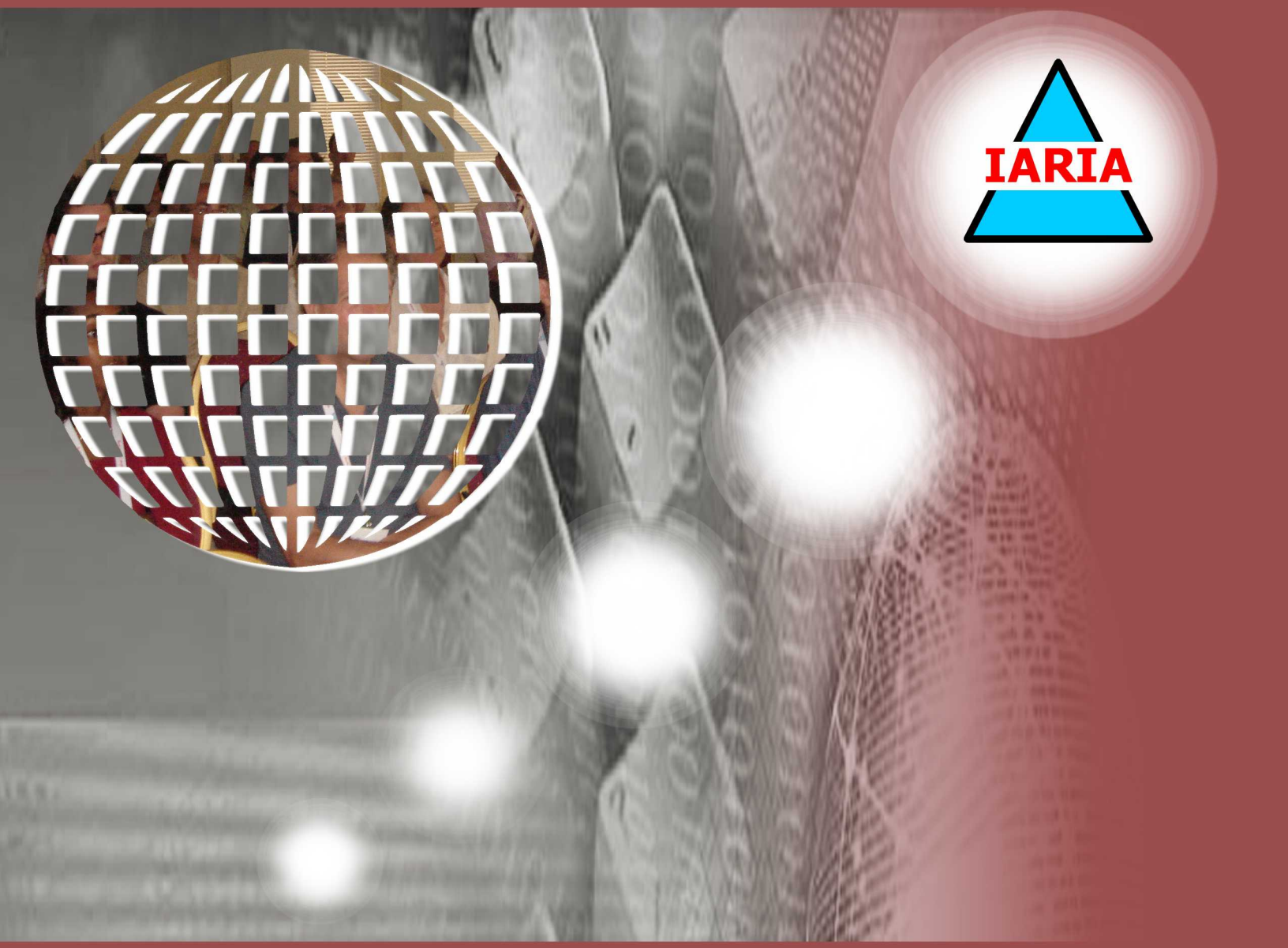
International Journal on Advances in Telecommunications



2018 vol. 11 nr. 1&2

The *International Journal on Advances in Telecommunications* is published by IARIA.
ISSN: 1942-2601
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Telecommunications
Volume 11, Number 1 & 2, 2018

