branching node routers store information on next destinations in special tables. Trivially, this kind of protocol does not resolve the scalability problem.

Explicit multicast routing protocols have been proposed that scale better with the number of multicast groups. When explicit routing is used the group forwarding information is stored in the header of the datagrams. The group information is generally collected by a particular router and this information should be available at the source to send the datagrams. So, this type of multicasting can be regarded as a source-based routing technique. Simple flat explicit multicast routing only encodes the set of destinations in the datagram headers. In the subsequently encountered routers, datagrams are forwarded using the header information by applying the locally available forwarding mechanism (often a unicast forwarding). Accordingly, there is no forwarding state information for the given groups in the forwarding tables.

The flat explicit routing protocols suffer from an important drawback: each intermediate router on the multicast route has to inspect the datagram header. The router should duplicate the datagram if there are several next hops to forward it toward the encoded destinations. This handling is obligatory even when the router is not a branching node of the multicast tree.

To avoid obligatory processing of the datagram headers in the intermediate routers, tree-based explicit multicast routing

protocols have been proposed [5] [6]. In these protocols, the source (or an appropriate route computation element) computes the tree spanning the destinations and stores the tree structure in the datagram headers. Note that the tree can be encoded entirely by its significant nodes (destinations and branching nodes), and the data forwarding between two successive significant nodes can be performed using unicast routing. This allows tree-based explicit protocols to forward datagrams faster than flat explicit protocols.

An indisputable drawback of explicit multicast routing resides in the traffic overhead due to the enlarged header size. Moreover the header size may differ between one route and another, and this is particularly true for encoded trees. The more significant intermediate (and so encoded) nodes a spanning tree contains, the longer the datagram header becomes. The generated header related traffic must be taken into account, even for route computation and optimization. So, the optimization of the communication cost needs a new formulation of the explicit multicast routing problem, which is significantly different from the classic Steiner problem [7].

When IP protocols are used, explicit multicast routing must cope with datagram fragmentation. Because the amount of encoded routing information in the headers can be significant it is possible that these datagrams will be fragmented and data and header will be unfortunately separated. To avoid bad IP fragmentation, the segmentation of the destination set into several sub-sets has been proposed for flat explicit routing protocols [8]. Using this technique each sub-set of destinations can be encoded separately by a ”small” datagram, which may be sent without fragmentation. However, the segmentation of multicast delivery trees for tree-based explicit protocols has not yet been investigated.

Because multicast datagrams can be fragmented and the multicast structure segmented, we analyze the optimality of routes with and without fragmentation. More precisely, we describe the optimal multicast structure, which generates the minimum communication cost per bit including the variable cost of the header transmission. Generally, by taking into account the header size limitation, this cost minimization corresponds to a constrained partial minimum spanning problem, which is NP-difficult even if the solution is a single tree.

Tree-based explicit multicast routing protocols can be solicited for different reasons and not only to tackle the scalability problem. Multicast communication may be constrained by a given policy of the source or of the application. The quality of service (QoS) requirement is one of the most frequently imposed constraints. Often the QoS is formulated on the basis of multiple criterion and the computation of feasible or optimal routes corresponds to a multi-constrained optimization. Finding the multicast graph respecting the defined QoS requirements and minimizing network resources is an NP-complete optimization task [9]. For example, Multicast Adaptive Multiple Constraints routing Algorithm (MAMCRA) [10] proposes the computation of routing structures constrained by multiple QoS criterion from the source to the destinations. In certain cases the result does not correspond to a tree but to a set of trees and paths rooted at the source and containing some cycles. Traditional IP multicast routing using a single IP address for the group cannot be used. Explicit routing is a good candidate to resolve the conflicts induced by the cycles. More generally, constrained multicast routing structures are tree-like structures called *hierarchies*. The use of this kind of structure for multicast routing in IP domain necessitates routing protocols that allow the crossing of branches (routes) in the same multicast route structure. The technique of tree-based explicit multicast routing also permits the encoding of hierarchical routing structures.

Another candidate for tree-based explicit multicasting is application level multicasting. Delivery trees can be computed at the application level and overlay links can be used among end systems handling the multicast packets. These solutions support naturally traffic engineering, can improve the reliability of multicast delivery, and facilitate secure group communications [11]. Generally, in traffic engineering solutions and QoS aware environments tree-based explicit multicasting may offer an interesting tunable multicast data delivery technique.

The present work focuses on the cost-optimal tree-based explicit multicast solutions taking into account the increased bandwidth usage due to the largest datagram headers. It is an extended version of [1], and it provides more detailed information about the problem formulation, the properties of the optimal solution and some algorithmic issues of the possible route computations.

The next section gives a rapid overview of the related work. The formulation of the tree-based multicast routing problem with minimum communication cost can be found in Section III. We demonstrate that the cost optimization of tree-based explicit routing is an NP-difficult computational problem. Some exact algorithms are presented in Section IV but these algorithm are very expensive. More practicable heuristic algorithms are also proposed for routing protocols. Our conclusions and perspectives close this initiative study.

II. Related Work

Initially, explicit multicast routing was proposed for small multicast groups (cf. Small Group Multicast in [12]) to decrease the number of multicast entries in routing tables.

Combined with the traditional IP multicast routing for large groups, scalability for all type of multicast can be achieved.

A simple approach to implementing explicit multicast routing is to simply store the set of destination addresses in each datagram. The basic protocol of this kind of flat explicit protocol is the Xcast protocol proposed in [13]. When an Xcast router receives an Xcast datagram, it performs a look-up for each valid destination in the header to determine the required next hop. Then it copies the incoming datagram to each required outgoing link. An improved version of this protocol is the protocol Xcast+, described in [14], which uses dedicated routers to reduce the header size. Simple explicit multicast routing eliminates tree construction and maintenance costs in the network and decreases the network control load. For these reasons it was also proposed for mobile ad hoc networks [15].

To resolve the main drawback of flat explicit routing protocols (which is the check of the destination list in each router) precomputed tree based explicit routing was proposed. The first tree-based explicit protocol was the ERM protocol proposed in [5]. In the ERM protocol the source encodes the IP addresses of the branching nodes and the destinations of the multicast tree in the datagram headers. Inside the routing domain, this header is analyzed and datagrams are routed using unicast forwarding mechanism. The protocol Linkcast, described in [6] improves ERM by proposing a new header encoding. Since the tree is encoded in the datagram header, a node can easily decide whether it is a branching node or not. Similarly, it is easy for a branching node to find its children. In [16] the trade-offs of the tree-based explicit routing protocol design are discussed and a performance analysis is presented. The analyzed metrics are the header size and the processing overheads. More detailed and appropriate tree information may reduce the processing overhead in return for larger header size and traffic overhead. The authors propose a modification to ERM called Bcast, which reduces the overhead of the protocol. In Bcast, a proactive bypassing mechanism helps to adjust the code size in response to inconvenient distribution of the receivers.

Using IP protocols, explicit multicast routing will inevitably experience datagram fragmentation. Because the amount of encoded routing information in the headers can be significant, it is possible that these datagrams will be fragmented and data and header will be unfortunately separated. The problem of IP fragmentation of multicast datagrams using flat explicit routing has been analyzed in [8]. The segmentation of the destination set into several sub-sets has been proposed to avoid cutting the headers in two. The optimal segmentation has also been analyzed and the authors have demonstrated that quasi-optimal communication cost can be obtained when header length is less than half the datagram size.

III. Communication Cost Optimization for Tree-Based Explicit Multicasting

In this section, we formulate the optimal tree based explicit multicast routing, which minimizes the communication cost (and not the cost of the used trees). We will show that communication cost minimization is a very hard problem when the traffic overhead due to explicit routing headers and segmentation must be taken into account. This problem corresponds to a special constrained Steiner problem with nonlinear cost function even if the maximal header size does not limit the tree. In the general case, when the limitation on the maximal header size should be taking into account, the problem becomes a special constrained partial spanning problem. In this case, the optimum corresponds to a special hierarchy: to a set of spanning trees.

Let $G = (V, E)$ be the undirected and connected graph corresponding to the network topology and $D \subset V$ the set of destinations of the multicast group originated at the source s. Let us suppose that the network topology is known at the source. Moreover, the size of the datagrams is limited by a value L_{max}.

We suppose that a homogeneous unicast routing mechanism exists in the routing domain and that this mechanism is known at the source. So, the source node can compare any spanning tree with the possible unicast routes in order to decide, which nodes of the tree should be encoded explicitly. Explicit multicast routing can then use the unicast routing mechanism between any two successive encoded nodes of the multicast tree. Evidently, the encoded tree must contain all nodes such that the set of unicast routes between them corresponds to the original tree. In the following, we call these nodes of the tree *significant nodes*. Figure 1 illustrates the encoding of the tree in a simple example. The source node s would send messages for the destination set $D = \{d_1, d_2, d_3, d_4, d_5\}$ using tree-based explicit multicast routing. Let us suppose that the unicast routing uses the shortest paths between any node pairs and the multicast tree is a partial spanning tree T as indicated in the figure. In this case the significant nodes are the nodes, which are indicated with a double line and correspond to the following parenthesized list:

$$a(c(d_1, d_2, g(d_3, d_4, d_5)))$$

The node a is significant, because the shortest path from s to c does not pass through a. So, to follow the route from s to c via a, a must be explicitly encoded.

To simplify, in the following the set of significant nodes S_T of a spanning tree T denotes the union of the branching

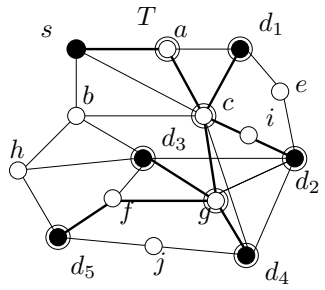


Figure 1. Significant nodes of a tree for unicast routing

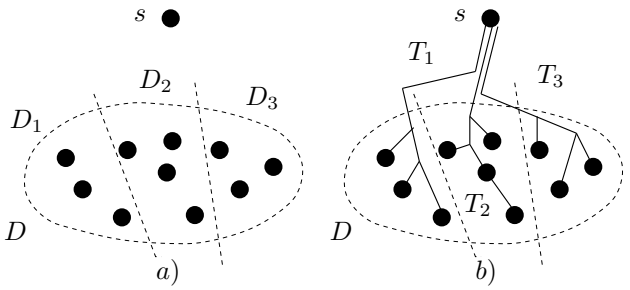


Figure 2. The optimization is a superposition of a) a partitioning and b) a minimum cost weighted spanning tree problem

nodes B_T of T and the destinations: $S_T = B_T \bigcup D$ (*i.e.*, the paths between these nodes are shortest paths).

We consider the minimization of the *total communication cost* as the objective of the tree-based explicit multicast optimization under the constraint on the maximal length of datagrams. We will show that this cost does not correspond to the sum of link costs as it is the case in simple multicast route computations. We distinguish two components of the communication cost: the cost of the transmitted payload and the cost of the overhead generated by the headers. The latter cost is proportional to the explicit routing header size. This header size depends on the number of encoded addresses and so on the structure proposed for the routing. We will show that the minimal cost routing structure is always a set of trees. Figure 2 illustrates the difficulty of the optimization. Generally, the optimal solution comprises several destination sub-sets (which should be spanned separately because of the constraints). Thus, the first question is related to the partitioning of the destination set (Figure 2/a). Then, for each sub-set of destinations a special minimum cost partial spanning tree should be built. This latter problem itself is NP-difficult (and can be seen as a special case of the Steiner problem, cf. in the next sub-section). The optimal routing problem is the superposition of these difficult optimization problems, since the cost of the partitioning and the spanning trees are inseparably related. If the trees resulting from the segmentation are large (in term of number of encoded nodes), then the payload in the datagrams is small and several datagrams should be sent to transmit the desired message. If the trees are small, then several trees are needed to cover the entire multicast group.

In this section, we first present the objective function of the minimal cost tree construction even if segmentation is not needed (one tree can cover the entire multicast group and can be encoded in the header without segmentation and it corresponds to the minimal cost solution). Then, we show that this optimization problem is NP-difficult. Secondly, we present the explicit multicast routing structure optimization in the case where header segmentation is required. We will show that the optimal routing structure corresponds to a set of trees and the problem remains NP-difficult.

A. Minimum cost tree considering the header length

In order to determine the objective function of the explicit routing optimization progressively, we first focus on the simple case where only one encoded spanning tree is needed to cover the destinations.

More precisely, this case is produced when

- the source has only one sub-tree for spanning all of the destinations
- there is sufficient space in the packet header to store the encoded version of this unique spanning tree.

So, first we consider a unique spanning tree that covers the entire set of destinations where the tree is encoded and stored entirely in the corresponding datagram headers. If a spanning tree has several sub-trees at the source, then the datagrams sent on each sub-tree have distinct multicast tree encoding. Such a spanning tree can be considered as a set of its sub-trees rooted at the source. For instance, if a tree T can be decomposed at its source into two (disjoint) sub-trees T_a and T_b, then we say that, from the point of view of tree encoding, a set of trees $\{T_a, T_b\}$ covers the destinations. The trees of this set are encoded separately in different packet headers. Figure 3/a and Figure 3/b illustrate respectively the cases when the source has only one sub-tree, and where two disjoint sub-trees cover the destination set.

Lemma 1. *If the segmentation of the destination set is not needed and all of the destination are accessible via the same neighbor node of the source, the optimal structure is a partial spanning tree.*

Proof: Without segmentation, the optimal solution is a connected sub-graph. The datagrams are sent on each link in this structure. Since the edges are positively evaluated, an eventual cycle increases the cost. So, the solution is a partial spanning tree. ■

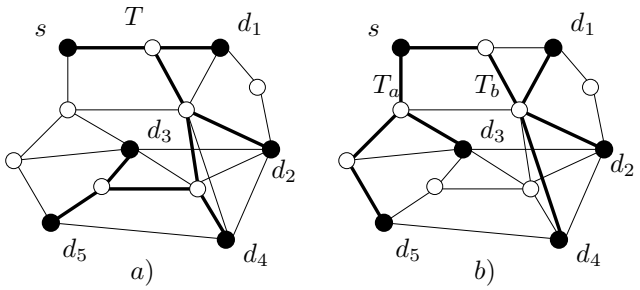


Figure 3. One or several sub-trees may be at the source

To cover the given set of destinations and the source with a single tree, different partial spanning trees can be found and enumerated (for example by a Steiner Tree Enumeration Algorithm, cf. [17]). Each tree contains a different set of significant nodes, corresponds to a specific header length and so involves a specific overhead and payload. The optimal solution is the tree, which minimizes the total communication cost. In the following, we talk about the partial minimum spanning tree for encoding (which is generally different from the Steiner tree of the given group).

To formulate the overhead generated by the explicit multicast headers, let us suppose that the significant nodes are encoded by their network addresses using l_a bytes, there are $k(T)$ significant nodes and the maximal size of messages is equal to L_{max} bytes. The encoding technique of the datagram header is out of scope of the paper. Only the impact of the encoded tree is analyzed, the rest of the header is considered to have a constant length. In this way, the size l_h of a header can be expressed by $l_h = k(T) \cdot l_a + c$, where c is the constant length of the rest of the header. Using datagrams with the maximum length, the maximum payload in a datagram corresponds to $l_p = L_{max} - k(T) \cdot l_a - c$ and to transmit a message of L bytes, $n_p = \left\lceil \frac{L}{L_{max} - k(T) \cdot l_a - c} \right\rceil$ datagrams must be used. So, the traffic generated by the transmission of the headers can be expressed by $L_h = n_p \cdot (k(T) \cdot l_a + c)$. The traffic corresponding to the transmission of the message of length L is

$$L_k = L + \left\lceil \frac{L}{L_{max} - k(T) \cdot l_a - c} \right\rceil \big(k(T) \cdot l_a + c\big) \quad (1)$$

Let us suppose that the communication uses a tree T of cost $d(T)$. Thus the total communication cost is

$$C_L(T) = L_k \cdot d(T) = \left(L + \left\lceil \frac{L}{L_{max} - k(T) \cdot l_a - c} \right\rceil \big(k(T) \cdot l_a + c\big) \right) \cdot d(T) \quad (2)$$

The optimization of the communication support should be independent from the message length L. The cost per bit better characterizes the cost of the communication and this cost should be minimized. The cost per bit can be obtained asymptotically as

$$C(T) = \lim_{L \to +\infty} \frac{C_L(T)}{L} = \lim_{L \to +\infty} \left(1 + \frac{\left\lceil \frac{L}{L_{max} - k(T) \cdot l_a - c} \right\rceil \big(k(T) \cdot l_a + c\big)}{L} \right) \cdot d(T) \quad (3)$$

Finally, the communication cost per bit using the tree T corresponds to

$$C(T) = (1 + \frac{k(T) \cdot l_a + c}{L_{max} - k(T) \cdot l_a - c}) d(T) = \frac{L_{max}}{L_{max} - k(T) \cdot l_a - c} d(T) \quad (4)$$

The optimal encoded partial spanning tree T^*_M is the tree, which minimizes this communication cost (Problem 1):

$$T^*_M : \arg \min_{T \in \mathcal{ST}} \frac{L_{max}}{L_{max} - k(T) \cdot l_a - c} d(T) \quad (5)$$

Theorem 1. *The optimization given by (5) is NP-difficult.*

Proof: Trivially, if a particular case of the problem given by (5) is NP-difficult, then the problem is NP-difficult. In the expression (5) the length $d(T)$ of the tree T is multiplied by a factor

$$f(T) = \frac{L_{max}}{L_{max} - k(T) \cdot l_a - c}, \quad (6)$$

that characterizes the tree (it depends on the number of significant nodes in the tree). Generally, this factor is different from one tree to another. Let l_a be chosen so that the factors $f(T)$ do not influence the choice of the optimal solution compared with the tree lengths. Concretely, for every pair (T_i, T_j) of possible spanning trees, such that $d(T_i) < d(T_j)$, let a value l^m_a be chosen, which guarantees that $C(T_i) < C(T_j)$. Taking into account the cost function, the condition for this can be expressed as

$$\frac{d(T_i)}{d(T_j)} < \frac{L_{max} - k(T_j) \cdot l^m_a - c}{L_{max} - k(T_i) \cdot l^m_a - c} \quad (7)$$

Since $\frac{d(T_i)}{d(T_j)} < 1$ such a value exists. With this value of l^m_a, the corresponding factors $f(T)$ do not influence the relation between the spanning trees: if $d(T_i) < d(T_j)$ then $f(T_i) \cdot d(T_i) < f(T_j) \cdot d(T_j)$. In this case, the shortest partial spanning tree from the set of all partial spanning trees is the solution of our problem. The selection of the minimum cost

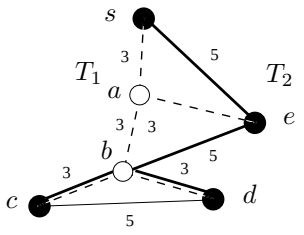


Figure 4. The impact of the header size on the optimal cost

partial spanning tree corresponds to the NP-difficult Steiner problem. ■

The simple example of Figure 4 illustrates the impact of the header size on the optimal cost solution. Let us suppose that the maximal datagram length L_{max} is equal to 20 bytes, that there is no additional constant information in the header ($c = 0$) and the addresses are encoded on $l_a = 2$ bytes. The node s is the source of the communication and c, d, e are the destinations. The minimal cost Steiner tree T_1 covering the source and the destinations is marked with dotted line on the figure and has a length $d(T_1) = 15$. As there are two branching nodes on the tree, the number of significant nodes is $k(T_1) = 5$. So the factor corresponding to this tree is $f(T_1) = 2$. The total communication cost implicated by the tree $f(T_1) \cdot d(T_1)$ is 30. When taking the header size into account, we obtain the tree T_2 represented by a bold line on the figure. This tree is longer ($d(T_2) = 16$) but there are less branching nodes, $k(T_2) = 4$ and the factor $f(T_2) = 1,67$. The total communication cost of this tree $f(T_2) \cdot d(T_2) = 26,67$ is less then the cost per bit using the encoded Steiner tree.

B. Minimum cost solution with header segmentation

When the multicast group is large and the number of significant nodes in the multicast tree is high, a single encoded tree cannot ensure the coverage of the destination set. The group should be segmented in the optimal solution. Let us notice that in some cases some segmentation may be naturally given by the sub-trees at the source (cf. Figure 3/b)). These sub-trees are edge disjoint. In other cases the solution may contain non-disjoint trees. An example can be found in Figure 5, where the number of encoded significant nodes is supposed to be limited to 4. In the given graph, the five destinations cannot be spanned by a unique spanning tree from the source s. Segmentation is necessary. The figure illustrates a segmentation where two non-disjoint trees span the destination set. The nodes c and d_3 belong to both trees. T_1 should be encoded as $T_1 = (d_3(f(d_5, d_4)))$ and T_2 corresponds to $T_2 = (d_3(c(d_1, d_2)))$ for routing. Note that the node d_3, which belongs to both trees is a destination node. That does not mean that d_3 should consume any message twice. On the contrary, this node must receive the message for local consumption only once and the second message must be transmitted to the next node without local consumption. In other worlds: the node d_3 is a destination in only one tree and serves as relay node in the other. So, an *exclusively served destination node set* is associated with each spanning tree of a segmented solution. This exclusively served destination node set contains the real destinations in the tree (and not the relays even if they are destinations in the original problem).