Ezendu Ariwa, London Metropolitan University, UK
Miguel Arjona Ramirez, São Paulo University, Brasil
Radu Arsinte, Technical University of Cluj-Napoca, Romania
Tulin Atmaca, Institut Mines-Telecom/ Telecom SudParis, France
Mario Ezequiel Augusto, Santa Catarina State University, Brazil
Marco Aurelio Spohn, Federal University of Fronteira Sul (UFFS), Brazil
Philip L. Balcaen, University of British Columbia Okanagan - Kelowna, Canada
Marco Baldi, Università Politecnica delle Marche, Italy
Ilija Basicevic, University of Novi Sad, Serbia
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Mark Bentum, University of Twente, The Netherlands
David Bernstein, Huawei Technologies, Ltd., USA
Eugen Borcoci, University "Politehnica"of Bucharest (UPB), Romania
Fernando Boronat Seguí, Universidad Politecnica de Valencia, Spain
Christos Bouras, University of Patras, Greece
Martin Brandl, Danube University Krems, Austria
Julien Broisin, IRIT, France
Dumitru Burdescu, University of Craiova, Romania
Andi Buzo, University "Politehnica" of Bucharest (UPB), Romania
Shkelzen Cakaj, Telecom of Kosovo / Prishtina University, Kosovo
Enzo Alberto Candreva, DEIS-University of Bologna, Italy
Rodrigo Capobianco Guido, São Paulo State University, Brazil
Hakima Chaouchi, Telecom SudParis, France
Silviu Ciochina, Universitatea Politehnica din Bucuresti, Romania
José Coimbra, Universidade do Algarve, Portugal
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Noel Crespi, Institut TELECOM SudParis-Evry, France
Leonardo Dagui de Oliveira, Escola Politécnica da Universidade de São Paulo, Brazil
Kevin Daimi, University of Detroit Mercy, USA
Gerard Damm, Alcatel-Lucent, USA
Francescantonio Della Rosa, Tampere University of Technology, Finland
Chérif Diallo, Consultant Sécurité des Systèmes d'Information, France
Klaus Drechsler, Fraunhofer Institute for Computer Graphics Research IGD, Germany
Jawad Drissi, Cameron University , USA
António Manuel Duarte Nogueira, University of Aveiro / Institute of Telecommunications, Portugal
Alban Duverdier, CNES (French Space Agency) Paris, France
Nicholas Evans, EURECOM, France
Fabrizio Falchi, ISTI - CNR, Italy
Mário F. S. Ferreira, University of Aveiro, Portugal
Bruno Filipe Marques, Polytechnic Institute of Viseu, Portugal
Robert Forster, Edgemount Solutions, USA
John-Austen Francisco, Rutgers, the State University of New Jersey, USA
Kaori Fujinami, Tokyo University of Agriculture and Technology, Japan
Shauneen Furlong , University of Ottawa, Canada / Liverpool John Moores University, UK
Ana-Belén García-Hernando, Universidad Politécnica de Madrid, Spain
Bezalel Gavish, Southern Methodist University, USA

Christos K. Georgiadis, University of Macedonia, Greece
Mariusz Glabowski, Poznan University of Technology, Poland
Katie Goeman, Hogeschool-Universiteit Brussel, Belgium
Hock Guan Goh, Universiti Tunku Abdul Rahman, Malaysia
Pedro Gonçalves, ESTGA - Universidade de Aveiro, Portugal
Valerie Gouet-Brunet, Conservatoire National des Arts et Métiers (CNAM), Paris
Christos Grecos, University of West of Scotland, UK
Stefanos Gritzalis, University of the Aegean, Greece
William I. Grosky, University of Michigan-Dearborn, USA
Vic Grout, Glyndwr University, UK
Xiang Gui, Massey University, New Zealand
Huaqun Guo, Institute for Infocomm Research, A*STAR, Singapore
Song Guo, University of Aizu, Japan
Kamal Harb, KFUPM, Saudi Arabia
Ching-Hsien (Robert) Hsu, Chung Hua University, Taiwan
Javier Ibanez-Guzman, Renault S.A., France
Lamiaa Fattouh Ibrahim, King Abdul Aziz University, Saudi Arabia
Theodoros Iliou, University of the Aegean, Greece
Mohsen Jahanshahi, Islamic Azad University, Iran
Antonio Jara, University of Murcia, Spain
Carlos Juiz, Universitat de les Illes Balears, Spain
Adrian Kacso, Universität Siegen, Germany
György Kálmán, Gjøvik University College, Norway
Eleni Kaplani, Technological Educational Institute of Patras, Greece
Behrouz Khoshnevis, University of Toronto, Canada
Ki Hong Kim, ETRI: Electronics and Telecommunications Research Institute, Korea
Atsushi Koike, Seikei University, Japan
Ousmane Kone, UPPA - University of Bordeaux, France
Dragana Krstic, University of Nis, Serbia
Archana Kumar, Delhi Institute of Technology & Management, Haryana, India
Romain Laborde, University Paul Sabatier (Toulouse III), France
Massimiliano Laddomada, Texas A&M University-Texarkana, USA
Wen-Hsing Lai, National Kaohsiung First University of Science and Technology, Taiwan
Zhihua Lai, Ranplan Wireless Network Design Ltd., UK
Jong-Hyouk Lee, INRIA, France
Wolfgang Leister, Norsk Regnesentral, Norway
Elizabeth I. Leonard, Naval Research Laboratory - Washington DC, USA
Richard Li, Huawei Technologies, USA
Jia-Chin Lin, National Central University, Taiwan
Chi (Harold) Liu, IBM Research - China, China
Diogo Lobato Acatauassu Nunes, Federal University of Pará, Brazil
Andreas Loeffler, Friedrich-Alexander-University of Erlangen-Nuremberg, Germany
Michael D. Logothetis, University of Patras, Greece
Renata Lopes Rosa, University of São Paulo, Brazil
Hongli Luo, Indiana University Purdue University Fort Wayne, USA
Christian Maciocco, Intel Corporation, USA

Dario Maggiorini, University of Milano, Italy
Maryam Tayefeh Mahmoudi, Research Institute for ICT, Iran
Krešimir Malarić, University of Zagreb, Croatia
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Herwig Mannaert, University of Antwerp, Belgium
Michael Massoth, University of Applied Sciences - Darmstadt, Germany
Adrian Matei, Orange Romania S.A, part of France Telecom Group, Romania
Natarajan Meghanathan, Jackson State University, USA
Emmanouel T. Michailidis, University of Piraeus, Greece
Ioannis D. Moscholios, University of Peloponnese, Greece
Djafar Mynbaev, City University of New York, USA
Pubudu N. Pathirana, Deakin University, Australia
Christopher Nguyen, Intel Corp., USA
Lim Nguyen, University of Nebraska-Lincoln, USA
Brian Niehöfer, TU Dortmund University, Germany
Serban Georgica Obreja, University Politehnica Bucharest, Romania
Peter Orosz, University of Debrecen, Hungary
Patrik Österberg, Mid Sweden University, Sweden
Harald Øverby, ITEM/NTNU, Norway
Tudor Palade, Technical University of Cluj-Napoca, Romania
Constantin Paleologu, University Politehnica of Bucharest, Romania
Stelios Papaharalabos, National Observatory of Athens, Greece
Gerard Parr, University of Ulster Coleraine, UK
Ling Pei, Finnish Geodetic Institute, Finland
Jun Peng, University of Texas - Pan American, USA
Cathryn Peoples, University of Ulster, UK
Dionysia Petraki, National Technical University of Athens, Greece
Dennis Pfisterer, University of Luebeck, Germany
Timothy Pham, Jet Propulsion Laboratory, California Institute of Technology, USA
Roger Pierre Fabris Hoefel, Federal University of Rio Grande do Sul (UFRGS), Brazil
Przemyslaw Pochec, University of New Brunswick, Canada
Anastasios Politis, Technological & Educational Institute of Serres, Greece
Adrian Popescu, Blekinge Institute of Technology, Sweden
Neeli R. Prasad, Aalborg University, Denmark
Dušan Radović, TES Electronic Solutions, Stuttgart, Germany
Victor Ramos, UAM Iztapalapa, Mexico
Gianluca Reali, Università degli Studi di Perugia, Italy
Eric Renault, Telecom SudParis, France
Leon Reznik, Rochester Institute of Technology, USA
Joel Rodrigues, Instituto de Telecomunicações / University of Beira Interior, Portugal
David Sánchez Rodríguez, University of Las Palmas de Gran Canaria (ULPGC), Spain
Panagiotis Sarigiannidis, University of Western Macedonia, Greece
Michael Sauer, Corning Incorporated, USA
Marialisa Scatà, University of Catania, Italy
Zary Segall, Chair Professor, Royal Institute of Technology, Sweden
Sergei Semenov, Broadcom, Finland