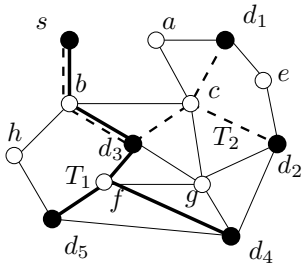


Figure 5. Two spanning trees with intersection

Lemma 2. *1) If the segmentation of the destination set is needed, the optimal explicit multicast routing structure is a set of partial spanning trees. 2) Each tree of this optimal set is rooted at the source and corresponds to a partial spanning tree minimizing the total communication cost for its exclusively served destinations.*

Proof: 1) The optimal solution Θ connects the destinations to the source. Since the maximum length of the datagrams limits the number of the significant nodes encoded in the headers, a single datagram header cannot be used for all destinations. A partitioning of the destination set is required. To connect a sub-set of the destinations to the source with a unique sub-graph in the optimal solution only a spanning tree can be used (cf. Lemma 1). So, the optimal solution Θ is a set of trees. 2) Each spanning tree in Θ should be a partial spanning tree minimizing the total communication cost relative to its exclusively served destinations. Let us suppose that a tree $T' \in \Theta$ does not minimize the communication cost relative to its exclusively served destination set $D_{T'}$. In this case, there is an other tree T'' minimizing the communication cost for the same destinations. So the overall solution containing T' cannot reach minimum cost. ■

After segmentation, each header contains a tree spanning a sub-set of the multicast group. Let us suppose that the segmentation results a set of trees $F = \{T_i, i = 1, ..k(F)\}$ spanning $\{s\} \cup D$ with not necessary disjoint trees T_i.

Here $k(F)$ indicates the number of trees in the segmented solution.

Trivially, to reach the destinations, each tree of the solution should be rooted at the source node. So, the minimum cost multicast route forms a set of trees routed at the source. This kind of set of trees is often called a "forest" in the literature. Recently, a new spanning structure was introduced in [18] to describe hierarchical structures which are, contrarily to trees, not obligatory exempt of redundancies. A *hierarchy* in a graph is a connected structure of consecutive nodes and edges that allows the some nodes and edges to be repeated such that for each node occurrence there is at most one predecessor node occurrence in the structure. In other worlds, a hierarchy is a tree-like structure permitting the repetition of the graph elements. A non-elementary path may contain a node several times. An elementary path is a path without repetition of the graph elements. The hierarchy is a more general concept than the tree concept and it may contain a node several times. Hierarchies without repetition are trees. Evidently, a set of trees routed at the same source node is a hierarchy since nodes and edges may be repeated in the set of the trees but each node occurrence has only one predecessor in the set. The source node can have several sub-hierarchies which are, in this particular case, spanning trees.

Generally, the different trees (sub-hierarchies) do not contain the same number of significant nodes. On the tree T_i, which has $k(T_i)$ significant nodes, the maximal payload per datagram is $p(T_i) = L_{max} - k(T_i) \cdot l_a - c$. It was shown in the last section that a tree T_i is optimal for the sub-set of destinations, if the total cost

$$C(T_i) = (1 + \frac{k(T_i) \cdot l_a}{L_{max} - k(T_i) \cdot l_a - c}) d(T_i) \qquad (8)$$

is minimal. Using the previously mentioned set of trees (or hierarchy) F, the total transmission cost of a message of L bytes corresponds to

$$C_L(F) = \sum_{i=1}^{k(F)} \left(L + \left\lceil \frac{L}{L_{max} - k(T_i) \cdot l_a - c} \right\rceil \cdot \big(k(T_i) \cdot l_a + c\big) \right) d(T_i) \qquad (9)$$

The optimal solution (which results in the minimum cost per bit when L tends to infinite) is a hierarchy (set of trees) F_M^* spanning $s \cup D$ (Problem 2) such as:

$$F_M^* : \arg \min_{F \in \mathcal{SF}} \sum_{i=1}^{k(F)} \frac{L_{max}}{L_{max} - k(T_i) \cdot l_a - c} d(T_i) \qquad (10)$$

where $\mathcal{SF}$ denotes the set of hierarchies spanning $s \cup D$, each hierarchy is composed of partial spanning trees. The complexity of this new problem is discussed later, at the end of the next sub-section. Here we propose first a simplification of the data fragmentation.

In the optimal solution presented above it is possible that the header length and the payload are different from one tree to another. The differing fragmentation of the same message depending on the different trees may significantly complicate the data transmission procedure at the source. Organizing multicast communication around a set of trees that use the same data transmission procedure facilitates the explicit routing protocol.

C. Minimum cost solution with homogeneous fragmentation

Generally, the different trees in the segmented solution do not contain the same number of significant nodes. On the tree T_i, which has $k(T_i)$ significant nodes, the maximum payload per datagram is $p(T_i) = L_{max} - k(T_i) \cdot l_a - c$. The fragmentation of the message of L bytes is optimal in T_i, if this maximum payload is applied in the tree. To obtain the maximum payload a customized fragmentation is needed on each tree. In each tree, the data should be sent using different fragments, which results in a very complicated transmission procedure at the source.

Homogeneous fragmentation constraint. To simplify the fragmentation task at the source, let us suppose that the source implements a common fragmentation algorithm and always sends the same content (payload or fragment) on the trees covering the multicast group.

To satisfy the Homogeneous fragmentation constraint the maximum number of significant nodes per tree is trivially:

$$k_{max}(F) = \max_{T_i \in F} k(T_i) \qquad (11)$$

Consequently in a simple data transmission procedure, each header contains at most $k_{max}(F)$ encoded significant nodes and the payload is the same in simultaneously sent datagrams. To transmit a message of length L, the source should use

$$n_p = k(F) \left\lceil \frac{L}{L_{max} - k_{max}(F) \cdot l_a - c} \right\rceil \qquad (12)$$

datagrams. Using the aforementioned hierarchy corresponding to a set of trees F the total cost of the communication is equal to

$$C_L(F) = \sum_{i=1}^{k(F)} \left(L + \left\lceil \frac{L}{L_{max} - k_{max}(F) \cdot l_a - c} \right\rceil \cdot \big(k_{max}(F) \cdot l_a + c\big) \right) d(T_i) \qquad (13)$$

The optimal hierarchy (which induces the minimum cost per bit) corresponds to the set of trees spanning $s \cup D$ (Problem 3) such that:

$$F_M^* : \arg \min_{F \in \mathcal{SF}} \sum_{i=1}^{k(F)} \frac{L_{max}}{L_{max} - k_{max}(F) \cdot l_a - c} d(T_i) \quad (14)$$

where $\mathcal{SF}$ denotes the set of hierarchies spanning $s \cup D$ composed only from trees under the mentioned constraint.

The difference between the optimization problems 10 and 14 resides in the factor $f(T)$, which weights the different trees in the sums. These weights are typical of each tree in Problem 10 but they have the same value within a partition in Problem 14. So, optimization 14 with homogeneous weights is more simple but the complexity of both problems is high.

Theorem 2. *The optimization (10) and (14) are NP-difficult.*

Proof: In both problems, the optimal solution corresponds to a set F of trees. The destination sub-sets spanned exclusively by the different trees in F correspond to a partition $P = \{D_i, i = 1, \ldots, k(F)\}$ of D. Each sub-set of destinations D_i in this partition is covered by a partial spanning tree $T_i \in F$. Trivially, the tree T_i is of the minimum cost per bit regarding D_i, and the result corresponds to the partition minimizing the total cost (the sum of the sub-set costs). So, the solution corresponds to the selection of the minimum cost partition and this problem corresponds to the well known set cover problem, which is NP-difficult [19]. Trivially, each partial spanning tree T_i in the solution should be a partial spanning tree of $\{s\} \cup D_i$ inducing the minimum cost per bit while respecting the constraints (otherwise there is a solution with less cost when using the minimum cost spanning tree instead of T_i). For example, to find the optimal cost partial spanning tree of $\{s\} \cup D_i$ in a given partition, the Homogeneous fragmentation constraint should be respected. A tree with minimum cost per bit must be computed while respecting the maximum homogeneous header size and thus while respecting the maximum number of significant nodes. This latter computation itself is a NP-difficult problem (it can be considered as a particular case of Theorem 1). Combined with the optimal partitioning Problems (10) and (14) are NP-difficult. ■

IV. Algorithms to Find Cost-Aware Explicit Multicast Routes

Without completeness, some basic ideas to find minimum cost and cost aware solutions for the tree-based explicit multicast routing can be found in this section. Since the problem is NP-difficult, exact algorithms are expensive. Cost-aware but non-optimal solutions can be obtained by heuristics taking reasonable (polynomial) execution time.

A. Exact and heuristic solutions of Problem (5)

In Problem (5) we suppose that a single spanning tree is sufficient to solve the problem.

1) Exact algorithms: Modified Spanning Tree Enumeration Algorithms and Topology Enumeration Algorithms (cf. [17]) can be used to find the optimal tree. In the original algorithms, the possible partial spanning trees are enumerated and the tree with minimal cost is selected as the solution. The cost of each tree is computed as the sum of the costs of its edges. In our case, as Formula 5 indicates, this cost is weighted by the factor $f(T)$, which can be unique for each tree. In order to solve our problem, the enumeration algorithms can be applied but the tree with the minimal weighted cost should be selected. Let us notice that, in some cases, this factor $f(T)$ can also be used to eliminate excess trees in the enumeration algorithms. Since the function $f(T)$ is concave and increases rapidly depending on the number $k(T)$ of significant nodes, the optimal solution is probably among the spanning trees having few branching nodes. The complexity of the exact enumeration algorithms is always exponential and in $\mathcal{O}(n^2 2^{n-d-1})$ (where n denotes the number of nodes and d is the number of destinations) [20].

2) Heuristic algorithms: Contrarily, shortest path based heuristics originally proposed to find a 2-approximation for the Steiner problem cannot guarantee the same approximation ratio for the optimization problem (5). Indeed, the "penalty" factor $f(T)$ (which is a function of the number of significant nodes) cannot be included in the shortest path based heuristics.

The following simple example illustrates that a shortest path based Steiner heuristic finds an arbitrarily bad solution for Problem (5). Let there be a topology, a source s and a set of destinations D given as shown in Figure 6. Let us suppose that all the edges have a unit cost and $d = |D|$. In this particular topology, the optimal tree T^* (represented by a dotted line) uses a unique branching node. Shortest paths between the multicast group members do not traverse this node. A shortest path based heuristic (*e.g.,* the Takahashi-Matsuyama heuristic [21]) constructs the tree T_h (the continuous line in the figure).

The costs are $d(T^*) = 2(d + 1)$ and $d(T_h) = 2d$ respectively. Since there are $2d$ significant nodes in T_h and $d + 1$ in T^*, the approximation ratio in this case can be expressed as

$$A = \frac{C(T_h)}{C(T^*)} = \frac{L_{max} - (d+1) \cdot l_a - c}{L_{max} - 2 \cdot d \cdot l_a - c} \cdot \frac{d}{d+1} \quad (15)$$

Increasing the group size d causes this ratio to increase rapidly and an upper-bound cannot be given.

To find trees with low communication cost, we propose a modified version of the Takahashi-Matsuyama algorithm.

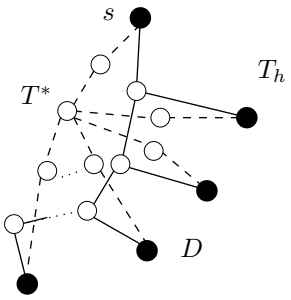


Figure 6. Shortest path based heuristics give an arbitrary solution

A simple objective function can be formulated if the costs (the cost of the usage of the edges and the overhead due to the headers) are expressed by additive metrics. Edge costs are basically additive. Moreover, the use of new branching nodes in the multicast tree can be penalized by additional cost factors. Let B_T be the set of branching nodes of the tree T and let us suppose that the inclusion of a new branching node $v \in B_T$, which is not a destination, corresponds to an additional cost $b(v)$. So, a partial spanning tree resulting in a low communication cost can be obtained by minimizing the sum of edge and node costs :

$$T_D^* : \arg\min_{T \in \mathcal{T}} (d(T) + \sum_{v \in B_T \setminus D} b(v)) \qquad (16)$$

This expression can be considered as an approximation of the communication cost. Trivially, similarly to the Steiner tree problem, this problem is also NP-difficult. The advantage of the formulation (16) is that simple and efficient Steiner heuristics can be adapted to resolve it. Starting from this modified problem formulation we propose a heuristic to compute advantageous partial spanning trees for explicit tree based multicast routing.

Avoidance of Branching node Creation (ABC) algorithm

Following the objective function given by (16), a simple algorithm can be designed by modifying the well-known Steiner heuristic proposed by Takahashi and Matsuyama [21]. In each step of the original algorithm, the nearest destination node is added to the tree using the shortest path from node to tree.

In the modified ABC algorithm, the creation of a new branching node in the tree is penalized. For this reason, the "distance" $\bar{d}(n, T)$ between the tree T and the node n is defined as

$$\bar{d}(n, T) = d(n, m) + \begin{cases} 0 & \text{if } m \in D \cup B_T \\ c & \text{otherwise} \end{cases} \qquad (17)$$

where $m \in T$ is the node connecting n to T, $d(n, m)$ is the distance from m to n and c is the penalty associated with creating a new branching node in the tree. This modification does not affect the favorable complexity of the algorithm. Figure 7 illustrates one step of the algorithm. Let us suppose that each edge has unit cost. The cost of the nodes in the tree T are indicated in the figure. To connect the node n to the tree, the algorithm does not use the shortest path (n, b) but an alternative (the path (n, m)), which connects n to the leaf node m. This connection results in a lower communication cost because new branching nodes are not created.

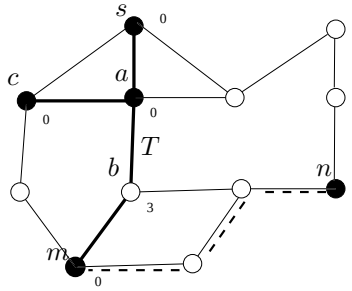


Figure 7. Add a new destination to the tree using the ABC algorithm.

To illustrate the performance of the ABC algorithm, simulation has been performed in the Eurorings topology, which has 43 nodes and 55 edges (cf. an example in [22]). In this topology, the Shortest Path Tree (SPT) algorithm, the Takahashi-Matsuyama (TM) heuristic and the ABC algorithm have been executed for different multicast requests the group size of which varied between 10 and 35. For each group size, 100 groups were generated randomly. Figure 8 shows the number of significant nodes in the computed multicast trees. Supposing a maximal packet size equal to $l_{max} = 1600$, addresses encoded in 128 bits and a constant part in the headers occupying 200 bytes, the communication cost corresponding to the three different trees is illustrated in Figure 9. In this network, the ABC algorithm reduces the communication cost by 10 - 20 % compared to the shortest path tree and the approximated Steiner tree using explicit routing.

B. Exact and heuristic solutions of Problems (10) and (14)

To the best of our knowledge, exact algorithms are not known that solve the recently formulated size-constrained minimum-cost partial spanning problem. Since a single tree is not always sufficient, Steiner Tree Enumeration Algorithms do not work. A trivial exact solution can be proposed as follows.

1) Exact algorithm: As demonstrated in Section III and illustrated with Figure 2, the optimal solution corresponds to an optimal partition of the destination set. So, exact algorithms solving the Set Cover Problem (cf. [19]) can be applied with the following adaptation: the cost associated to

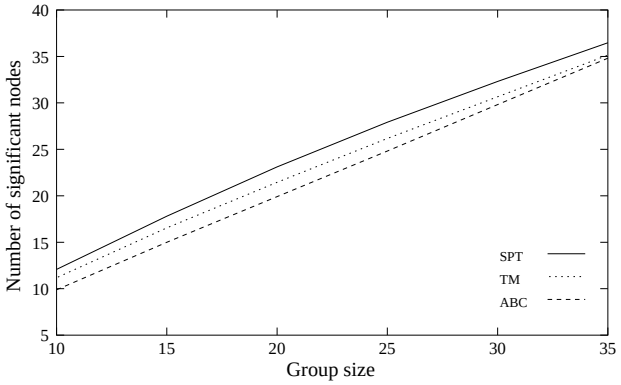


Figure 8. The number of significant nodes in the multicast tree.

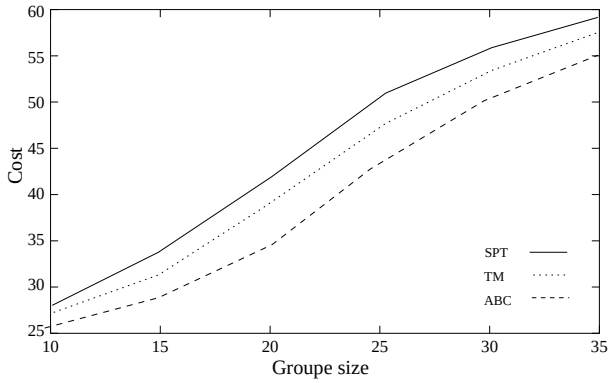


Figure 9. The communication cost associated with the multicast tree.

a sub-set of destinations in the partition is the communication cost of the partial spanning tree minimizing this cost.

Most of the exact algorithms to solve the Set Cover Problem are brute force and dynamic programming based algorithms [23]. In both cases, the associated optimal communication cost per bit must be computed for each sub-set of destination nodes in the partition. For this, a simple modified Steiner Tree Enumeration Algorithm (cf. [17]) can be used as indicated in the previous sub-section.

Let k_{max} be the maximal number of significant nodes in our size-limited spanning problem. The maximal number k of destinations in a spanning tree respecting the size constraint is given by $k(k-1) = k_{max}$. If P_D^k denotes the number of k-limited partitions of D, then the exact algorithm complexity is bounded with $\mathcal{O}(P_D^k 2^{k+1} n^2)$.

2) Some heuristic algorithms: Since the exact computation is very expensive, only heuristic algorithms can compete for potential use in networks. Heuristic solutions can be obtained in two different manners.

- The heuristics in the first group aim to *directly* build a set of trees with respect to the size constraint (moreover, the algorithms can eventually balance the size of the trees).
- The second type of algorithms works in three phases to compute the final solution:
  1) at first a low cost partial spanning tree is computed (regarding the overhead generated by the headers)
  2) then this unique spanning tree is segmented into several trees when the size constraint is exceeded
  3) the size of the trees may also be balanced.

A simple algorithm in the first category can be obtained by modifying the ABC algorithm proposed in the last sub-section.

ABC algorithm with respect to the size constraint

The modification of the ABC algorithm presented in the previous section consists of the insertion of the size constraint. Let k_{max} be the maximum number of significant nodes. In the modified version, the destination associated with the lowest additive cost (in term of edge cost and new significant node creation cost) is added to the tree if and only if the number of significant nodes in the tree under construction is less than k_{max}. Otherwise, a new tree is created by connecting the nearest unspanned destination to the source node.

The second class of heuristics can be designed as follows.

Spanning tree segmentation

- At first, a partial spanning tree computation algorithm is used to compute a tree spanning the destination set (for example the original algorithm of Takahashi and Matsuyama or the original ABC algorithm can be used for this purpose). This unique tree does not necessarily respect the size constraint.
- In the second phase (which is the segmentation of the unique spanning tree), this low cost tree is segmented by distributing the destinations between several sub-trees taking the size constraint into account.
- If the tree set contains unbalanced numbers of significant nodes in the different trees, then a final balancing algorithm can be applied to obtain a balanced tree set.

In the following, we present our proposals for tree segmentation and charge balancing. In the segmentation problem, a tree spanning the entire destination set is given but the number of significant nodes exceeds the size upper bound k_{max}. The result of the segmentation is a set of trees; each tree in the set corresponds to a sub-tree of the delivery tree and the number of significant nodes in each tree is less than the size constraint. The segmentation can also be considered as a particular case of the Set Cover Problem. A heuristic segmentation approach has two potential objectives:

- minimize the number of k_{max}-limited trees

- minimize the overall cost of the set of k_{max}-limited trees covering the original tree.

The solutions obtained by the two different objectives can be different as illustrated in Figure 10, where the first figure shows the original delivery tree. Let us suppose that $k_{max} = 5$. Figure 10/b) presents the result when the number of trees is minimized. There are two trees to span the 8 destination nodes and the total length of this solution is equal to 18. Figure 10/c) illustrates the minimal cost solution under the constraint k_{max}. In this case, there are three trees and the cost is equal to 15.

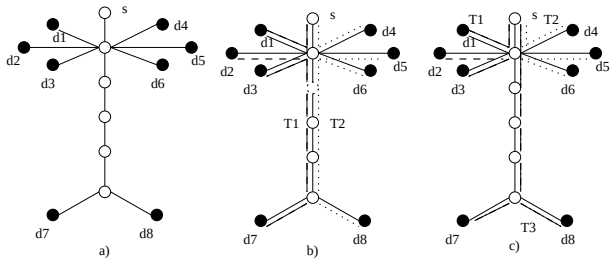


Figure 10. *Segmentation with two different objectives*

In the following, we propose a heuristic solution for this particular Set Cover Problem.