Dimitrios Serpanos, University of Patras and ISI/RC Athena, Greece
Adão Silva, University of Aveiro / Institute of Telecommunications, Portugal
Pushpendra Bahadur Singh, MindTree Ltd, India
Mariusz Skrocki, Orange Labs Poland / Telekomunikacja Polska S.A., Poland
Leonel Sousa, INESC-ID/IST, TU-Lisbon, Portugal
Cristian Stanciu, University Politehnica of Bucharest, Romania
Liana Stanescu, University of Craiova, Romania
Cosmin Stoica Spahiu, University of Craiova, Romania
Young-Joo Suh, POSTECH (Pohang University of Science and Technology), Korea
Hailong Sun, Beihang University, China
Jani Suomalainen, VTT Technical Research Centre of Finland, Finland
Fatma Tansu, Eastern Mediterranean University, Cyprus
Ioan Toma, STI Innsbruck/University Innsbruck, Austria
Božo Tomas, HT Mostar, Bosnia and Herzegovina
Piotr Tyczka, ITTI Sp. z o.o., Poland
John Vardakas, University of Patras, Greece
Andreas Veglis, Aristotle University of Thessaloniki, Greece
Luís Veiga, Instituto Superior Técnico / INESC-ID Lisboa, Portugal
Calin Vladeanu, "Politehnica" University of Bucharest, Romania
Benno Volk, ETH Zurich, Switzerland
Krzysztof Walczak, Poznan University of Economics, Poland
Krzysztof Walkowiak, Wroclaw University of Technology, Poland
Yang Wang, Georgia State University, USA
Yean-Fu Wen, National Taipei University, Taiwan, R.O.C.
Bernd E. Wolfinger, University of Hamburg, Germany
Riaan Wolhuter, Universiteit Stellenbosch University, South Africa
Yulei Wu, Chinese Academy of Sciences, China
Mudasser F. Wyne, National University, USA
Gaoxi Xiao, Nanyang Technological University, Singapore
Bashir Yahya, University of Versailles, France
Abdulrahman Yarali, Murray State University, USA
Mehmet Erkan Yüksel, Istanbul University, Turkey
Pooneh Bagheri Zadeh, Staffordshire University, UK
Giannis Zaoudis, University of Patras, Greece
Liaoyuan Zeng, University of Electronic Science and Technology of China, China
Rong Zhao , Detecon International GmbH, Germany
Zhiwen Zhu, Communications Research Centre, Canada
Martin Zimmermann, University of Applied Sciences Offenburg, Germany
Piotr Zwierzykowski, Poznan University of Technology, Poland

International Journal on Advances in Telecommunications
Volume 11, Numbers 1 & 2, 2018

CONTENTS

Panagiotis Spapis, Huawei German Research Center, Germany
Chan Zhou, Huawei German Research Center, Germany

pages: 76 - 86

Energy-efficient Live Migration of I/O-intensive Virtual Network Services Across Distributed Cloud Infrastructures Leveraging Renewable Energies

Ngoc Khan Truong, Department of Applied Computer Science - Fulda University of Applied Sciences, Germany
Christian Pape, Department of Applied Computer Science - Fulda University of Applied Sciences, Germany
Sven Reißmann, Datacenter - Fulda University of Applied Sciences, Germany
Thomas Glotzbach, Department of Electrical Engineering and Information Technology - Hochschule Darmstadt University of Applied Sciences, Germany
Sebastian Rieger, Department of Applied Computer Science - Fulda University of Applied Sciences, Germany

pages: 87 - 100

Using Cisco VIRL and GNS3 to Improve the Scale-out of Large Virtual Network Testbeds in Higher Education

Sven Reißmann, Datacenter - Fulda University of Applied Sciences, Germany
Sebastian Rieger, Department of Applied Computer Science - Fulda University of Applied Sciences, Germany
Christoph Seifert, Department of Applied Computer Science - Fulda University of Applied Sciences, Germany
Christian Pape, Department of Applied Computer Science - Fulda University of Applied Sciences, Germany

Study of a Battery-less Near Field Communicating Sensor Network with ContactLess Simulator

David Navarro, Cédric Marchand, Laurent Carrel
Université de Lyon, ECL, INSA Lyon, UCBL, CPE
INL, UMR5270
F-69134, Ecully, France
david.navarro@ec-lyon.fr, cedric.marchand@ec-lyon.fr, laurent.carrel@ec-lyon.fr