Maximal Common Path First algorithm

The Maximal Common Path First (MCPF) algorithm proposes a new alignment of the destinations in the spanning tree T. To achieve this it uses a new metric $\kappa(d_i, d_j)$ between two destinations d_i and d_j corresponding to the number of common edges of the paths from s to d_i and from s to d_j in T.

$$\kappa(d_i, d_j) = |path(s, d_i) \cap path(s, d_j)|.$$

Using this metric, a complete graph (a special metrical closure) can be computed for the destinations. Figure 11/b) illustrates the metrical closure of the tree presented in Figure 11/a).

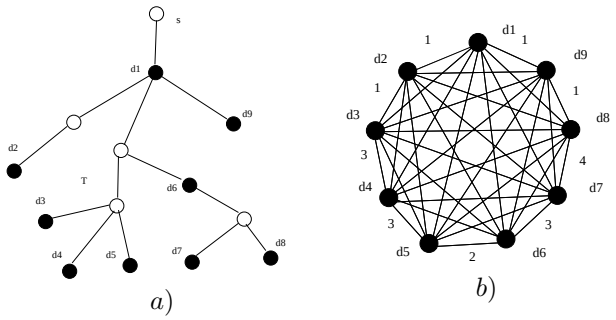


Figure 11. *The metrical closure of the destinations with the new metric*

The MCPF algorithm computes a k_{max}-limited spanning forest in the metrical closure. It is based on the well-known Prim's algorithm and consists of extending a tree T_i started at the source until no new destinations can be added. At each step, the destination with the maximal number of common edges is added to T_i, If a tree is saturated according to the size constraint, a new tree is initiated from the source to the next destination. The pseudo-code of MCPF is given by Algorithm 1.

Algorithm 1 MCPF algorithm using the Prim approach

Require: a tree T spanning the multicast group (s, D), the maximum number k_{max} of significant nodes
Ensure: a set $F = \{T_i, i = 1, \ldots, p\}$ such that each tree T_i has no more than k_{max} significant nodes
Initialization
Build the metrical closure $\bar{G}$ of the set of members D, using the "distance" κ
$F \leftarrow \emptyset$
$i \leftarrow 1$
$T_i \leftarrow$ a new tree initialized with the source s
repeat
  (d, m) is an edge of $\bar{G}$ of maximum value, such as d is in D and m is in T_i
  if $T_i \cup path(d, m)$ has no more than k_{max} branching nodes **then**
    connect d to T_i
    $D \leftarrow M \backslash \{d\}$
    recompute the cost of the edges in $\bar{G}$
  else
    $F \leftarrow F \cup T_i$
    $i \leftarrow i + 1$
    $T_i \leftarrow$a new tree initialized with the source s
  end if
until $D = \emptyset$
$F \leftarrow F \cup T_i$

Let d denote the number of destinations and t the number of trees after segmentation. Let us suppose that the destinations are distributed uniformly in the trees and there are $\lceil d/t \rceil$ destinations per tree. In the worst case, there are $\lceil d/t \rceil - 1$ branching nodes per tree to cover $\lceil d/t \rceil$ destinations. So

$$2 \left\lceil \frac{d}{t} \right\rceil - 1 \leq k_{max} \tag{18}$$

This relation gives the following approximated upper-bound of the number of trees after segmentation :

$$\frac{2d}{k_{max} + 1} \leq t \tag{19}$$

To examine the real number of trees after segmentation with the MCPF algorithm, simulations in the Eurorings topology have been executed. For each group size 100 groups have been generated randomly and a multicast tree has been computed using the Takahashi-Matsuyama algorithm. The size limit k_{max} on the headers has been set to 20. Figure 12 shows the observed number of trees per group after the segmentation.

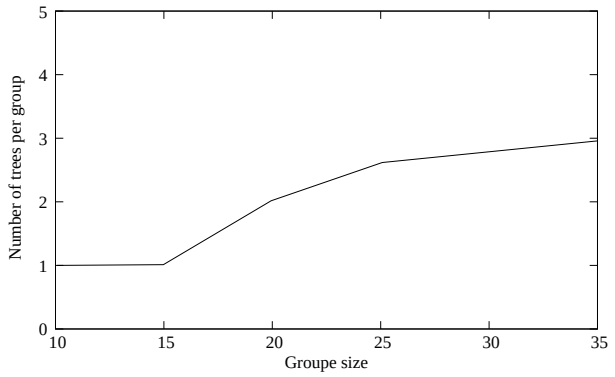


Figure 12. The number of trees after the segmentation by the MCPF algorithm.

The set of trees should be balanced if necessary. For example, using the MCPF algorithm, the resulting trees contain a number of significant nodes near to the given limit k_{max} except the last tree, which may contain only a few members. The equilibrium of the number of significant nodes in the different trees decreases the maximal length of the explicit routing header information and so increases the payload. So, the balancing operation can decrease the total cost of multicast communication.

Member Switching Algorithm

To balance the trees, the Member Switching algorithm considers the tree having the largest encoding. It removes a destination from it, then adds this destination to the tree having the smallest encoding. This process is repeated until the encoding of the largest tree is close to the encoding of the smallest tree.

The worst-case time complexity of the algorithm is $\mathcal{O}(|D| \cdot |N|)$, where $|D|$ is the number of destinations and $|N|$ is the number of nodes of the graph.

V. Conclusions and Future Work

Explicit multicast routing is an alternative solution to resolve the scalability of multicast routing in IP. Flat explicit routing protocols generate significant overhead in routers due to the intensive processing of the datagram headers. Tree-based explicit routing could simplify the task of the routers by encoding the multicast tree in the datagrams and by using conventional unicast data forwarding between the significant nodes of the tree. The computation of the multicast route corresponding to the minimum communication cost per bit is a hard optimization problem. We formulated and illustrated this optimization in several cases: when the multicast group can be spanned with only one tree but also when several trees are needed for the group due to limitations on header size. In this latter case, we introduced the important homogeneous message fragmentation constraint to avoid complicated data transmission procedures at the source. The optimization problems are NP-difficult in these aforementioned cases and well known Steiner heuristics cannot guarantee limited cost solutions. To illustrate the introduced problems, some exact algorithms were presented but they are very expensive. For explicit multicast routing, we also proposed heuristics providing low cost, explicitly encoded multicast routes. These algorithms find approximate solutions in polynomial execution time. In particular, the ABC algorithm permits the construction of multicast trees with low communication cost when the tree should be encoded in the packet headers. If the number of significant nodes is high, tree segmentation and balancing can be performed with good performance using the presented MCPF and Member Switching algorithms.

Algorithm 2 Member Switching Algorithm

Require: a set of trees $F = \{T_i, i = 1, \ldots, p\}$
Ensure: the balanced set F
  repeat
    $T_s \leftarrow$the tree of F of smallest encoding
    $T_l \leftarrow$the tree of F of largest encoding
    if $encoding(T_l) > encoding(T_s) + 2$ **then**
      remove the destination d from T_l such as the significant father of d in T_l has the lowest degree
      add the member d to T_s
    end if
  until $encoding(T_l) \leq encoding(T_s) + 2$

References

[1] M. Molnár, “On the Optimal Tree-Based Explicit Multicast Routing,” in *Second International Conference on Communication Theory, Reliability, and Quality of Service CTRQ, IARIA*, Colmar, France, July 2009.

[2] S. E. Deering, “Multicast Routing in a Datagram Internetwork,” Ph.D. dissertation, Stanford University, December 1991.

[3] I. Stoica and al., “REUNITE: A Recursive Unicast Approach for Multicast,” in *INFOCOM 2000*, March 2000.

[4] L. Costa, S. Fdida, and O. Duarte, "Hop By Hop Multicast Routing Protocol," in *ACM SIGCOMM'01)*, August 2001.

[5] J. Bion, D. Farinacci, M. Shand, and A. Tweedly, "Explicit Route Multicast (ERM)," IETF, Draft, June 2000, draft-shand-erm-00.txt.

[6] M. Bag-Mohammadi, S. Samadian-Barzoki, and N. Yazdani, "Linkcast: Fast and Scalable Multicast Routing Protocol," in *IFIP Networking*, 2004, pp. 1282–1287.

[7] F. K. Hwang, D. S. Richards, and P. Winter, "The Steiner Tree Problem," *Annals of Discrete Mathematics*, vol. 53, 1992.

[8] A. Boudani, A. Guitton, and B. Cousin, "GXcast: Generalized Explicit Multicast Routing Protocol," in *IEEE Symposium on Computers and Communications (ISCC)*, June 2004.

[9] F. Kuipers and P. V. Mieghem, "Conditions that Impact the Complexity of QoS Routing," *IEEE/ACM Transactions on Networking*, vol. 13/4, 2005.

[10] ——, "MAMCRA: A Constrained-Based Multicast Routing Algorithm," *Computer Communications*, vol. 25/8, pp. 801–810, 2002.

[11] T. Braun, V. Arya, and T. Turletti, "Explicit routing in multicast overlay networks," *Computer Communications*, vol. 29, no. 12, pp. 2201–2216, August 2006.

[12] R. Boivie, N. Feldman, and C. Metz, "Small Group Multicast: A New Solution for Multicasting on the Internet," *IEEE Internet Computing*, vol. 4, no. 3, pp. 75–79, May 2000.

[13] R. Boivie, N. Feldman, Y. Imai, W. Livens, D. Ooms, and O. Paridaens, "Explicit multicast (Xcast) basic specification," IETF, Draft, June 2004, draft-ooms-xcast-basic-spec-06.txt.

[14] M.-K. Shin, Y.-J. Kim, K.-S. Park, and S.-H. Kim, "Explicit Multicast Extension (Xcast+) for Efficient Multicast Packet Delivery," *Electronics and Telecommunication Research Institute (ETRI) Journal*, vol. 23, no. 4, December 2001.

[15] L. Ji and M. S. Corson, "Explicit Multicasting for Mobile Ad Hoc Networks," *Mobile Networks and Applications*, vol. 8, pp. 535–549, 2003.

[16] M. Bag-Mohammadi and S. Samadian-Barzoki, "On the efficiency of explicit multicast routing protocols," in *ISCC '05: Proceedings of the 10th IEEE Symposium on Computers and Communications*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 679–685.

[17] S. L. Hakimi, "Steiner's Problem in Graphs and its implications," *Networks*, vol. 1, pp. 113–133, 1971.

[18] M. Molnár, "Hierarchies for Constrained Partial Spanning Problems in Graphs," IRISA, Rennes, France, Tech. Rep. 1900, 2008.

[19] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher, Eds. Plenum Press, 1972, pp. 85–103.

[20] P. Winter, "Steiner problem in networks: A survey," *Networks*, vol. 17, pp. 129–167, 1987.

[21] H. Takahashi and A. Matsuyama, "An approximate solution for the Steiner problem in graphs," *Mathematica Japonica*, vol. 24, no. 6, pp. 573–577, 1980.

[22] J. Moulierac, A. Guitton, and M. Molnár, "Multicast Tree Aggregation in Large Domains," in *IFIP Networking, Springer-Verlag, LNCS*, no. 3976, January 2006, pp. 691–702.

[23] F. V. Fomin, F. Grandoni, and D. Kratsch, "Measure and Conquer: Domination - A Case Study," in *32nd International Colloquium on Automata, Languages and Programming (ICALP 2005), Springer LNCS vol. 3580*, 2005, pp. 191–203.

Comparison of Packet Switch Architectures and Pacing Algorithms for Very Small Optical RAM

Onur Alparslan, Shin'ichi Arakawa, Masayuki Murata
Graduate School of Information Science and Technology
Osaka University
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
{*a-onur,arakawa,murata*}*@ist.osaka-u.ac.jp*

***Abstract*—One of the difficulties with optical packet switched (OPS) networks is buffering optical packets in the network. The research on optical RAM presently being done is not expected to achieve a large capacity soon. However, the burstiness of Internet traffic causes high packet drop rates and low utilization in small buffered OPS networks. In this article, we investigate and compare optical-buffered switch architectures and pacing algorithms for minimizing the buffer requirements of OPS switches. We simulate two mesh topologies (NSFNET and Abilene) for goodput and packet drop rate comparisons and optimization of XCP parameters. We show that XCP-based pacing algorithm with a shared buffered switch architecture yields high TCP goodput and low packet drop rate in a core OPS network when very small optical RAM buffers are used.**

***Keywords*-small buffer, OPS, optical RAM, optical switch**

I. Introduction

A well-known problem in realizing optical packet switched (OPS) networks is buffering. Recent advances in optical networks such as dense wavelength division multiplexing (DWDM) have allowed us to achieve ultra-high data-transmission rates in optical networks. This ultra-high speed of optical networks has made it necessary to do some basic operations like buffering and switching in the optical domain instead of the electronic domain due to high costs and limitations with electronic buffers. However, the lack of high-capacity optical RAM makes it difficult to buffer enough optical packets in OPS networks [1]. According to a rule-of-thumb [2], the buffer size of a link must be $B = RTT \times BW$, where RTT is the average round trip time of flows and BW is the bandwidth of the output link, to achieve high utilization with TCP flows. However, as this requires a huge buffer size in optical routers due to the ultra-high speed of optical links, this buffer size is unfeasible.

The only available solution that can currently be used for buffering in the optical domain is using fiber delay lines (FDLs), where contended packets are switched to long FDLs so that they can be delayed. However, FDLs pose severe limitations such as signal attenuation, and bulkiness. Most FDL architectures lack the real $O(1)$ reading operation of RAM as it may not be possible to access a packet circulating an FDL until the packet departs the fiber and arrives back to the switch, which causes extra delays depending on the FDL length. Moreover, all-optical RAMs, which can solve the problems with FDLs, are still being researched (e.g., NICT project [3]) and this may become available in the near future. Furthermore, optical RAMs are expected to have a lower rates of power consumption, which is a serious problem with electronic RAMs. However, optical RAMs are not expected to attain large capacities. Therefore, it is necessary to decrease the buffer requirements of OPS networks to make use of optical RAMs.

Appenzeller et al. [4] recently demonstrated that when there are many TCP flows sharing the same link, a buffer sized at $B = \frac{RTT \times BW}{\sqrt{n}}$, where n is the number of TCP flows passing through the link, is sufficient to achieve high utilization. However, there should be many flows on a link to significantly decrease the buffer requirements of ultra-high-speed optical networks. Enachescu et al. [5] proposed that $O(\log W)$ buffers are sufficient where W is the maximum congestion window size of flows when packets are sufficiently paced by replacing TCP senders with paced TCP [6] or by using slow access links. TCP pacing is defined as transmitting ACK (data) packets according to special criteria, instead of immediately transmitting packets when data (ACK) packets arrive [6]. Paced TCP is usually implemented by evenly spreading out the transmission of a window of data packets over a round-trip time. However, the $O(\log W)$ buffer size depends on the maximum congestion window size of TCP flows, which may change. Moreover, using slow access links, which is an extreme way of applying node pacing, is not ideal when there are applications that require large amounts of bandwidth on the network. Furthermore, replacing TCP senders of computers with paced versions can be difficult. Furthermore, this proposal was based on the assumption that most IP traffic is from TCP flows. Theagarajan et al. [7] demonstrated that even small quantities of bursty real-time UDP traffic can increase the buffer requirements of well-behaved TCP traffic on the same link. Therefore, it may be better to design a general architecture for an OPS network that can achieve high utilization in a small buffered OPS network independent of the number of TCP or UDP flows, and that does not require a strict limit to be placed on the speed of access links, and that does not require the

sender or receiver TCP and UDP agents of computers using the network to be replaced.

We recently proposed [8] an all-optical OPS network architecture that can achieve high utilization and a low packet drop ratio by using small buffering. We took into consideration an OPS domain where packets entered and exited the OPS domain through edge nodes. We proposed using an Explicit Congestion Control Protocol (XCP)-based [9] intra-domain congestion control protocol to achieve high utilization and a low packet-drop ratio with small buffers. XCP is a new congestion control algorithm using a control-theory framework, which was specifically designed for high-bandwidth and large-delay networks. XCP was first proposed by Katabi et al. [9] as a window-based algorithm for reliably controlling congestion and transmission. We selected the XCP framework because it allows the utilization level of each wavelength to be individually controlled. Moreover, there is no need to modify the TCP and UDP agents of computers or limit the speed of access links to decrease burstiness.

Another difficulty in attaining OPS networks is the switching fabric, which is usually one of the biggest factors determining overall router cost. Many switching fabric architectures like MEMS, optomechanical, electrooptic, thermooptic, and liquid-crystal based switches have been proposed for optical switching [10]. However, the number of switching elements in the fabric increases together with the overall cost, and crosstalk and insertion losses as the number of ports for the switch increases. In our previous papers, we evaluated the performance of our proposed architecture with UDP and TCP-based traffic and output buffering [8][11][12]. In Alparslan et al. [1], we proposed and investigated different optical-buffered switch architectures to further minimize the size of the optical switching fabric of core nodes while achieving higher goodput with small optical RAM buffers. We evaluated the optical RAM requirements of our proposed architecture on a mesh NSFNET topology with TCP traffic. We also compared the performance of our architecture with paced TCP, which is the solution generally proposed for small buffered networks. Our simulations revealed that the average goodput of standard TCP flows in our proposed architecture could even surpass the goodput of paced TCP when buffers were small.

This article discusses the extension of our work and presents our results in Alparslan et al. [1] verified by simulating an Abilene network, which is a larger topology with a higher nodal degree than NSFNET. We optimized XCP parameters on both NSFNET and Abilene topologies to demonstrate what effect XCP parameters have on overall performance. Moreover, we introduce one more metric, which is the overall packet-drop rate in a small buffered core network, that enables better comparison of switch architectures and algorithms.

The rest of the paper is organized as follows. Section II describes the XCP algorithm and switch architectures. Section III describes the simulation methodology and presents the simulation results. Finally, we conclude the paper and describe future work that we intend to do in Section IV.

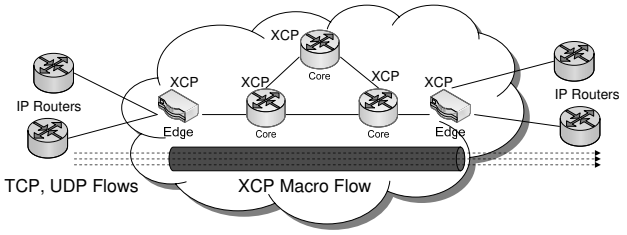


Figure 1. XCP pacing

II. ARCHITECTURE

This section describes the XCP algorithm and switch architectures.

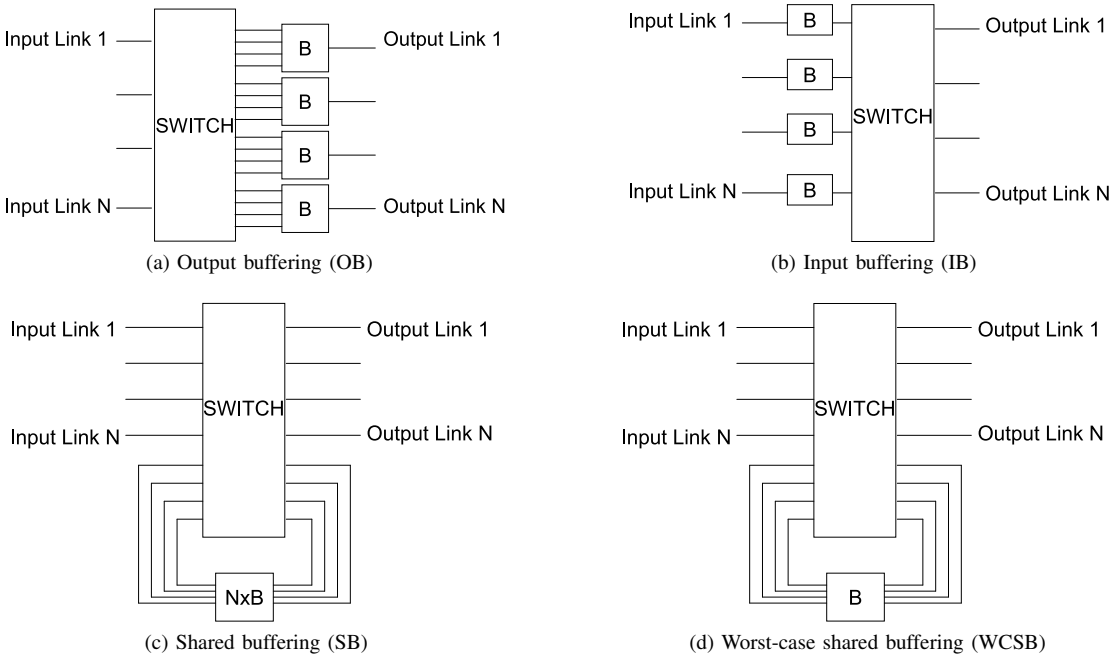
A. Optical Rate-based Paced XCP

XCP is a new congestion control algorithm that has been specifically designed for high-bandwidth and large-delay networks. XCP makes use of explicit feedback received from the network. Core routers are not required to maintain per-flow state information. Each XCP core router updates its control decisions calculated with an Efficiency Controller (EC) and a Fairness Controller (FC) when timeout of a per-link control-decision timer occurs.

EC controls input aggregate traffic to maximize link utilization. A required increase or decrease in aggregate traffic for each output port is calculated by using the equation, $\Phi = \alpha \cdot S - \beta \cdot Q/d$, where Φ is the total amount of required change in input traffic, α and β correspond to spare bandwidth-control and queue-control parameters, and d is the control-decision interval. S is the spare bandwidth that is the difference between the link capacity and input traffic in the last control interval. Q is the persistent queue size.

After EC has calculated the aggregate feedback Φ, FC fairly distributes this feedback to flows according to AIMD-based control. However, convergence to fairness may take a long time when Φ is small. Bandwidth shuffling, which redistributes a small amount of traffic among flows, is used to solve this problem. The amount of shuffled traffic is calculated by $h = max(0, \gamma \cdot u - |\Phi|)$, where γ is the shuffling parameter and u is the rate of aggregate input traffic in the last control interval.

In Alparslan et al. [8], we proposed optical rate-based paced XCP, which is a modified version of XCP adapted to work as an intra-domain traffic shaping and congestion control protocol in an OPS network domain. In our architecture, when there is traffic between an edge-source destination node pair, a rate-based XCP macroflow is created as shown in Figure 1, and the incoming TCP and UDP packets of



(a) Output buffering (OB)

(b) Input buffering (IB)

(c) Shared buffering (SB)

(d) Worst-case shared buffering (WCSB)

Figure 2. Switch architectures

this edge pair are assigned an XCP macroflow similar to TeXCP [13]. The edge nodes of the OPS network apply leaky-bucket pacing to the macroflows by using the rate information provided by XCP to minimize burstiness.

In our optical rate-based paced XCP, XCP feedback is carried in separate probe packets that XCP sender agents only send once in every control period. As there is no feedback information carried in the header of data packets, there is no need to calculate per-packet feedback in core routers unlike in the original XCP [9]. We separated the control channel and data channels. Probe packets are carried on a separate single control wavelength that is sufficiently slow to only carry probe packets. The low transmission rate to control wavelength allows electronic conversion to be applied to update the probe feedback and buffer the probe packets in an electronic RAM in case of a contention.

When a probe packet of macroflow i arrives at a core router, the XCP agent responsible for controlling the wavelength of i calculates positive feedback p_i and negative feedback n_i for macroflow i. Positive feedback is calculated as

$$p_i = \frac{h + max(0, \Phi)}{N} \tag{1}$$

and negative feedback is calculated as

$$n_i = \frac{u_i \cdot (h + max(0, -\Phi))}{u}, \tag{2}$$

where N is the number of macroflows on this wavelength, u_i is the traffic rate of flow i estimated and sent by the XCP sender in the probe packet, and h is the shuffled bandwidth. N can be estimated by counting the number of probe packets received during the last control interval. Another possible method is using the number of LSPs if GMPLS is available [13]. The control interval is the maximum RTT in the network. The control interval can be selected to be a bit longer than the maximum RTT to compensate for the processing and buffering delays in control packets. Feedback, which is the required change in the flow rate, is calculated as $feedback = p_i - n_i$. If this feedback is smaller than that in the probe packet, the core router replaces the feedback in the probe packet with its own feedback. Otherwise, the core router does not change the feedback in the probe packet.

B. Switch Architectures

In this paper, we compare the performance of output buffering (OB), input buffering (IB), shared buffering (SB) and worst-case shared buffering (WCSB), as shown in Figure 2. The internal speedup is 1 in all switches, which means the line rates are equal in both inside and outside the switch. Switch size is shown as $I \times O$, where I and O are the number of input and output ports, respectively. Output buffering has a large switch size of $N \times N^2$ to prevent internal blocking where N is the nodal degree as seen in Figure 2a. It has buffer size B at each output link. As input buffering has a switch size of only $N \times N$, as seen in Figure 2b, it has the smallest switches. However, a well-known problem with input buffering is head-of-line blocking, which limits the utilization that can be accomplished. We applied virtual output queuing (VOQ) scheduling, which is the

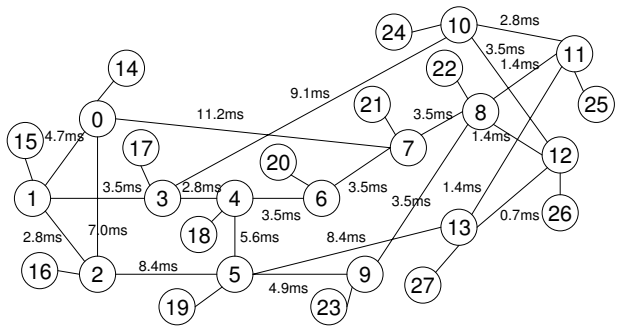


Figure 3. NSFNET

generally proposed solution in the literature, by dividing each input buffer into N sub-queues to minimize the head-of-line blocking problem [14]. Input buffering has buffer size B at each input link. Shared buffering and worst-case shared buffering have a switch size of $2N \times 2N$. As shared buffering has a single buffer with a size of $N \cdot B$, this is in direct proportion to its capacity and the nodal degree. Shared buffering has the same total buffer size per node as input and output buffering to enable fairer comparisons with them at the same B value in the simulations. As worst-case shared buffering has a single buffer with a size of B independent of the nodal degree, it has a buffer capacity of only a single link in input or output buffering.

III. Evaluation

This section describes the simulation methodology and presents the simulation results. We investigated and compared optical-buffered switch architectures and pacing algorithms for minimizing the buffer requirements of OPS switches. We first simulated a wide range of XCP parameters to find the optimum XCP parameters for two mesh topologies. In the second step, we simulated both topologies with four different switch architectures to compare their performance under standard TCP traffic, XCP-paced standard TCP traffic, and paced TCP traffic.

A. Simulation Settings

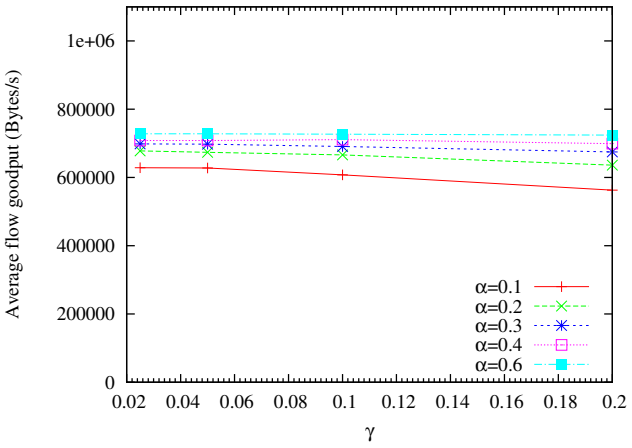
The proposed network architecture and buffering models were implemented over the *ns* simulator version 2.32 [15]. The simulator used cut-through packet switching and buffering for data wavelengths. There was a single store-and-forward switching slow control wavelength dedicated to probe packets. The edge nodes had 0.25 s of electronic buffering, which is commonly used on the Internet. However, core routers only used small optical RAM for buffering optical packets on data wavelengths. The contention of probe packets on the control-wavelength was resolved by an electronic RAM as O/E/O conversion was not a problem in the control-wavelength due to its low speed. As the IP datagram for TCP data (ACK) packets was 1500 Bytes (40 Bytes), the TCP data (ACK) segment was 1480 Bytes (20 Bytes). Maximum segment size (MSS) was 1500 Bytes. The size limit of the congestion window was 20 packets. The target utilization (TU) parameter of XCP was set to 100% at both core and edge links to maximize link utilization.

B. NSFNET Results

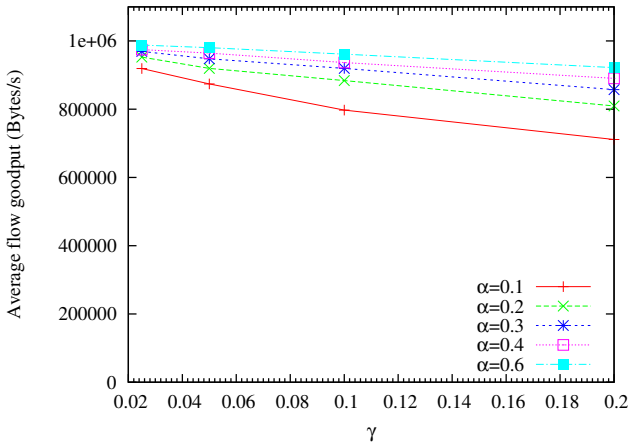
Figure 3 plots the simulated NSFNET topology. The nodes numbered from 0 to 13 are the core nodes and the rest are the edge nodes connected to the core nodes. All links (including edge and core links) had a single data wavelength and the same XCP target utilization. We selected a propagation delay for links between the core and edge nodes of 0.1 ms. We selected an XCP control period for core routers and the sending interval of probe packets for edge routers of 50 ms by taking extra processing and queuing delays in the core routers into account. The capacity of the data and XCP control wavelengths were set to 1 Gbps for the former and 100 Mbps for the latter. TCP Reno traffic was applied between the edge nodes of the network. Throughout the simulation, 1586 TCP flows entered the network between randomly selected edge-node pairs according to a Poisson arrival process. The total simulation duration was 40 s. Only the simulation results in the last 5 s were used for the evaluation.

1) Optimization of XCP Parameters: XCP's α, β and γ parameters were chosen as 0.2, 0.056, and 0.1 in Alparslan et al. [1]. In this paper, we first optimized XCP parameters to demonstrate what effect XCP parameters had and to increase overall performance. Figure 4 plots the average goodput of TCP flows under XCP pacing and packet drop rate at the core nodes when shared buffering is used. The x-axis is the γ parameter and each line plots a different α value. Figures 4a and 4c use a small shared buffer with a size of 1 KByte per link, while Figure 4b uses a larger shared buffer with a size of 100 KBytes per link. As no packets were dropped in the core when the shared buffer size per link was 100 KBytes, this has not been plotted. Figures 4a and 4c indicate that when the buffer is small, the average flow goodput and core packet drop rate increase with increasing α values, while they are relatively less sensitive to the γ parameter. Therefore, there is a tradeoff between the drop rate and goodput as we change α. This is an expected result, because XCP behaves more aggressively and utilizes links faster with an increasing α value, which causes more frequent buffer overflows and thus packet drops as a side effect in the core routers. However, XCP may become unstable when α is too large. Katabi et al. [9] proved that the system is stable independent of delay, capacity, or the number of sources when α and β satisfy

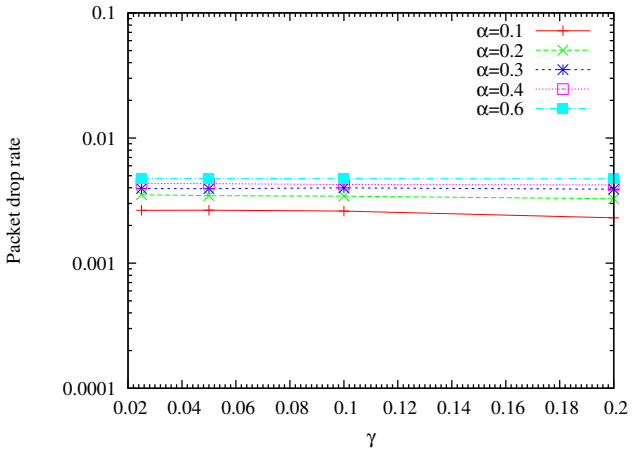
$$0 < \alpha < \frac{\pi}{4\sqrt{2}} \tag{3}$$



(a) Goodput (1 KByte buffer per link)



(b) Goodput (100 KBytes buffer per link)



(c) Packet drop rate in the core nodes (1 KByte buffer per link)

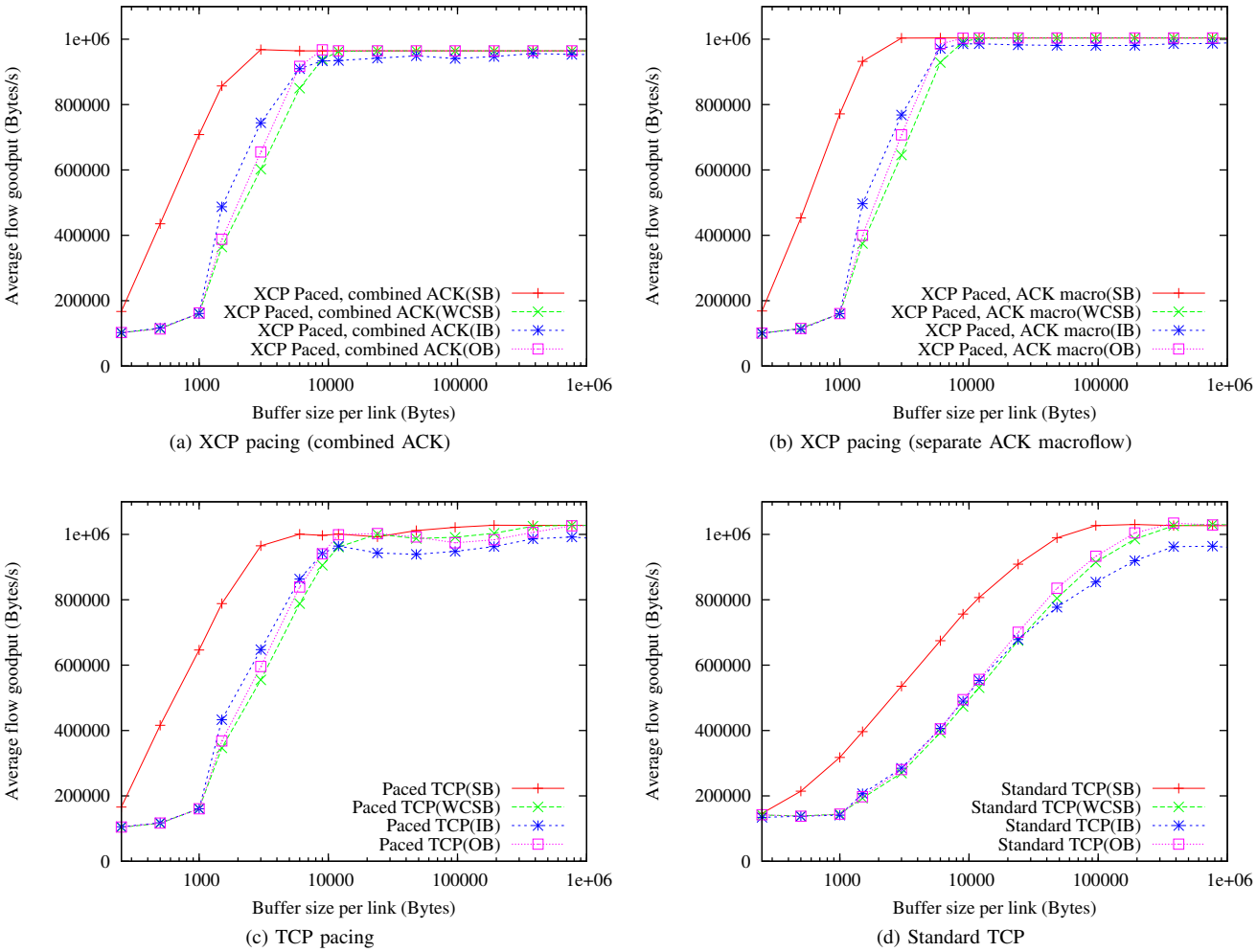
Figure 4. Optimization of parameters in NSFNET

and

$$\beta = \alpha^2\sqrt{2}. \quad (4)$$

Figure 4b shows that when the buffer is larger, goodput becomes more sensitive to the γ parameter. As γ decreases or α increases, the average goodput increases. The reason for this is that XCP encounters a well-known max-min fairness problem under various special conditions. XCP's mechanism to control congestion in a multi-bottleneck environment can cause a flow to receive an arbitrarily small fraction of its max-min allocation, which may cause some bottleneck links to be under-utilized [16]. XCP may end up utilizing only 80% of the bottlenecked bandwidth in the worst case with the default XCP parameters in Katabi et al. [9]. As they give a high goodput in Figure 4, we chose α=0.4 and β=0.226, which are the default values suggested in Katabi et al. [9]. However, we chose γ=0.1, which is half the default value in Katabi et al. [9]. Although choosing a lower γ than the default value decreases the speed of fairness convergence, XCP achieves better max-min fairness and higher worst-case link utilization with higher average goodput as seen in Figure 4b.

2) Comparison of Switches and Algorithms: After XCP parameters were optimized, we simulated the NSFNET topology by using the four different switch architectures shown in Figure 2. The switch architecture was a parameter in the simulations, which was applied to all nodes in the network. We compared the performance of the switch architectures under standard TCP, paced TCP, and XCP-paced standard TCP traffic. Figure 5 plots the average goodput of TCP flows on different switch and network architectures based on the optical RAM buffer size per link. In all the figures, the x-axis is the buffer size per link, which is designated as B in Figure 5 on a log scale and the y-axis is the average TCP goodput on a linear scale. Figure 5a plots the TCP goodput when our XCP pacing algorithm was applied to standard TCP Reno traffic. We



(a) XCP pacing (combined ACK)

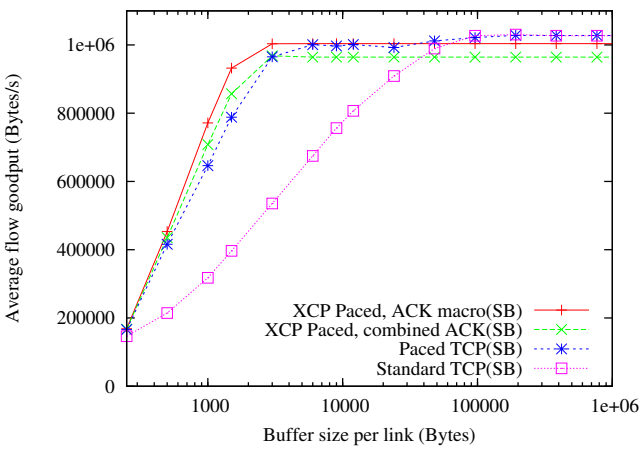
(b) XCP pacing (separate ACK macroflow)

(c) TCP pacing

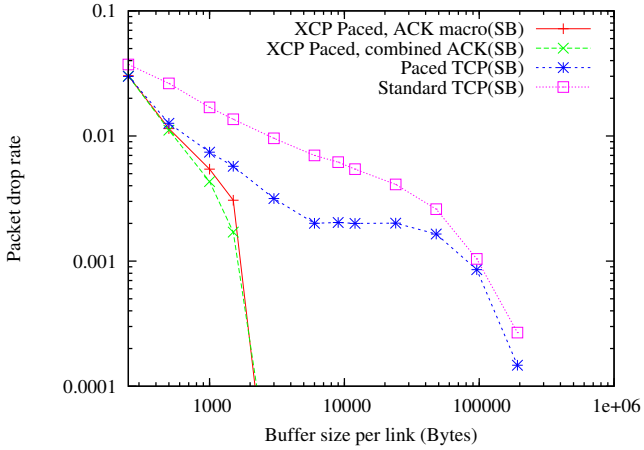
(d) Standard TCP

Figure 5. Goodput comparison of algorithms and switch architectures in NSFNET

can see that input, output, and worst-case shared buffered switches yield almost the same goodput when the buffer size is less than a 1 MSS or more than around 6 MSS. However, shared buffering yields a much higher goodput than the others even though its per node buffer capacity is the same as that for input and output buffering. If we can use a single shared buffer instead of splitting it into input or output links, it clearly yields much higher goodput as a small buffer capacity is being used with maximum efficiency. When we compare input and output buffering, we can see that when the link buffer is between 1–6 MSS, input buffering yields higher goodput while using the smallest switch in switch architectures. This result was expected, because input buffering can handle packet contentions better than output buffering when the input traffic is sufficiently smooth. For example, let us assume that we have a switch with only single packet capacity output buffers. When there is contention with five packets arriving from five input links that are going to the same output link, if the buffers and links are idle, one packet will be sent to the output link, one packet will be buffered in the output buffer, and the remaining three packets will be dropped as there is no more buffer left. However, if we use an input buffered switch, one packet will be sent to the output link and the other four packets will be buffered at the input ports. As buffered packets can be sent to the output link as the link becomes idle, its tendency to drop packets when there is a contention is lower than that for output buffering. Input buffering greatly benefits from pacing as it smooths the packet arrival from its link, which gives it time to drain its queue. When we check the goodput of worst-case buffering, we can see that it is very close to output buffering even though the whole switch in worst-case buffering has the same buffer capacity of only a single link in output buffering. In other words, worst-case shared buffering yields almost the same goodput as output buffering with a much smaller buffer capacity per node.



(a) Goodput



(b) Core packet drop rate in the core nodes

Figure 6. Comparison of algorithms in NSFNET with shared buffering

The packet level tracing of simulations with our XCP pacing algorithm revealed that there was actually still room for improvement. We saw that the well-known ACK compression problem [17] had caused some utilization inefficiencies, which decreased the average goodput. It is possible in our architecture to solve this problem and increase utilization by simply using separate XCP macroflows for TCP ACK packets [12]. Figure 5b plots the TCP goodput when our optical rate-based paced XCP architecture was used with separate XCP macroflows for TCP ACK packets on the same wavelength. We can see that TCP goodput becomes higher than XCP with the combined ACK architecture in Figure 5a.