Abstract - **Active tags and sensor nodes are an important part of the devices involved in the Internet of Things and in some cases it is impossible to power them using standard power supply or even battery. In such a case, battery-less smart sensors are needed. Design this kind of smart system is not an easy task and many different considerations must be taken into account. To face autonomy problems, battery-less sensor is a serious alternative. Energy harvesting is one of the main issue and it is mandatory to simulate the behavior of the system before designing and manufacturing it. In this article, we present ContactLess Simulator (CLS). It is developed in order to simulate contactless powered smart systems such as Near Field Communication (NFC) devices. CLS has been written using visual C# and is thus very flexible and easily portable. More precisely, this article focusses on a battery-less electronic systems: an autonomous NFC smart sensor. To design such a system, the energy budget has to be explored as the central point. CLS was developed exactly for this goal and this article describes a realistic test case to demonstrate it. The test case corresponds to a battery-less sensor node. It is composed of a microcontroller unit, a temperature sensor and a NFC circuit for communication and energy harvesting.**

Keywords-Simulation; Modelling; NFC; Sensor node; Microcontroller; MCU; Energy harvesting.

I. INTRODUCTION

This paper is an extended work of [1], about autonomous Wireless Sensor Networks energy study.

The Internet of Things (IoT) is now a well-known ecosystem, continuously growing (five times more are expected be connected in the next few years), where small and smart objects interact trough communicating networks. Even if these networks can be wired or wireless, the trend is to choose wireless communication in most cases. This is explained since small devices using wireless communications gives faster and easier installation and deployment. Among these objects, sensor nodes sense physical data in order to send information or actuate.

As shown in Fig. 1, they are usually composed of a microcontroller, a sensor, a communication device and a battery or an energy harvesting system.

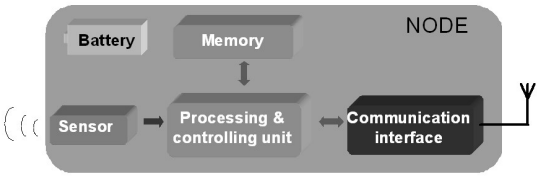


Figure 1. Typical Architecture of a smart object.

Concerning wireless communications, different solutions exist depending on the size of the network and on the communication range. Local Area Networks (LAN) are small networks whith small number of connected objects (up to few hundreds), they use short range communications (up to 300m). In this scope of networks, Bluetooth, Zigbee or Wifi are used for communication. At the contrary, Wide Area Networks are considered for big networks where more than 1000 objects are involved and where long range communication such as LoRa or mobile communications (3G, LTE, etc.) are used. Energy consumption consideration divides all these strategies of communication in two groups. The first one correspond to WAN which are often plugged with powerful supply, AC power supply or heavy acid-like battery. At the contrary, LAN and LoRa (based on the Low Power WAN) can be powered with small batteries and their autonomy can go up to years with lithium batteries and a smart usage.

It is even possible to design battery-less sensor node (energy-constrained electronic systems) in LAN and new trends target energy harvesting in the environment and smart management of this precious energy to create completely autonomous systems. We consider 2 kinds of harvesting: natural and artificial energy sources. Natural sources are directly given by the Nature, such as seismic vibrations, temperature, or solar rays. Many electronic systems are developed to consider these energy sources, such as for example Seebeck-effect systems or mini-photovoltaic panels [14]. Non-natural sources include for example machine vibrations (often coupled with piezoelectric elements [18]), ambient radiofrequency waves (through radiofrequency to continuous power converters), or magnetic fields such as wireless powering systems.

Wireless powering systems exist since several years, and are nowadays widespread in powering systems, such as inductive charging stations for smartphones [10] or vehicles wireless chargers [16]. Moreover, certain lightweight systems communicate at the same time they power the object. It is the case in RadioFrequency IDentification systems (RFID). They are composed of an emitter and a receiver, called tag. The emitter sends radiofrequency waves in order to power a tag and communicate at the same time. A classical tag answers its identification number. Near Field Communication (NFC) is based on RFID. It permits very short communications at a high frequency, in a full peer to peer mode. NFC inherits characteristics from RFID, network and smart card. It is suitable for secure communications; as an example, smartphone payment is possible with NFC.

This paper focuses on NFC battery-less objects. As Energy is the major constraint of these systems, a deep study has to be led at every stage. To achieve this goal, it is an absolute necessity to simulate the system before design and production. As these systems are electronic and communicating objects, they can be studied at different levels: low-level (hardware, software) or high-level (protocol, network). Many studies involve hardware platforms, like [3]. Low-level simulations focus on hardware or radiofrequency aspects. For example, [2] describes a MATLAB-Simulink model of a radiofrequency transceiver in order to provide a quick evaluation of the performances according to noise and non-linearity of each individual block in the transceiver. [6] and [17] present studies on the physical link (emitter, air, receiver) and radiofrequency propagation aspects. [8] gives a MATLAB Simulink NFC model for radiofrequency modulation study.