Figure 5c plots the TCP goodput when paced TCP Reno is used without XCP control. We can see that its goodput pattern is very similar to that in Figures 5a and 5b when the buffer size per link is less than around 6 MSS. When the buffer size per link was larger than 6 MSS, the simulations yielded some varying results. Figure 5d plots the TCP goodput when standard TCP Reno was used without XCP control. We can see that it has much lower goodput than the simulated paced architectures had. More than 10-fold buffering is necessary to achieve utilization that is as high as that of paced architectures. When the buffer is small, the goodput of input buffering is almost the same as that of output buffering, which indicates that pacing is necessary to surpass the goodput of output buffering.

As shared buffering has the highest goodput for all simulated architectures, we did a general comparison of algorithms with shared buffering. Figure 6 plots the average goodput and core packet drop rate of XCP-paced standard TCP, paced TCP, and standard TCP on a shared buffered switch architecture based on the optical RAM buffer size per link. We can see that when the buffer is small, XCP pacing methods yield higher goodput and lower packet drop rates in the core nodes than paced and standard TCP.

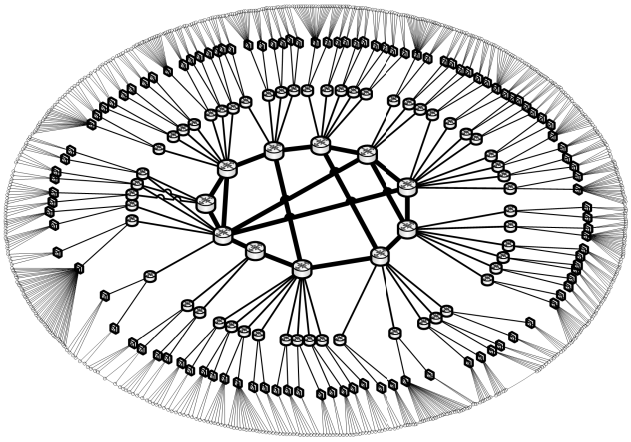
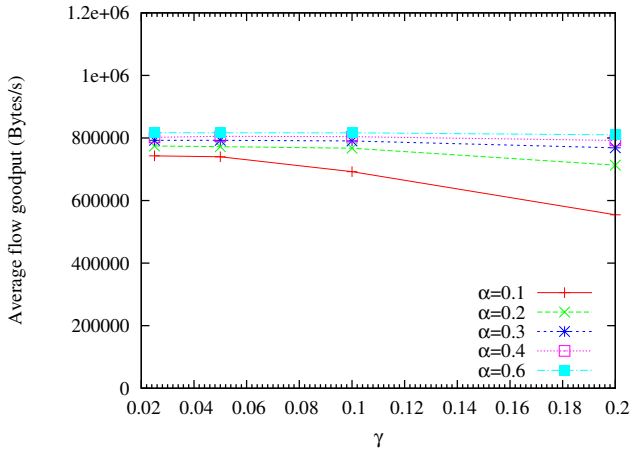


Figure 7. Abilene topology

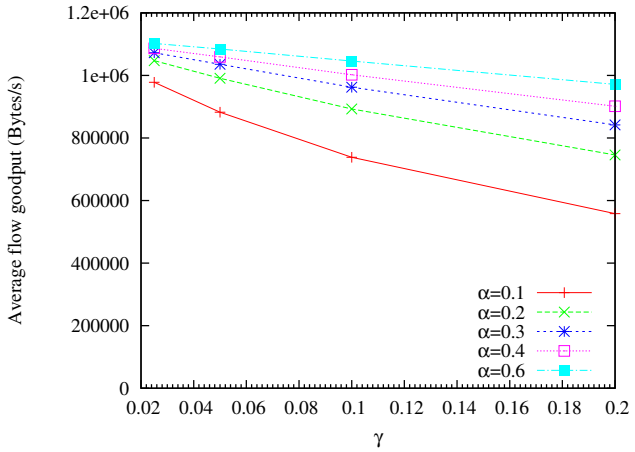
C. Abilene Topology Results

Abilene is an Internet backbone network for higher education and part of the Internet2 initiative. The Abilene-inspired topology in Figure 7 from Li et al. [18] was used in the simulations. The topology has a total of 869 nodes that are divided into two groups of 171 core nodes and 698 edge nodes. A total of 2232 TCP flows started randomly and sent traffic between randomly selected edge node pairs. The total simulation time was 40 s. There was a single data wavelength on the links. The propagation delay of the edge and core links corresponded to 0.1 ms and 1 ms. All the links had a 1 Gbps capacity.

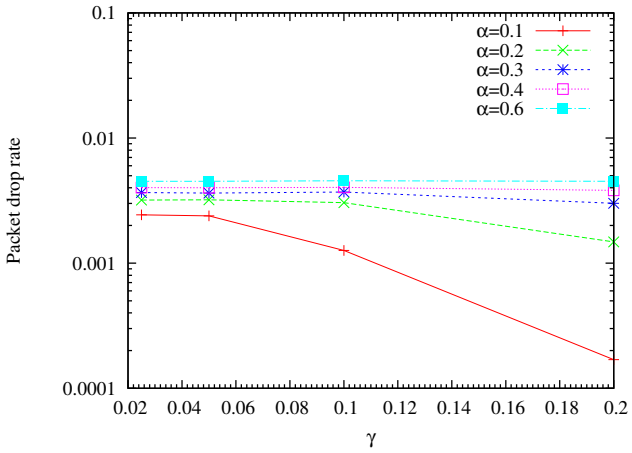
1) Optimization of XCP Parameters: First, we simulated a range of α and γ parameters to optimize the XCP parameters on the Abilene topology. Figure 8 plots the average



(a) Goodput (1 KByte buffer per link)



(b) Goodput (100 KBytes buffer per link)



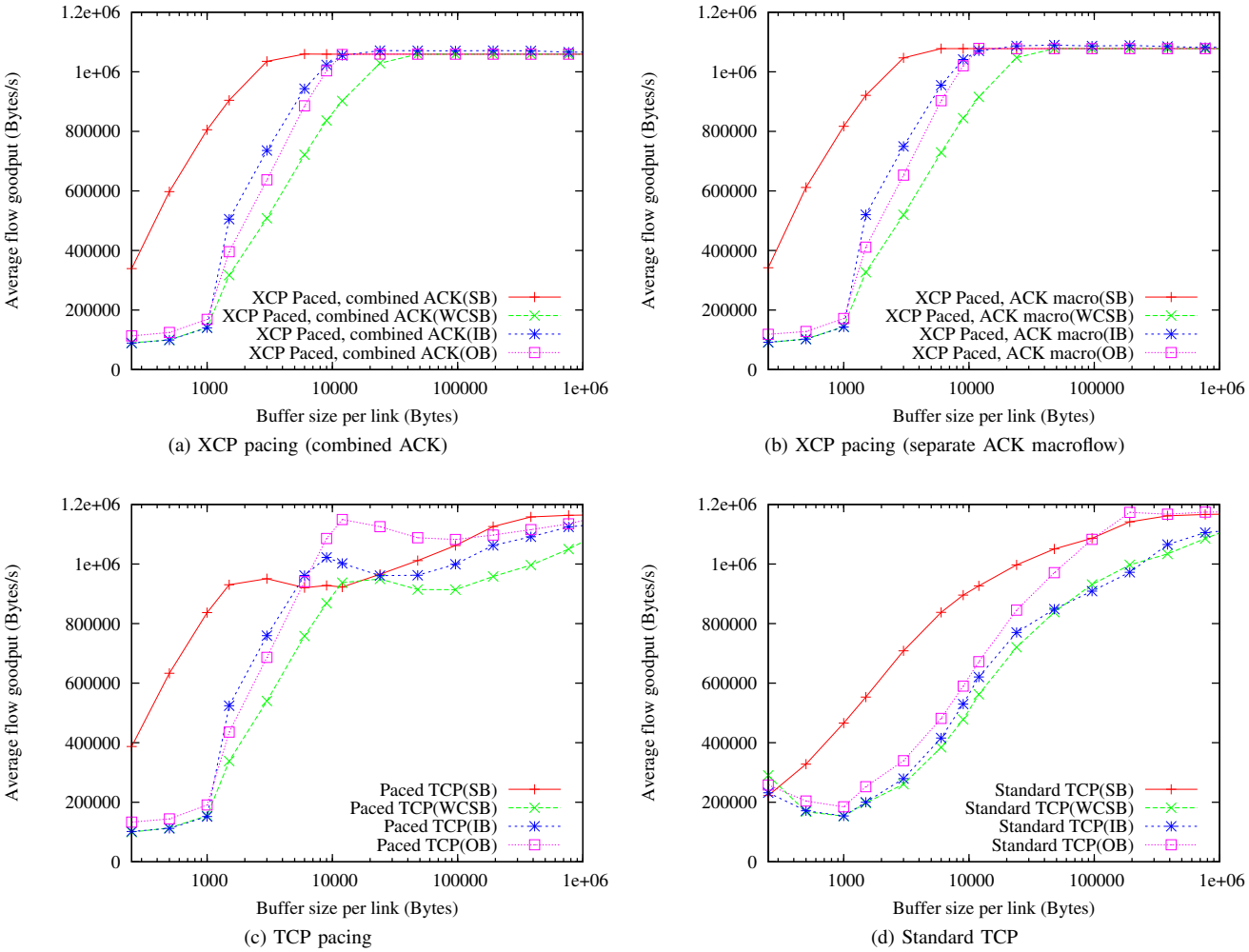
(c) Packet drop rate in the core nodes (1 KByte buffer per link)

Figure 8. Optimization of parameters in Abilene topology

goodput of TCP flows and packet drop rate at the core nodes when shared buffering was used. We can see that both the average goodput and packet drop rate in the core nodes increase with increasing α as occurs in optimizing the NSFNET topology. However, under-utilization due to the max-min fairness problem is more visible in the Abilene topology in Figure 8b. Changing the α or γ parameter yields a wider change in goodput than in NSFNET. Figure 8 shows that the α, β, and γ values selected in Sec. III-B give high goodput, so we chose the same values as those in NSFNET. If the time for fairness convergence is not a concern, a lower γ value can be chosen to further increase goodput.

2) Comparison of Switches and Algorithms: Figure 9 plots the average goodput of TCP flows on different switch and network architectures based on the optical RAM buffer size per link. We can see that the goodput plots of XCP pacing in the Abilene topology in Figures 9a and 9b are similar to those of the NSFNET simulations in Figures 5a and 5b. However, the goodput gap between worst-case shared buffering and output buffering is higher due to the higher nodal degree of the Abilene topology. Figure 9c shows that as we increase the buffer size, the goodput of TCP pacing in the Abilene topology yields even wider fluctuations than those in NSFNET. It seems that output buffering can handle the paced TCP traffic better than other switch architectures and even surpasses the goodput of shared buffering greatly when the buffer is small. Figure 9d shows that output buffering with standard TCP yields a performance boost over input buffering in the Abilene topology, because output buffering can handle bursty TCP traffic in a node with a high nodal degree better.

As a last step, we did a general comparison of algorithms in the Abilene topology with shared buffering. In Figure 10, we can see that when the buffer is very small, XCP pacing methods yield almost the same goodput as TCP pacing. However, when the shared buffer is larger than 1 MSS,



(a) XCP pacing (combined ACK)

(b) XCP pacing (separate ACK macroflow)

(c) TCP pacing

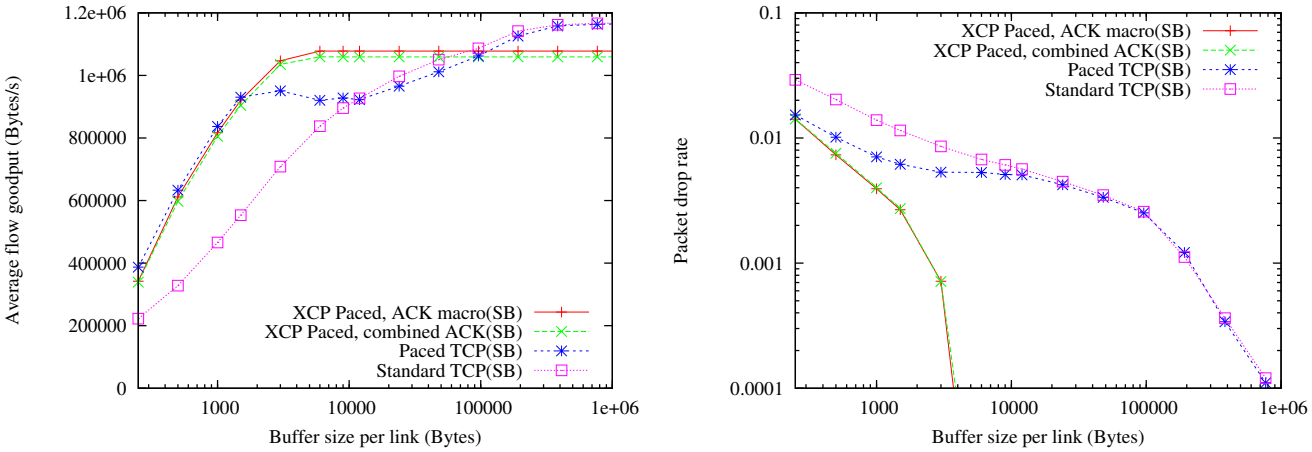
(d) Standard TCP

Figure 9. Goodput comparison of algorithms and switch architectures in Abilene topology

the goodput of TCP pacing stalls and XCP pacing methods yield higher goodput than TCP pacing. XCP pacing methods reach their maximum goodput when around 2–3 MSS per link of shared buffer is used. However, paced TCP and standard TCP require around 30–60 MSS per link of shared buffer to reach the goodput of XCP pacing methods. When the buffer is larger, standard TCP and paced TCP achieve slightly higher utilization than XCP due to the max-min fairness problem with XCP, which causes some bottleneck links to become under-utilized. However, as our aim is to increase performance with small buffers, the difference in goodput with such large buffers is not a concern. Optical RAM is not expected to attain a large capacity, so shared buffering is good. Figure 10b shows that XCP pacing yields a lower packet drop rate in the core nodes than paced or standard TCP just like in the NSFNET simulations. When we compare the two XCP pacing methods in Figure 10, we can see that their goodput and packet drop rates are very close, which indicates that the ACK compression problem in the Abilene topology is much less than that in NSFNET.

IV. Conclusion and Future Work

We investigated and compared optical-buffered switch architectures and pacing algorithms for minimizing the buffer requirements of OPS switches. By using two mesh topologies, our simulations revealed that even under bursty TCP traffic, using our architecture based on optical rate-based paced XCP, which is a modified version of XCP adapted to work as an intra-domain traffic shaping and congestion control protocol in an OPS network domain, could yield equal or higher TCP goodput and lower packet drop rates in the core nodes than using paced TCP, which is the solution that has generally been proposed in the literature. Moreover, simulations in the Abilene topology revealed that the goodput of paced TCP might exhibit some fluctuating behaviors, which adversely affect its performance even with relatively small buffers.



(a) Goodput

(b) Packet drop rate in the core nodes

Figure 10. Comparison of algorithms in Abilene topology with shared buffering

There are many advanced switch architectures in the literature like combined input output buffered switches, but most of them have have high scheduling complexity, which may become a bottleneck at ultra high speed of optical networks, so we limited our work to simpler architectures with lower scheduling complexity. When we compared the small buffered switch architectures simulated in this work, we could see that shared buffering yielded much higher TCP goodput than input or output buffering as it used small buffer capacity in the node more effectively. When the traffic was paced, input buffering yielded higher TCP goodput than output buffering while using much smaller switches. In NSFNET topology, where the nodal degree of nodes was small, worst-case shared buffering yielded almost the same goodput as output buffering with a much smaller buffer capacity per switch. Therefore, output buffering looks as though it is the worst choice as a switch architecture for small buffered optical networks with a low nodal degree.

Overall, the combination of applying XCP pacing and using shared buffered switches generally yielded the highest performance in terms of goodput and packet drop rate for small buffered OPS networks. Our XCP pacing proposal only operates at the edge/core routers of OPS domains and there are still no optical-RAM-buffered OPS networks deployed on the Internet, so it can be applied to OPS networks when they become commercially available, while paced TCP solution is harder to deploy as this requires replacing the TCP stack of computers on the Internet.

As a future work, we will work on the max-min fairness problem of XCP, which causes some bottleneck links to be under-utilized. Our work has revealed that shared buffering requires much less buffering than input and output buffering, but the required buffer size is not clear. Therefore, we will try to formulate the relationship between the number of wavelengths, traffic type, nodal degree, and the required shared buffer size. We intend to simulate other state-of-the-art congestion control algorithms and pacing methods with different types of traffic to gain a broader understanding of their performance in different switch architectures in future work.

ACKNOWLEDGMENTS

This work was partly supported by the National Institute of Information and Communications Technology (NICT).

REFERENCES

[1] O. Alparslan, S. Arakawa, and M. Murata, “Packet switch architectures for very small optical RAM,” in *Proceedings of The First International Conference on Evolving Internet (INTERNET 2009)*, 2009, pp. 106–112.

[2] C. Villamizar and C. Song, “High performance TCP in ANSNET,” *Computer Communication Review*, vol. 24, no. 5, pp. 45–60, 1994.

[3] T. Aoyama, “New generation network(NWGN) beyond NGN in Japan,” Web page: http://akari-project.nict.go.jp/document/INFOCOM2007.pdf, 2007.

[4] G. Appenzeller, J. Sommers, and N. McKeown, “Sizing router buffers,” in *Proceedings of ACM SIGCOMM*, 2004, pp. 281–292.

[5] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden, “Part III: Routers with very small buffers,” *ACM SIGCOMM Computer Communication Review*, vol. 35, pp. 83–90, 2005.

[6] L. Zhang, S. Shenker, and D. D. Clark, “Observations on the dynamics of a congestion control algorithm: The effects of two-way traffic,” in *Proceeedings of ACM SIGCOMM*, 1991, pp. 133–147.

[7] G. Theagarajan, S. Ravichandran, and V. Sivaraman, “An experimental study of router buffer sizing for mixed TCP and real-time traffic,” in *Proceedings of IEE ICON*, 2006.

[8] O. Alparslan, S. Arakawa, and M. Murata, "Rate-based pacing for small buffered optical packet-switched networks," *Journal of Optical Networking*, vol. 6, no. 9, pp. 1116–1128, September 2007.

[9] D. Katabi, M. Handley, and C. Rohrs, "Congestion control for high bandwidth-delay product networks," in *Proceedings of ACM SIGCOMM*, 2002, pp. 42–49.

[10] G. Papadimitriou, C. Papazoglou, and A. Pomportsis, "Optical switching: Switch fabrics, techniques, and architectures," *Journal of Lightwave Technology*, vol. 21, no. 2, pp. 384–405, 2003.

[11] O. Alparslan, S. Arakawa, and M. Murata, "Rate-based pacing for optical packet switched networks with very small optical RAM," in *Proceedings of IEEE Broadnets*, September 2007, pp. 300–302.

[12] ——, "XCP-based transmission control mechanism for optical packet switched networks with very small optical RAM," *Photonic Network Communications*, vol. 18, no. 2, pp. 237–243, October 2009.

[13] S. Kandula, D. Katabi, B. Davie, and A. Charny, "Walking the tightrope: Responsive yet stable traffic engineering," in *Proceedings of ACM SIGCOMM*, 2005, pp. 253–264.

[14] N. McKeown and T. E. Anderson, "A quantitative comparison of iterative scheduling algorithms for input-queued switches," *Computer Networks*, vol. 30, no. 24, pp. 2309–2326, 1998.

[15] S. McCanne and S. Floyd, "ns Network Simulator," Web page: http://www.isi.edu/nsnam/ns/, July 2002.

[16] S. Low, L. Andrew, and B. Wydrowski, "Understanding XCP: equilibrium and fairness," in *INFOCOM*, 2005, pp. 1025–1036.

[17] J. C. Mogul, "Observing TCP dynamics in real networks," in *Proceedings of ACM SIGCOMM*, 1992, pp. 305–317.

[18] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the Internet's router-level topology," in *Proceeedings of ACM SIGCOMM*, 2004, pp. 3–14.

On Designing Semantic Lexicon-Based Architectures for Web Information Retrieval

Vincenzo Di Lecce, Marco Calabrese
DIASS
Politecnico di Bari – II Faculty of Engineering
Taranto, Italy
e-mail: {v.dilecce, m.calabrese}@aeflab.net

Domenico Soldo
myHermes S.r.l.
Taranto - Italy
e-mail: domenico.soldo@myhermessrl.com

***Abstract*—In this work, a novel framework for designing Web Information Retrieval systems with particular reference to semantic search engines is presented. The key idea is to add the semantic dimension to the classical Term-Document Matrix thus having a three-dimensional dataset. This enhancement allows for defining a lexico-semantic user interface where the query process is performed at the conceptual level thanks to the use of a Semantic Lexicon. WordNet Semantic Lexicon is used here as golden ontology for handling polysemy and synonymy, hence it is useful for disambiguating user queries at the semantic level. A layered multi-agent system is employed for supporting the design process. Particular emphasis is given to formal system knowledge representation, the interface layer managing user-system interaction and the markup layer performing the semantic tagging process.**

Keywords-component; information retrieval; semantic lexicon; WordNet; MAS; semantic query

I. INTRODUCTION

Since its advent, the World Wide Web (hereinafter *WWW* or simply *the Web*) has increased dramatically in size and number of interlinked resources. This trend enforces search engine developers to adopt Web document indexing techniques, which exchange scalability for fair precision/recall performances. Surveying the literature of the latest years it is easy to notice the growing consensus about the need for involving semantics in retrieval systems. Approaches that employ low-level features as indexing parameters are prone in fact to a number of pitfalls, like the inherent ambiguity of polysemous query words. At the present time, commercial search engines provide relevant responses if the user is good enough when submitting the *right* query. Therefore, the access to high-quality information on the Web may be still problematic for unskilled users.

Traditional search engines are conceptually based on a term-document look-up table (also known as Term-Document Matrix or TDM for short). Lexical terms conveyed by the user query play the role of entries; documents populate a (ranked) list of weblinks that match user query terms according to a given metric. The user is required to discern among the given options and choose the one that is supposed to be closest to his/her intentions.

A more sophisticated type of Web information retrieval systems is represented by meta-search engines, which relay user query to several search engines, collect their responses and finally propose them to the user according to certain criteria. Meta-search engines, however, have still to deal with the problem of mixing information coming from different sources, which is an awkward task to accomplish, unless some semantic approach is pursued.

Semantic search engines should attempt to understand the user query at the ontology level. They should also offer a pictorial representation of the retrieved dataset, letting the user have the impression to move within a semantic search space. In order to be really effective, they require a strong theoretical knowledge model, which has to be sufficiently robust to allow for indexing heterogeneous data scattered across the Web.

In this work, a novel framework for designing textual information retrieval systems with particular reference to semantic search engines is presented. A semantic dimension is added to the classical term-document matrix thus having a three-dimensional dataset. In this view, the user is forced to adopt a new semantic query paradigm, which is closer to human understanding than to traditional keyword-based techniques. The query process is performed at the conceptual level thanks to the use of a Semantic Lexicon considered as a golden ontology useful for the sense disambiguation task. A layered Multi-Agent System (MAS) is employed for supporting the whole design process.

This article is an extension of a previous work presented in the ICIW 2009 Conference [1] specifically focused on semantic tagging of Web resources using MAS architecture. In the present paper, the critical point of including semantics in text retrieval systems is handled under a more general and complete perspective that involves Web ontology modeling. The final aim is to bridge the gap between traditional search engines based on term-document indexing and emerging semantic requirements by means of a suitable model, which embeds terms, documents and semantics into a single knowledge representation.

The outline of the paper is as follows: Section II reports related work in Information Retrieval with particular reference to search engines and semantic tagging aspects; Section III describes WordNet architecture [2] and its usefulness for the scope of this work; Section IV proposes the new three-dimensional information retrieval framework; Section V presents the used multi-agent system architecture; Section VI comments the carried out experiments and

prototypal implementations; conclusions are sketched in Section VII.

II. Related Work

Information Retrieval (IR) is finding material (usually documents) of unstructured nature (usually text) satisfying an information request from within large collections (usually stored on computers) [3]. Automated IR systems are conceptually related to object and query. In the context of IR systems, an object is an entity, which keeps or stores information in a database, i.e. in a structured repository. User queries are then matched to objects stored in the database. A document is, therefore, an opportune collection of data objects.

Often the documents themselves are not kept or stored directly in the IR system, instead they are represented in the system by document surrogates automatically generated by the same IR system by means of a document analysis. Nowadays there are two approaches to document analysis: statistical and semantic.

The statistical approach was initially proposed by Lhun. In 1958 he wrote: "*It is here proposed that the frequency of word occurrence in an article furnishes a useful measurement of word significance. It is further proposed that the relative position within a sentence of words having given values of significance furnish a useful measurement for determining the significance of sentences. The significance factor of a sentence will therefore be based on a combination of these two measurements*" [4]. It is interesting to note that this approach is still used in many modern IR systems.

On the other hand, a Semantic Information Retrieval system exploits the notion of *semantic similarity* (based on lexical and semantic relations) between concepts to determine the relevancy of a certain document. One way of incorporating semantic knowledge into a representation is mapping document terms to ontology-based concepts. In [5], for example, a formal ontology-based model for representing Web resources is presented. Starting from semantic Web standards as well as established ontologies the authors reformulate the IR task into a data retrieval task assuming that more expressive resources and query models allow for a precise match between content and information needs. In this work instead, the term-concept mapping is provided by a golden ontology expressed in the form of a Semantic Lexicon like WordNet. The usefulness of this choice will be explained throughout the text further on.

A. Traditional IR techniques

The most widespread and popular applications of IR are Web search engines. They are designed to answer to a human query with an HTML page containing a ranked list of links to Web sites or documents. Every traditional Web search engine represents each retrieved webpage in its own search space by using a set of sentences that are considered as relevant to the user query. The relevancy of the retrieved documents is essentially dependent upon the chosen metric and the ranking strategy. As far as now, the most common document retrieval approach is searching for word-to-word correspondences (after stemming and stop-word procedures) between the set of query keywords and the set of document terms. Although the query search may be restricted by using Boolean and/or operators (thus providing a more selective filtering on the search space), the quality of document retrieval is significantly affected by the ranking strategy. A simple comparison among the principal Web search engines shows in fact how different the retrieved document could be, even in response to the same user query word.

The well-known Page Rank Algorithm [6] has been one of the keys to success for the Google Web search engine. It represents undoubtedly one of the most single important contributions to the field of IR in the latest years. The Page Rank Algorithm employs a fast convergent and effective random-walk model for ranking graph nodes like hyperlinked Web resources [7]. It is based on the bright assumptions that weblinks may be interpreted as "votes" given from the source page to the destination page. The vote expressed by a link is in fact weighted by the "reference" (Page Rank value) of the pages from where the links come, in accordance with the formula provided by the authors:

$$PR(A) = (1-d) + d * \left(\frac{PR(T_1)}{C(T_1)} + \ldots + \frac{PR(T_n)}{C(T_n)} \right) \quad (1)$$

where $PR(X)$ function gives the Page Rank value of page X, A is the webpage pointed by T_1, T_2,... T_n webpages, $C(T_i)$ is the number of links outgoing from page T_i and finally d is a properly set constant value. The previous formula is recursive. By highly ranking the most referenced pages, Page Rank represents a good prior filter to the enormous heterogeneous search space. In addition to this, the simple graphic view provided by Google home page can be easily understood by a great variety of users. In many real cases, Google apparent precision, however, can be partially ascribed to the poor syntax underpinning the user query and to the self-influence it has had on users in the way they formulate the query. Everyone can experience how much the retrieval performances decrease with more complex human-like queries.

The adoption of more sophisticated retrieval functions can help reduce the misbalance now pending on the ranking algorithms.

B. Semantic IR techniques

To overcome the limits of the traditional approaches, new semantics based techniques are being investigated in the latest years, although there is no ground-breaking technology at the moment that can be considered sufficiently mature to compete with traditional IR systems on a large scale. In the preface to the proceedings of a late international workshop on semantic search held in 2008 [8] it is explicitly stated: *"…the representation of user queries and resource content in existing search appliances is still almost exclusively achieved by simple syntax-based descriptions of the resource content and the information need such as in the predominant keyword-centric paradigm."*. A recent study [9] shows that

retrieval performances are still low for both keyword-based search engines and the semantic search engines.

Provided this, one of the most relevant semantic techniques which has had a number of useful applications in various fields spanning from information discovery to document classification is Latent Semantic Indexing (LSI). LSI implements a strictly mathematical approach based on applying Single Value Decomposition (SVD) to the TDM. SVD decomposes the TDM into the product of three matrices:

$$TDM = T_0 * S_0 * D_0' \quad (2)$$

T_0 and D_0 are the matrices of left and right singular vectors and S_0 is the diagonal one with its elements representing singular values in decreasing order. D_0' is the matrix transpose of D_0. Taking only the largest values offers a good approximation of the original TDM, thus reducing the whole search space to a relatively smaller "concept space" called LSI [10]. From this point of view, LSI is more powerful than a traditional document search algorithm: it overcomes the limits of Boolean query allowing for clustering documents semantically. LSI represents a right compromise between simplicity and good retrieval performance (measured as recall/precision values). This makes it a powerful, generic technique able to index any cohesive document collection in any language. It can be used in conjunction with a regular keyword search, or in place of one, with good results. Unfortunately, LSI suffers from scalability problems since large document sets require heavy computing on massive matrices. Furthermore, although it has been shown that LSI is able to handle correctly data structured into taxonomic hierarchies [11], it is not suited to make these taxonomies explicit in the search results. In other words, LSI is a good tool for finding semantic similarities, but it clusters output data in a *flat* (nonhierarchical) manner.

To deal with both semantics and lexical issues, a more comprehensive approach than TDM-based techniques is needed. This work grounds on the idea that, for an efficient information retrieval, lexical forms must be endowed with semantic tags in order to disambiguate their meaning. To carry out the disambiguation task, WordNet Semantic Lexicon is used.

C. Semantic Tagging

Semantic tagging (or markup) is conceived to define metadata for describing a given resource. A tag can be interpreted as a placeholder that helps user (human or computer) understands the context in which to interpret the tagged resource. An HTML tag, for example, lets browser interpret how to render a webpage; an XML tag instead allows for defining an entity name in a syntactically structured way. However, despite the initial enthusiasm around this new (meta)language [12], XML alone proved to be insufficient for most ontology-driven applications. XML in fact supplies a well-defined syntax (which is a desirable for data integration) but lacks in providing semantics. For example, XML does not resolve the lexical ambiguity that may arise when two applications share data having the same tag names, unless a Document Type Definition (DTD) or a XML Schema Definition (XSD) file is attached.

The authors are confident that any kind of semantic application cannot exist without prior defining the knowledge representation model, which is suitable and sufficient to express the given problem ontology. In the semantic engineering process, the ontology that conceptualizes (ideally in the best way) the common body knowledge is generally called *golden* or *gold standard* ontology. Its counterpart is the *individual* ontology, which strongly depends on the person who actually performs the ontology engineering process.

Generally, Web ontology modeling requires an engineering effort that can be yielded only by experts with the aid of auxiliary ontology editing tools [13][14]. In the last decade much attention has been devoted to designing layered XML-based languages such as RDF(S) [15], DAML-OIL and OWL [16], all based on formal semantics. The final attempt was to find out a good compromise among expressiveness, inferential capabilities and computability to use in the Web context.

The gap between software engineering methodologies based on the above languages and real-world ontology modeling is still a debated issue [17]. Web ontology representations have to deal with a spectrum of drawbacks spanning from language inherent ambiguity to context dependency, presence of incoherent statements, scattered pieces of information, difficulty in ontology matching and so on. It seems that all these issues have twofold reason: they lay both on the semantic (ontology) level and on the lexical (language) level. Consequently, semantic annotation of Web resources is prone to produce weak structure metadata. This is particularly true for collaborative (wiki) approaches [18] where personal conceptualizations are rather difficult to be mapped one another. Although such collaborative environments represent a challenge for the research community, they are still tailored to generic semantic services [19]. A top-down solution is to provide the tagging system with a well-defined and widely-accepted ontology: choosing the right ontology may be demanding in complex environment like the whole Web.

This work employs an agent-based architecture model for supporting the whole information retrieval process, from the user interface to the semantic tagging of Web resources. The agents that perform the annotation task use a Semantic Lexicon (hereinafter SL) as their golden ontology. In its actual implementation, the chosen SL was WordNet 3.0.

III. USING WORDNET AS GOLDEN ONTOLOGY

The golden ontology paradigm focuses on comparing how well a given ontology resembles the gold standard in the arrangement of instances into concepts and the hierarchical arrangement of the concepts themselves [20]. A copious literature exists on golden ontologies [21][22][23]. In [24] Di Lecce and Calabrese address the new emerging approach of SL-based systems for modeling semantic Web applications. Starting from a preliminary survey on the different use of the

concepts 'taxonomy' and 'ontology' in the literature, they identify SL as a good mediator between the two extremes. The authors also provide a SL-based abstract model suitable for multi agent system implementation. According to the authors' view, an indicative exemplar for the SL class is WordNet [2].

WordNet is a SL purposely engineered for text mining and information extraction. For example, it has been used to carry out Word Sense Disambiguation (WSD), for an overview of such characteristic the reader can refer to [25][26]. WordNet is referred to in the literature in several ways:

- Lexical Knowledge Base [27][28]
- Lexical Taxonomy [29][30]
- Lexical Database [31][32]
- Machine Readable Dictionary [33][34]
- Ontology [35][36]
- Semantic Lexicon [1][37]

Although, the above definitions can be considered synonyms, they emphasize different aspects of the same object. In this paper, only the latter definition will be used, since it accounts for the two elements (lexicon and semantics) which are relevant for the IR task, as explained forth. In this view, an important WordNet feature supplied by its underpinning data model is the capability of handling polysemy and synonymy. To this end, the concept of 'Sense Matrix' is introduced.

A. Defining the Sense Matrix

Two prominent causes of language ambiguity are polysemy and synonymy. Synonymy decreases recall and polysemy decreases precision, leading to poor overall retrieval performances [38]. It is interesting to note that synonymy represents a lexical relation among word forms while polysemy occurs when the same lexical form has multiple meanings. To define the relation among lexical and semantic entities at a finer grain, the definition of *Sense Matrix* is due. Thereby, a (*feasible*) *sense* is defined as particular element of such a matrix. Formally:

Def. (Sense Matrix). If *L represents the set of lexical entities and C the set of concepts of a given SL, a Sense Matrix S is defined as the matrix* $L \times C$ *such that* $S[i,j]=1$ *if* $(l_i, c_j) \in SL$ *and* $S[i,j]=0$ *otherwise. The set of feasible senses is defined as:*

$$FS := \{ s_{ij} \mid S[i,j]=1 \} \quad (3)$$

Throughout the text only feasible senses will be considered.

The concept of Sense Matrix is not new in the literature. In 2006 Swen [39] introduces almost the same notion. There is however some difference in terminology. The term 'sense' for Swen corresponds to our 'concept', thus, for Swen, a 'sense' is a term-document matrix. Our model can be considered as a specification to that of Swen assuming that senses are provided by a golden ontology.

It is noteworthy that *S* induces a binary matrix *M* on the Cartesian product *L* x *C* that is generally called 'lexical matrix' in the literature [40][41]. In [42], the lexical matrix is presented as an integral part of the human language system. Since there is no preference between the two dimensions represented by *M* (lexical and semantic), the authors prefer to refer to *M* as a Sense Matrix. This matrix can be considered as the base computational support for dictionary-based retrieval systems. Actually, it works as a look-up table that allows for switching from one dimension to another. An illustrative example of matrix *M* is provided in Table I.

TABLE I. EXAMPLE SENSE MATRIX . SENSES ARE DEFINED AS MATCHES BETWEEN LEXICAL ENTITIES (ROWS) AND CONCEPTS (COLUMNS)

SENSE MATRIX		CONCEPTS			
		c_1	c_2	c_3	c_4
LEXICON	l_1	0	0	0	1
	l_2	0	1	1	0
	l_3	1	1	0	0

B. WordNet data model

WordNet is organized around the idea of synsets, i.e. group of cognitive synonyms, each one representing a specific concept in a given context. Synsets are interlinked by means of conceptual-semantic and lexical relations. Any synset pertains to the concept layer, i.e. it is an instance of the set of concepts. The one-to-one relation between the synset and the word form produces the *sense*. Hence, a synset can be defined as the union of senses sharing the same concept entity (i.e. synonyms).

The 'sense' table combines tuples of the 'word' table with tuples of the 'synset' table. According to SL definition, the three tables define respectively the Sense Matrix, the set of lexical entities and the set of concept entities.

The WordNet taxonomic hierarchies (comprising the set of lexical and semantic relations) are covered by the two tables 'lexlinkref' and 'semlinkref'. Semlinkref defines only semantic relations, while lexlinkref defines lexico-semantic relations. In other words, lexlinkref provides recursive relations over the set of senses. An index to all kind of relations is contained in the 'linkdef' table.

WordNet is an ongoing project, since minor bugs and refinements characterize new version releases (in this work WordNet 3.0 was finally adopted). An excerpt of WordNet 3.0 class diagram is reported in Figure 1.

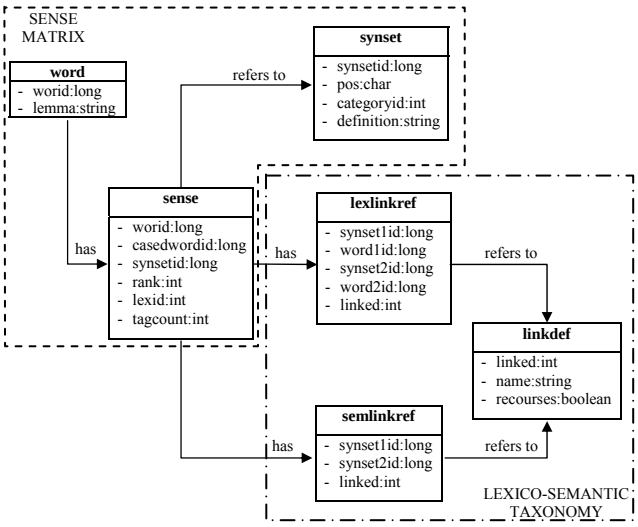


Figure 1. An extract of WordNet data model. Tables belonging to sense matrix and lexico-semantic taxonomy are grouped separately.

C. Formal Knowledge Representation

A formal representation of WordNet SL can be adapted from Dellschaft and Staab [21]. Referencing to a previous work on the subject [43], they provide a simple, but formal definition of a *Core Ontology* as the triplet:

$$CO = <C, root, \leq_C> \quad (4a)$$

where C represents the set of concepts, *root* the uppermost superordinate concept and $\leq_C$ a partial order on C (hence a taxonomical relation). Core Ontology seems to be an effective representation because it synthesizes the different layers constituting an ontology [44]. As a consequence of the introduction of the Sense Matrix, the Core Ontology definition used in this paper slightly changes. The set C is substituted by the set of feasible senses *FS*:

$$CO = <FS, root, \leq_C> \quad (4b)$$

It is evident that Core Ontology definition still has a taxonomical structure (actually a directed acyclic graph form)[1].

In WordNet, nouns, verbs, adjectives and adverbs can be considered as four lexical categories (also known as *part-of-speech* in the literature) each one defining a corresponding sub-ontology. Thus, the following partition generally[2] holds for a generic SL:

$$SL = \overline{O}_n \oplus \overline{O}_v \oplus \overline{O}_{adj} \oplus \overline{O}_{adv} \quad (5)$$

[1] This complies with some relations like hypernymy/hyponymy and may be not sufficient for others where cyclic relations may occur. For the aim of this paper however, DAG structures only are considered. More complex grap-like structures are left to future work on the subject.

[2] It can happen that some relations (especially the morphological ones, like derivational forms) make this assumption not valid. In this sense, the provided partition should be intended as an opportune simplification for a working hypothesis.

The overline is used to stress that the ontology is a golden one.

WordNet considers different semantic and lexical relation among concepts such as hyponymy/hypernym, meronym/holonym, antonyms, entailment and so on. Some relations are specific to certain categories like entailment for verbs; moreover, there are some relations having a single-rooted structure while some others are not. $\overline{O}_n$ is the only one having one single root (the synset conceptualizing the 'entity' lexical entry) hence, it suits the formal (4b) definition perfectly.

Notation. For the sake of conciseness, the following notation is introduced:

$$O_{conc}^{rel} \quad (6)$$

where *rel* represents a lexico-semantic relation, i.e. one element of the set {*hyponymy, hypermymy, holonymy, meronymy,...*}and *conc* is one of the four lexical categories, i.e. one element of the set {*nouns, verbs, adjectives, adverbs*}. O_n^{hyper} , for example, indicates an ontology defining hypernymy (relation) among nouns (concept nodes). In this paper only hypernymy has been employed in the considered golden ontology:

$$\overline{O}_n = \overline{O}_n^{hyper} \quad (7)$$

In [26] more WordNet relations are used with the aim of building *semantic graphs*. The authors adopt these structures in the Structural Semantic Interconnection (SSI) algorithm for the word sense disambiguation task. The semantic context is used in each iteration of the algorithm to disambiguate the lexical terms. Thus the accuracy of the algorithm is strictly related to the chosen context. The major difference between SSI and our approach is that, in the latter, the context is not an input for the sense disambiguation system.

IV. PROPOSED IR MODEL

The perfect search engines should respond to user query by listing exactly what the user actually queried for. Provided that this desirable situation is an ideal one, it is more feasible to reason about what current search engines *generally* do. They provide a ranked list of websites matching the user query according to a given algorithm. Upon search engines response, user chooses the website to browse, occasionally coming back to the search engine webpage to submit another query (Figure 2 reports an UML representation of the whole mechanism). Since the most of currently available search engines are not semantic-based, they index Web documents in a way similar to the Sense Matrix reported in Table I. User query is performed only at the lexical layer, being exposed to misinterpretation due to erroneous synonymy and polysemy interpretation. This means that the semantic gap is left totally to the user understanding.

In fact, in general a small text preview is fed back to the user, to let him/her decide the best option, basing on the semantics he/she gives to the displayed preview. This is an elegant way of bridging the semantic gap: the lack of this approach is that semantics is pushed in the query-response mechanism only from the user side. However, this "try and look" paradigm can be overcome in the light of the proposed IR system (Figure 3 depicts an UML representation of the mechanism characterizing the proposed IR system).

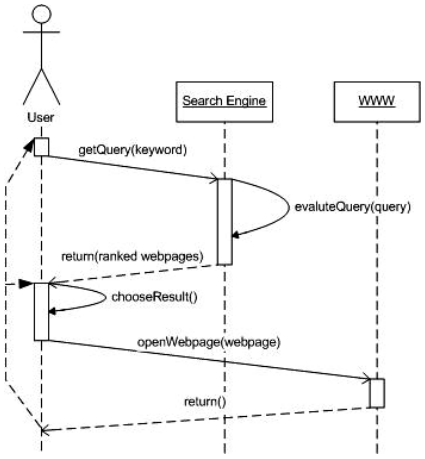


Figure 2. UML representation of the mechanism characterizing traditional search engines. The searching process starts with the user's query. Keywords are the entry points for the search engine alghoritm.

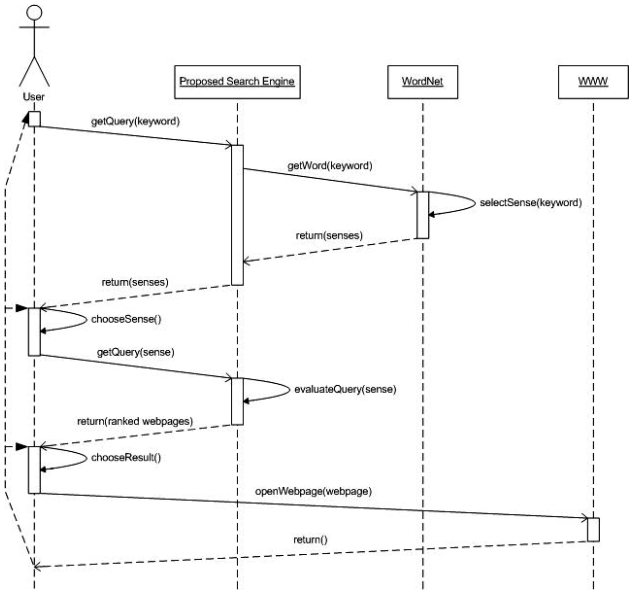


Figure 3. UML representation of mechanism characterizing the proposed semantic search engine. As in figure 2, the searching process starts with the user's query. In this case, instead, senses related to the keywords are the entry points for the proposed semantic search engine alghoritm.

In the proposed system, as in traditional keyword-based systems, the user enters one or more keywords that he/she considers as significant for the kind of document he/she is searching for. Contrarily to retrieval systems based on term-document matrix, our IR system queries the golden ontology in order to get all possible senses related to (lexical) user query. The sense is explained by means of a short gloss, which is actually a meta-description of the sense itself. Once that the user has chosen the sense he/she wants to search for, the system retrieves all documents previously indexed by that sense.

A. New Browsing Paradigm

Our approach is based on a three dimensional dataset comprising respectively term, synset and document dimensions (Figure 4).

User query begins at the lexical level and then moves towards semantics thus becoming a two-step request/reply process:

1. The first step is a traditional keyword-based query performed at the lexical level. The system replies by listing the possible related senses.
2. The second step consists in user choosing the right sense thus entering the 'semantic browsing mode' which also allows for selecting the semantically indexed documents.

In this new framework, documents are indexed by senses. This does not affect the chosen document ranking criteria (like Page Rank) since dimensions are orthogonal. The real difference from traditional IR models is that user moves within a semantic space, eventually deciding to open a sense-related webpage (as examples related to a gloss in a dictionary).

System response in the first step is not possible unless some sense disambiguation technique is applied. For a previously published sense disambiguation technique, the readers may refer to the work of Di Lecce et al. [25].

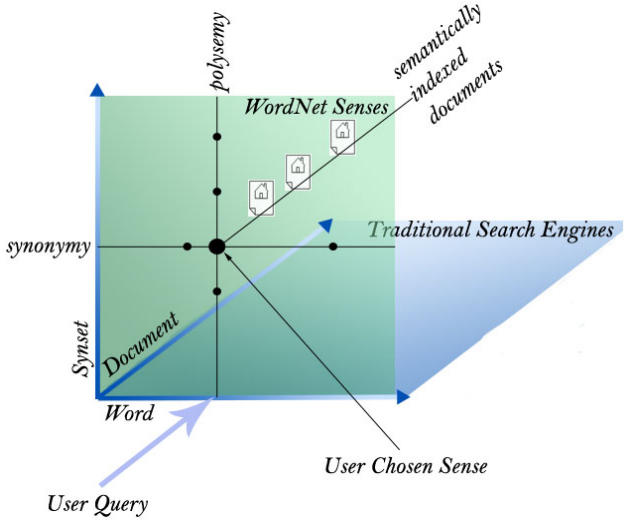


Figure 4. The proposed 3-dimensional IR model. It can be considered as the 3d space projection of the tuple *f(t, s, d)* characterized by term, synset and document.

V. SEMANTIC LEXICON-BASED MAS

MAS design has been gaining the attention of research community for many years [45]. Software agents are designed to cooperate (either with other agents or with humans) for managing the system knowledge base (KB) in different situations. In this paper a MAS implementation that employs WordNet as golden ontology is used to support the design of the proposed SL-based information retrieval system. The used MAS architecture is a hierarchical one [46] and is composed of the following layers:

- **Interface Layer**: it responds to user query. User may be human or computer such as crawlers and parsers;
- **Brokerage Layer**: it mediates among computational resources according to environment constraints;
- **Markup Layer**: it performs the tagging and other related activities.
- **Knowledge Layer**: it manages system knowledge base.

This agent-based approach is scalable because many features can be added to the SL-based system without affecting the underpinning model. For example, an inference engine may be added to the system in order to inference on new semantic relations among concept words. Tests in this direction are currently under way. Their aim is to assess the feasibility of domain-specific search engines that would enhance domain browsing and document retrieval.

The used MAS architecture has been inspired by previous works in other fields (see for example [46]). An overview of the proposed MAS architecture is depicted in Figure 5.

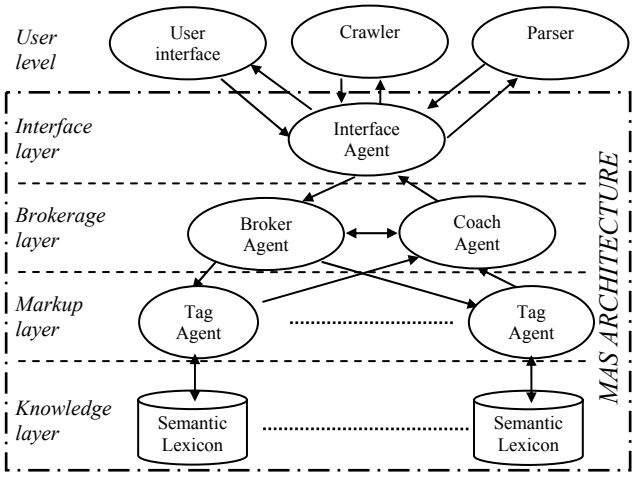


Figure 5. Multilayered MAS architecture used for semantically tagging Web resources. It is noteworthy the use of SL as golden ontology for the system knowledge.

Hereinafter an insight into the proposed MAS architecture is provided. Each different layer is described, with particular emphasis to the interface layer which handles user-system interaction, on the markup layer, which provides the Web resources semantic tagging process, and on the knowledge layer formally before presented.

A. User Level

It is the top level of the hierarchy and interacts with the MAS. It represents the communicational channel from and towards the system's environment. The user level is suitable composed of the three following elements: *user interface* – i.e. the human-machine interface; *spider* – a computer program that analyses the taxonomy structure of considered websites; *parser* – a computer program that extracts the relevant information from Web documents. The crawling/parsing processes are thoroughly described in [47]. Instead, a prototypal version of the user interface has been developed and presented in this paper (Figure 8 represents its actual implementation).

B. Interface Layer

The semantic browsing options handled by the interface agent are synthesized by the following Extended-BNF representation.

```
<interface> ::= <frame_header> <frame_www>
                <frame_semantics>
<frame_header> ::= {<sense>}
<frame_www> ::= {<href>}
<frame_semantics> ::=  [<forward_sense>]
                       [<backward_sense>]
<forward_sense> ::= {<sense>}
<backward_sense> ::= {<sense>}
<sense> ::= <word> <gloss>
```

The interface is composed of three frames:

1. Header: reporting the considered sense i.e. word-synset pair. The sense is described by means of the gloss associated to it;
2. WWW (traditional browsing): lists all Web resources indexed by the current sense;
3. Semantics (semantic browsing): allowing the user to move within the semantic space;

The choice of the extended version of the Backus Normal Form is due to the need to easily represent the cardinality for both elements sense and href. While curly brackets indicate the cardinality of a symbol, the square brackets represent the optional element in the derivation rule. *Forward_sense* and *backward_sense* are the parts of semantic space linked to the header sense. A graphical illustration explaining the BNF is presented in the next Section.

It is noteworthy that this interface shows recursive characteristics. The user can perform semantic browsing moving towards similar concepts in the query refinement process. In the experiment Section a screenshot of the prototypal implementation is commented in more detail.

C. Brokerage Layer

The MAS is triggered by user query submitted to the interface agent. Once the query has been correctly decoded, the interface agent leaves control to the Brokerage Layer. This layer is managed by two agents: broker and coach.

Broker Agent analyses which Tag Agents can satisfy the requirement. It manages all inbound communication coming from the Interface Agent. Starting by one query, it relays user service request to available resources of the lower layer, according to the chosen scheduling policy.

Coach Agent receives message from all the Tag Agents, collects and ranks the results. Next, it sends a message to the Broker Agent to inform that service request has been fulfilled.

Both agents of this layer are poorly detailed because that goes beyond the scope of this paper.

D. (Semantic) Markup Layer

In [25], a WSD algorithm was proposed to find the nearest common WordNet subsumer among words extracted from two link texts (also known as anchortexts). These couples of textual descriptions are taken from any possible pair of inbound and outbound links of a given webpage. If a concept subsumer (which is a synset) is found, lemmas, which lexicalize it, are used to tag the webpage.

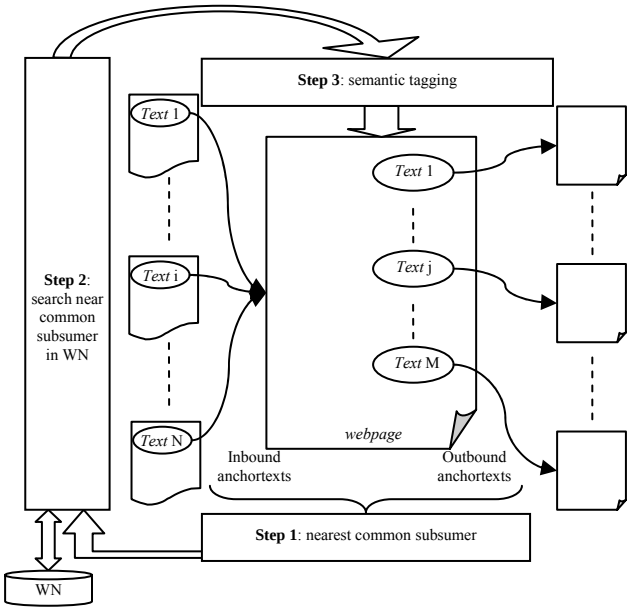


Figure 6. Actions performed by a Tag Agent. For any couple of inbound and outbound text links (step 1) the nearest common subsumer is searched in the WordNet database (step 2) according to the semantic relation pertaining the agent (e.g. hypernymy). If such synset element is found, its related lemmas are used to tag the corresponding webpage (step 3).

A *tag agent* (Figure 6 helps explaining how Tag Agent actually works) repeatedly performs this tagging activity on the list of webpages received by the *Broker Agent*.

For modularity purposes, each Tag Agent searches for semantic tags exploiting one of the possible semantic relations provided by the chosen SL. There can be one or more Tag Agents for hypernymy, others for holonymy and so on.

E. Knowledge Layer

Starting from the notations given in Section III.C, the system KB can be then formally expressed as follows:

$$KB = \overline{O}_n \bigcup_i \left\{ O_n^{href} \right\}_i \quad (8)$$

$\left\{ O_n^{href} \right\}_i$ represents the individual ontology of the *i*-th inspected webpage; *n* represents the nouns (webpage semantic tags) identified by the tagging agent, finally *href* represents the HTML hyperlinks connecting tagged webpages. As shown in Figure 5 the Knowledge Layer can be split in many Semantic Lexicon units as much as Tag Agents exist in the upper layer. This ensures more flexibility and scalability of the system too.

VI. EXPERIMENTS

Our experiments have been carried out on a data set of 48 distinct websites clustered in four semantic domains. Table II reports the number of analyzed websites for each semantic domain.

TABLE II. NUMBER OF INSPECTED WEBSITES FOR THE CHOSEN SEMANTIC DOMAINS

Semantic Domain	# inspected websites
University	17
Low-cost airline	10
Seaport	8
Airport	13

The crawling and the parsing phases have been limited to the analysis of the crawled first one hundred webpages for each website. This choice ensured the coverage of the main taxonomical structures of the inspected websites (general categories).

A. Prototypal Interface Implementation

According to the previously presented E-BNF representation a prototypal Web-based user interface has been implemented. Apache, PHP, MySQL and Ajax technologies have been used for this scope. Figure 8 shows an example screenshot of the user interface developed for the proposed semantic search engine. The interface can be divided into the following three frames:

1. *header frame*: it supports the user in the sense disambiguation process as specified in the UML of Figure 3. Thanks to this frame, the user selects the right sense in the synset-word (WordNet) plain depicted in Figure 4.

2. *www frame*: this area lists the hyperlinks indexed by the sense that results from the user query. By clicking one of the listed hyperlinks the user is redirected to the corresponding webpage as in traditional search engines (UML of Figure 2).
3. *semantics frame*: it allows the user for browsing the WordNet plain. The user is supposed to be in the sense chosen in header frame and can move forward or backward towards "neighbour" senses. Given two senses *a* and *b* they are considered here as neighbours if they exhibit a common subsumer. When the user selects a new sense the interface is reset (i.e. the user "is moved" to the new sense), thus showing a recursive behaviour.

B. Semantic tagging experiments

The semantic tagging process has been applied to the experiment set according to what was explained in the previous section. The considered semantic relations were hypernymy extracted from WordNet 3.0 release. The inspection depth in the WordNet taxonomy for finding the nearest common subsumer was thresholded to 5. This choice was affected by these reasons:

- Higher depth level in taxonomy accounts for very general concepts that are conceptually distant from the analyzed domain
- Computational effort may increase more than proportionally as depth level increases.

To evaluate the proposed architecture the presented results refer to the hypernym (*IS-A*) relation. Two different evaluations have been carried out during the test process. One of these is related to the evaluation of human agreement with the automatic markup system (qualitative test). The other one evaluates the amount of results given by the proposed system regarding the completeness of information (quantitative test).

Quantitative test. Figure 7 depicts the coverage index (in percentage) of the semantic markup grouped by the semantic domains defined in Table II. This value is useful to understand how much of the WordNet taxonomical structure is retrieved in the link-based architecture of a website. Coverage index (*ci*) has been computed for any webpage according to the simple formula:

$$ci\% = \frac{\#t}{\#w} \qquad (9)$$

where the numerator stands for the number of tagged webpages and the denominator equals the total number of inspected webpages for the website. Then, data have been grouped by domain. A box plot representation is adopted to have a synoptic view of mean, variance, minimum and maximum values for each domain. Moreover, it also represents outliers for the data set. It is noteworthy that nearly 25% of webpages are tagged by IS_A relations. As shown in Figure 7 the semantic domains of University, Seaport and Airport have a *ci* near to 28%, while semantic domain of low-cost airline has a lower *ci*. This is because inspected low-cost airline websites provide a flat cross-domain semantic structure (many heterogeneous services like car rental, hotel booking, tours, etc.).

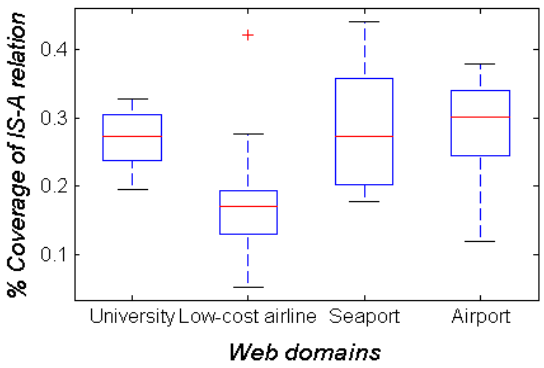


Figure 7. Coverage of IS-A relation for the considered semantic domains. The coverage values are considered in percentage. Any represented element is characterized by a continuous line (mean value) in a rectangle (variance range) and by two broken line ending with horizontal lines (minimum and maximum coverage for the data set).

Qualitative test. Table III reports an extract from the semantic mark-up process on the Manchester University website. Yet considering IS-A relation, the table is characterized by the uniform resource location (URL) of webpage, the lemmas associated to the sense markup and the anchortexts that caused the webpage to be tagged. The results are grouped by the URL identifier, in order to underline all sense markups assigned to a Web document. Table IV reports the semantic neighbors in WordNet semantic plain for data listed in Table III.

C. Evaluation of the proposal

The current semantic search panorama is quite a fragmented one. Although a lot of proposals can be found especially in recent years [49], they are still tailored to solving engineering aspects rather than being focused on performance. This conclusion can be fairly drawn by observing that even in the preface of the recent SemSearch 2008 International Workshop on semantic search [50] one of the major questions pointed out is: how can semantic search systems be evaluated and compared with standard IR systems?

Generally, small-sized comparison among the two approaches result in traditional IR systems largely outperforming semantic counterparts, at least for the recall performance. In [51] an attempt is made to fuse the two approaches by preserving the moderate recall of traditional system with the improved precision of the semantic-based ones. However, overall performances are still quite low and evaluation is confined to restricted datasets. In fact, one of the major pitfalls of corpus-based evaluation is the cost associated to the annotation. While in natural language

processing (NLP) and word sense disambiguation (WSD) such datasets already exist and allow for good performances [52], in the general framework of Semantic Web it is currently impossible to think to a wide-covering semantic annotation for Web resources, at least until new standards like OWL will be sufficiently spread. In the meanwhile (which the authors think will last for many years ahead), the solution should be based on using available information at the maximum possible extent but from a different perspective, possibly by using new user-system interaction paradigms.

With reference to this paper, the focus was on the architecture that may leverage the simple mechanism of knowledge extraction and semantic annotation from the linked structure of the Web. This is an easy to run process which contributes to building a skeleton of semantic structures to which append (index) the crawled web pages. This allows the user to exploit a different navigation paradigm based on surfing a semantic graph rather than a web graph, thus reducing the semantic gap between the user and the retrieval system. Such an enhancement shifts the problem of retrieval to sense tagging and semantic disambiguation. More detailed numerical assessments of the proposed semantic tagging technique along with WSD aspects can be found in [47].

VII. CONCLUSION

In a recent survey on existing semantic search technologies [48], the authors categorize the 35 reviewed systems under three main facets: query, system, result. They conclude that a next step for the semantic search community is to foster the use of semantics in each of the three places. They also point out three main hinders to the evolution of semantic approaches: (1) lack of evaluation of semantic search algorithms, (2) lack of user evaluation of user interfaces, (3) lack of API and middleware support. The approach proposed in this work attempts to provide a way out to all these points by proposing a holistic framework centered on the idea of the SL-based architecture for Web IR. The main idea is to enlarge traditional TDM indexing structure up to a third dimension by adding a semantic layer. In this new model the user experiences a novel query paradigm which requires two consecutive steps: first to identify the sense related to documents he/she is querying for and then to access the semantically indexed document. In the line of a previous work specifically focused on semantic tagging of Web resources, this article proposes a MAS approach to Web IR design. Particular emphasis has been given to the interface layer managing user-system interaction and the markup layer performing the semantic tagging process. Since the proposed approach is highly modular, enlarging the experiment set will be the subject of our prospective research on this matter. The user interface will be also enriched and optimized in order to be effective for an extended number of inspected websites.

ACKNOWLEDGMENT

The authors thank Prof. Roberto Peluso for his useful remarks that helped us to better define several concepts during the writing of the paper. The authors would also like to thank the anonymous referees which allowed for improving, by their useful comments, the overall quality of the paper.

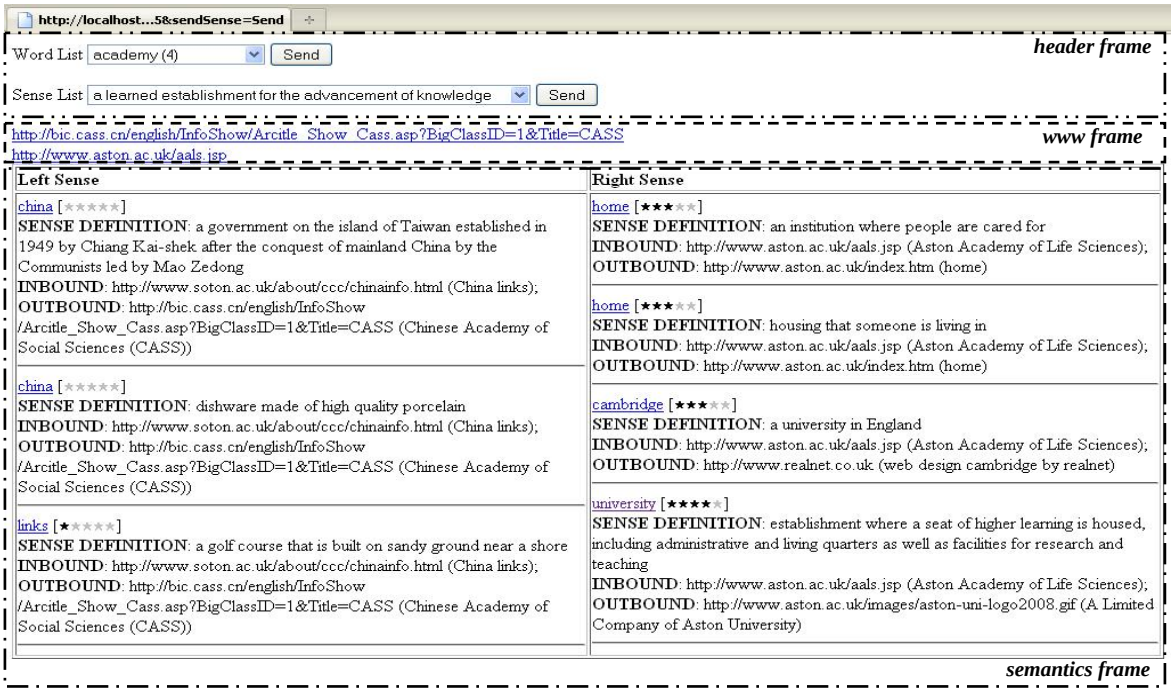


Figure 8. Screenshot of the prototypal interface. The three frames described in the text are confined in separate blocks. In the "semantics frame" each sense is quoted with a semantic relevance degree (star icons). Sense definition, along with original inbound and outbound anchortexts are also provided.

TABLE III. EXCERPT FROM THE SEMANTIC TAGGING PROCESS APPLIED TO MANCHESTER UNIVERSITY WEBPAGES ON NOVEMBER 2008 [1]. THE FIRST COLUMN REFERS TO THE URL OF THE TAGGED RESOURCES. THE NEXT COLUMN SHOWS THE FOUND SEMANTIC TAGS. THE 3RD AND 4TH COLUMNS REPORT THE INBOUND AND OUTBOUND ANCHORTEXTS RESPECTIVELY. IN PARTICULAR THE LEXICAL ENTRIES (WORDS) THAT PRODUCED THE SEMANTIC TAG ARE CAPITALIZED AND BOLDED. IT IS NOTEWORTHY THAT THE SAME WEBPAGE MAY BE REFERRED TO BY MORE THAN A COUPLE OF ANCHORTEXTS; HENCE IT MAY BE ANNOTATED BY MORE SENSE TAGS.

Url	WordNet Sense Tag {# synset_id}	Anchortext 1	Anchortext 2
http://www.eps.manchester.ac.uk	ability, power {105616246}	computer **SCIENCE**	**FACULTY** of engineering and physical sciences
	bailiwick, discipline, field, field of study, study, subject, subject area, subject field {105996646}	computer **SCIENCE**	faculty of **ENGINEERING** and physical sciences
	body {107965085}	physics and astronomy **SCHOOL** of	**FACULTY** of engineering and physical sciences
http://www.langcent.manchester.ac.uk	body {107965085}	languages linguistics and cultures **SCHOOL** of	**UNIVERSITY** language centre
	construction, structure {104341686}		
	educational institution {108276342}		
	building, edifice {102913152}	languages linguistics and cultures **SCHOOL** of	university language **CENTRE**
	cognitive content, content, mental object {105809192}	find an academic department or **DISCIPLINE**	language **CENTRE**
http://www.manchester.ac.uk/aboutus/jobs/research	work {100575741}	**JOB** opportunities	**RESEARCH** jobs
http://www.manchester.ac.uk/aboutus/manchester/sport	activity {100407535}	**ART** and museums in manchester	**SPORT**
	activity {100407535}	nightlife and **ENTERTAINMENT**	**SPORT**
	diversion, recreation {100426928}		
	activity {100407535}	**NIGHTLIFE** and entertainment	**SPORT**
	diversion, recreation {100426928}		
http://www.manchester.ac.uk/aboutus/structure	artefact, artifact {100021939}	**ART** and museums in manchester	university **STRUCTURE**
	body {107965085}	**GOVERNANCE**	**UNIVERSITY** structure
	construction, structure {104341686}	**UNIVERSITY** structure	university **STRUCTURE**
	construction, structure {104341686}	university **STRUCTURE**	**UNIVERSITY** structure
	construction, structure {104341686}	**SUPPORT** services	**UNIVERSITY** structure
	construction, structure {104341686}	**SUPPORT** services	university **STRUCTURE**
	construction, structure {104341686}	chancellors of the **UNIVERSITY**	university **STRUCTURE**

TABLE IV. SUBSUMPTION HIERARCHY FOR NEIGHBOUR SENSES EXTRACTED FROM TABLE III. SENSE NEIGHBOURS ARE REPORTED IN THE LEFT COLUMN, WHILE THE RIGHT COLUMN ACCOUNTS FOR FIRST OR SECOND LEVEL COMMON SUBSUMER. SOME SENSE NEIGHBOURS SHARE THE SAME SYNSET SUBSUMER BOTH AT FIRST LEVEL AND SECOND LEVEL SYNSET DISTANCE.

Neighbour WordNet Sense	**WordNet Synset Common Subsumer**	
	First Level Distance	Second Level Distance
'Science', {105636887}	{105616246}	
'Faculty', {105650329}		
'Science', {105999797}	{105996646}	
'Engineering', {106125041}		
'School', {108275185}	{107965085}	
'Faculty', {108287586}		
'School', {108275185}	{107965085}	
'University', {108286163}		
'School', {102913152}	{104146050}	{104341686}
'University', {103297735}	{104511002}	
'School', {108277393}	{108276342}	
'University', {108286569}		
'School', {104146050}	{102913152}	
'Centre', {102993546}		
'Discipline', {105999266}	{105996646}	{105809192}
'Centre', {105921123}	{105809192}	
'Job', {100576717}	{100575741}	
'Research', {100633864}	{100636921}	{100575741}
'Art', {100908492}	{100933420}	{100407535}
'Sport', {100582388}	{100433216}	
'Entertainment', {100426928}	{100429048}	{100407535}
'Sport', {100582388}	{100433216}	
'Entertainment', {100429048}	{100426928}	
'Sport', {100523513}		
'Nightlife', {100426928}	{100582388}	{100407535}
'Sport', {100431292}	{100433216}	
'Nightlife', {100431292}	{100426928}	
'Sport', {100523513}		
'Art', {103129123}	{102743547}	{100021939}
'Structure', {104341686}	{100021939}	
'Governance', {108164585}	{107965085}	
'University', {108286163}		
'University', {103297735}	{104511002}	{104341686}
'Structure', {104341686}	{104341686}	
'Support', {104361095}	{104360501}	{104341686}
'University', {103297735}	{104511002}	
'Support', {104361095}	{104360501}	{104341686}
'Structure', {104341686}	{104341686}	

REFERENCES

[1] V. Di Lecce, M. Calabrese, and D. Soldo, "Semantic Lexicon-Based Multi-Agent System for Web Resources Markup", In Proceedings of the Fourth International Conference on Internet and Web Applications and Services (ICIW 2009), May 2009, Mestre, Italy (ISBN: 978-0-7695-3613-2), pp. 143-148.

[2] C. Fellbaum, WordNet: An electronic lexical database, MIT Press, Cambridge, (1998).

[3] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, (2008).

[4] H. P. Luhn, "The automatic creation of literature abstracts", IBM Journal of Research and Development, Vol. 2, No. 2, pp. 159-165, (1958).

[5] D. T. Tran, S. Bloehdorn, P. Cimiano, and P. Haase, "Expressive Resource Descriptions for Ontology-Based Information Retrieval", In Proceedings of the 1st International Conference on the Theory of Information Retrieval (ICTIR'07), October 2007, Budapest, Hungary, pp. 55-68.

[6] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", In Proceedings of the Seventh Internation World-Wide Web Conference (WWW 1998), April 1998, Brisbane, Australia, Computer Networks and ISDN Systems, Vol. 30, No. 1-7, pp. 107-117.

[7] A. Esuli and F. Sebastiani, "Page Ranking WordNet Synsets: An application to Opinion Mining", In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 2007, Prague, Czech Republic, pp. 424-431.

[8] S. Bloehdorn, M. Grobelnik, P. Mika, and D. T. Tran (editors), Preface to the Proceedings of the International Workshop on Semantic Search, located at the 5th European Semamntic Web Conference (ESWC 2008), June 2008, Tenerife, Spain.

[9] D. Tumer, M. A. Shah, and Y. Bitirim, "An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia", In Proceedings of the Fourth International Conference on Internet Monitoring and Protection (ICIMP 2009), May 2009, Mestre, Italy (ISBN: 978-1-4244-3839-6) , pp. 51-55.

[10] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis", Journal of the American Society for Information Science, Vol. 41, No. 6, pp. 391-407, (1990).

[11] A. Graesser, A. Karnavat, V. Pomeroy, and K. Wiemer-Hasting, "Latent Semantic Analysis Captures Causal, Goal-oriented, and Taxonomic Structures", In Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society (CogSci 2000), August 2000, Philadelphia, PA, USA, pp. 184–189.

[12] R. Khare and A. Rifkin, "XML: a door to automated Web applications", in Internet Computing, IEEE, Vol. 1, Issue 4, July/August 1997, pp. 78-87.

[13] M. Vargas-Vera, et al., "MnM: Ontology-driven tool for semantic markup", In Handschuh, Mr Siegfried and Collier, Mr Niegel and Dieng, Miss Rose and Staab, Dr Steffen, Eds., Proceedings of the Workshop on Semantic Authoring, Annotation & Knowledge Markup (SAAKM 2002), Lyon, France, July 2002, pp. 43-47.

[14] K. Siorpaes and M. Hepp, "MyOntology: the marriage of ontology engineering and collective intelligence", Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), at the 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria, June 2007, pp. 127-138.

[15] S. Luke, L. Specter, and D. Rager, "Ontology-based knowledge discovery on the World Wide Web", In A. Franz & H. Kitano (Eds.), Working Notes of the Workshop on Internet-Based Information Systems at the 13th National Conference on Artificial Intelligence (AAAI96), AAAI Press, Portland, Oregon, August 1996, pp. 96-102.

[16] G. Antoniou and F. van Harmelen, "Web Ontology Language: OWL", in S. Staab and R. Studer, Handbook on Ontologies in Information Systems, Springer-Verlag, pp. 76-92, (2003).

[17] M. Diouf, K. Musumbu, and S. Maabout, "Methodological aspects of semantics enrichment in model driven", In Proceedings of the Third International Conference on Internet and Web Applications and Services, 2008 (ICIW '08), Athens, Greece, June 2008, pp. 205-210.

[18] R. Abbasi, S. Staab, and P. Cimiano, "Organizing resources on tagging systems using T-ORG", In Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), at the 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria, June 2007, pp 97-110.

[19] C. Lange, "Towards scientific collaboration in a semantic wiki", In Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), at the 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria, June 2007, pp. 119-126.

[20] J. Brank, D. Mladenic, and M. Grobelnik, "Gold standard based ontology evaluation using instance assignment", In Proceedings of the 4th Workshop on Evaluating Ontologies for the Web (EON2006), May 2006, Edinburg, Scotland.

[21] K. Dellschaft and S. Staab, "On how to perform a gold standard based evaluation of ontology learning", Proceedings of the 5th International Semantic Web Conference, (ISWC 2006) , November 2006, Athens, GA, USA, pp. 173-190.

[22] S. Farrar and D. T. Langendoen, "A linguistic ontology for the semantic Web", GLOT International Vol. 7, No. 3, March 2003, pp. 97-100.

[23] E. Zavitsanos, G. Paliouras, and G. A. Vouros, "A Distributional Approach to Evaluating Ontology Learning Methods Using a Gold Standard", 3rd Workshop on Ontology Learning and Population (OLP3), at the 18th European Conference on Artificial Intelligence (ECAI 2008), July, 2008, Patras, Greece.

[24] V. Di Lecce and M. Calabrese, "Taxonomies and ontologies in Web semantic applications: the new emerging semantic lexicon-based model", Proceedings of the IEEE International Conference on Intelligent Agents, Web Technologies and Internet Commerce (IAWTIC'08), December 2008, Vienna, Austria, (ISBN: 978-0-7695-3514-2), pp. 277-283.

[25] V. Di Lecce, M. Calabrese, and D. Soldo, "A Semantic Lexicon-based Approach for Sense Disambiguation and Its WWW Application", International Conference on Intelligent Computing (ICIC 2009), September 2009, Ulsan, Korea, pp. 468-477.

[26] R. Navigli and P. Velardi, "Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation", In IEEE Transactions on Patter Analysis and Machine Intelligence, Vol. 27, No. 7, July 2005, pp. 1075-1086.

[27] R. Basili, et al., "Knowledge-Based Multilingual Document Analysis", Proceedings of the International Conference On Computational Linguistics (COLING 2002) on SEMANET: building and using semantic networks - Volume 11, August 2002, Taipei, Taiwan, pp. 1-7.

[28] D. Inkpen, "Building A Lexical Knowledge-Base of Near-Synonym Differences", In Proceedings of the Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations (NAACL 2001), June 2001, Pittsburgh, PA, USA, pp. 47-52.

[29] J.J. Jiang and D.W. Conrath, "Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy", In Proceedings of the International Conference on Research in Computational Linguistics (ROCLING X), September 1997, Taipei, Taiwan, pp. 19-33.

[30] A. Gangemi, N. Guarino, A. Oltramari, and R. Oltramari, "Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level" In Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS-2001), October 2001, Ogunquit, Maine, USA, pp. 285-296.

[31] G. Miller, "WordNet: a lexical database for English.", Communications of the ACM, Volume 38, Issue 11 , pp.39-41 (1995).

[32] N. Ordan and S. Wintner, "Representing Natural Gender in Multilingual Databases", International Journal of Lexicography, Vol. 18, No. 3, pp. 357-370 (2005).

[33] J. Kegl, "Machine-readable dictionaries and education." Walker, Donald E., Antonio Zampolli and Nicoletta Calzolari, eds., Automating the Lexicon: Research and Practice in a Multilingual Environment, Oxford University Press, New York, NY, USA, pp. 249 – 284 (1995).

[34] Y. Hayashi and T. Ishida, "A Dictionary Model for Unifying Machine Readable Dictionaries and Computational Concept Lexicons", In Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006), May 2006, Genoa, Italy, pp.1-6.

[35] V. Snasel, P. Moravec, and J. Pokorny, "WordNet Ontology Based Model for Web Retrieval", Proc. Of International Workshop on Challenges in Web Information Retrieval and Integration, (WIRI '05), April 2005, Tokyo, Japan, pp. 220-225.

[36] E. Nichols, F. Bond, and D. Flickinger, "Robust ontology acquisition from machine-readable dictionaries", In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2005), August 2005, Edinburgh, Scotland, pp. 1111–1116.

[37] T. Qian, B. Van Durme, and L. Schubert, "Building a Semantic Lexicon of English Nouns via Bootstrapping", In Proceedings of the NAACL HLT Student Research Workshop and Doctoral Consortium, June 2009, Boulder, CO, USA, pp. 37–42.

[38] G. L. Kowalski and M. T. Maybury, "Information Storage and Retrieval Systems. Theory and Implementation", Springer, The Information Retrieval Series , Vol. 8, 2nd ed., (2000).

[39] B. Swen, "Sense Matrix Model and Discrete Cosine Transform", In Proceedings of the first Asia Information Retrieval Symposium (AIRS 2004), October 2004, Beijing, China, LNCS 3411, Springer-Verlag, Berlin, Heidelberg, pp. 202-214.

[40] N. Ruimy, P. Bouillon, and B. Cartoni, "Inferring a Semantically Annotated Generative French Lexicon from an Italian Lexical Resource", in Bouillon and Kanzaki (eds), Proceedings of the Third International Workshop on Generative Approaches to the Lexicon, May 2005, Geneva, Switzerland, pp. 27-35.

[41] B. Magnini, C. Strapparava, F. Ciravegna, and E. Pianta, A Project for the Construction of an Italian Lexical Knowledge Base in the Framework of WordNet, IRST Technical Report #9406-15, (1994).

[42] N. L. Komarovaa and M. A. Nowak, "The Evolutionary Dynamics of the Lexical Matrix", Bulletin of Mathematical Biology, Vol. 63, No. 3, May 2001, pp. 451-485, Springer.

[43] A. Maedche and S. Staab, "Measuring similarity between ontologies", In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web (EKAW '02), October 2002, Siguenza, Spain, pp. 251-263.

[44] J. Brank, M. Grobelnik, and D. Mladenić, "A survey of ontology evaluation techniques", In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005), at 7th International Multi-conference on Information Society (IS'05), October 2005, Ljubljana, Slovenia, pp. 166-169.

[45] M. Wooldridge and N. R. Jennings, "Agent theories, architectures, and languages: a survey", Intelligent Agents, Series Lecture Notes in Computer Science, Subseries Lecture Notes in Artificial Intelligence, Vol. 890, No. 8, Springer-Verlag, 1995, pp. 1-39.

[46] A. Amato, V. Di Lecce, C. Pasquale, and V. Piuri, "Web agents in an environmental monitoring system", Proceedings of the International Symposium on Computational Intelligence for Measurement Systems and Applications (CIMSA 2005), July 2005, Taormina, Italy, pp. 262-265.

[47] V. Di Lecce, M. Calabrese, and D. Soldo, "Fingerprinting Lexical Contexts over the Web", Journal of Universal Computer Science, vol. 15, no. 4 (2009), pp. 805-825.

[48] M. Hildebrand, J. van Ossenbruggen, and L. Hardman, "An analysis of search-based user interaction on the semantic Web", Technical report. Information Systems. Centrum voor Wiskunde en Informatica (NL) (2007).

[49] M. Hildebrand, J. van Ossenbruggen and L. Hardman. "An Analysis of Search-based User Interaction on the Semantic Web". Hildebrand, REPORT INS-E0706 MAY 2007. Centrum voor Wiskundeen Informatica Information Systems.

[50] S. Bloehdorn, M. Grobelnik, P. Mika, T. T. Duc, Preface of SemSearch 2008, CEUR Workshop Proceedings, online at CEUR-WS.org/Vol-334/

[51] F. Giunchiglia, U. Kharkevich, I. Zaihrayeu , "Concept Search: Semantics Enabled Syntactic Search", SemSearch 2008, CEUR Workshop Proceedings, Vol-334, pp.109-123.

[52] R. Navigli, "Word Sense Disambiguation: a Survey". ACM Computing Surveys, 41(2), ACM Press, 2009, pp. 1-69.

www.iariajournals.org

International Journal On Advances in Intelligent Systems
ICAS, ACHI, ICCGI, UBICOMM, ADVCOMP, CENTRIC, GEOProcessing, SEMAPRO, BIOSYSCOM, BIOINFO, BIOTECHNO, FUTURE COMPUTING, SERVICE COMPUTATION, COGNITIVE, ADAPTIVE, CONTENT, PATTERNS, CLOUD COMPUTING, COMPUTATION TOOLS
issn: 1942-2679

International Journal On Advances in Internet Technology
ICDS, ICIW, CTRQ, UBICOMM, ICSNC, AFIN, INTERNET, AP2PS, EMERGING
issn: 1942-2652

International Journal On Advances in Life Sciences
eTELEMED, eKNOW, eL&mL, BIODIV, BIOENVIRONMENT, BIOGREEN, BIOSYSCOM, BIOINFO, BIOTECHNO
issn: 1942-2660

International Journal On Advances in Networks and Services
ICN, ICNS, ICIW, ICWMC, SENSORCOMM, MESH, CENTRIC, MMEDIA, SERVICE COMPUTATION
issn: 1942-2644

International Journal On Advances in Security
ICQNM, SECURWARE, MESH, DEPEND, INTERNET, CYBERLAWS
issn: 1942-2636

International Journal On Advances in Software
ICSEA, ICCGI, ADVCOMP, GEOProcessing, DBKDA, INTENSIVE, VALID, SIMUL, FUTURE COMPUTING, SERVICE COMPUTATION, COGNITIVE, ADAPTIVE, CONTENT, PATTERNS, CLOUD COMPUTING, COMPUTATION TOOLS
issn: 1942-2628

International Journal On Advances in Systems and Measurements
ICQNM, ICONS, ICIMP, SENSORCOMM, CENICS, VALID, SIMUL
issn: 1942-261x

International Journal On Advances in Telecommunications
AICT, ICDT, ICWMC, ICSNC, CTRQ, SPACOMM, MMEDIA
issn: 1942-2601