High-level simulations consider protocols, network and communication performance. RFIDSIM [9] is a more complete simulator that considers physical link and protocol. It provides a realistic physical layer and permits a multi-interface and multi-channel analysis. Others higher-level simulators focus on communication protocol and communication performance, such as the well-known NS-3 simulator. NFC models and protocol study have been developed over NS-3 [5].

However, no simulator takes the energy as the central point. A new system-level simulator, called Contact-Less Simulator [1], has been developed in order to precisely study these electronic systems from the energy point of view. This paper details this new simulator, presents new features and a real test case. CLS simulator considers energy harvesting from NFC emitter and energy balance according to the tag electrical consumption. The wireless-supplied tag is not only composed of a classical NFC circuit, but of a more complex smart system. It is possible to configure a sensor node and an application, the simulation then gives the answer if the node is well designed or not.

The rest of this article is organized as follow: Section II describes the NFC system considered in this study and details its signal transmission and typical architecture. Section III presents the simulator interface, the different possible configurations and details implemented models. Section IV describes the NFC hardware used as test case and shows simulation results. Finally, Section V concludes the paper and gives future improvement that will be included inside de simulator.

II. CONSIDERED NFC SYSTEM

A. Electronic system

A typical NFC system is composed of an emitter and a tag. It is shown in Fig. 2.

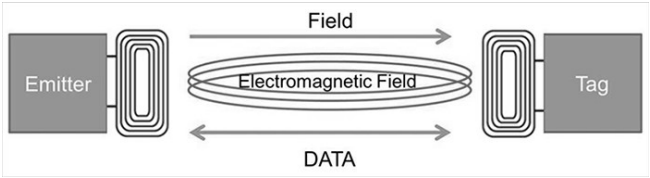


Figure 2. Typical Architecture of NFC system [13].

A tag is often composed of a NFC circuit that comprises 2 main sub-systems: energy harvesting part (for supply) and data decoding part (for communication). The energy harvesting block converts the received electromagnetic field into usable energy in order to supply the circuit. The data decoding block demodulates the signal in order to recover the bit-stream. As Energy is self-powered in passive tags, communication has to be initiated so that they are supplied in order to answer.

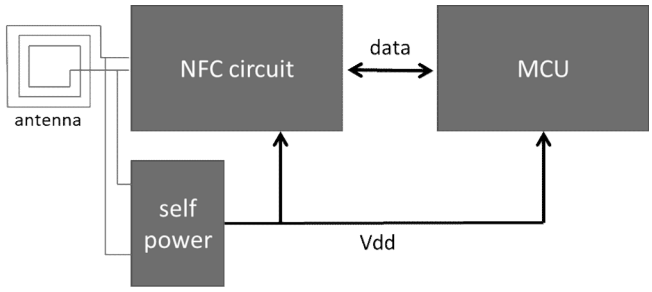


Figure 3. Classical NFC Tags

Fig. 3 shows the architecture of a classical tag composed of a microcontroller unit (MCU). When the tag is powered, the microcontroller executes the program as shown in Fig. 4.

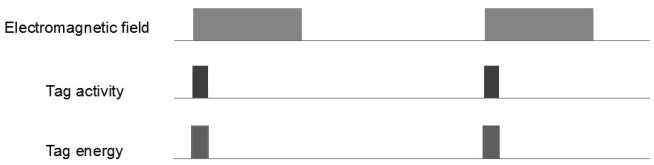


Figure 4. Classical use of NFC Tags

In this study, we consider a smarter tag comprising a microcontroller unit, a temperature sensor and a smart energy management block. Thus, this tag is called sensor node or wireless sensor node because of NFC (wireless) communications. Furthermore, we focus on a complete battery-less smart system where the only energy source comes from electromagnetic field during NFC communications between emitter and tag. We choose the ST-Microelectronics M24LR04E NFC chip for our test case. This chip is compatible with 13.56 MHz NFC ISO 15693 and ISO 18000-3 mode 1. In addition, it has an energy harvesting analog output, which makes it possible to supply other circuits on the board (i.e., microcontroller, sensor). The global considered system is shown in Fig. 2 and the tag architecture is detailed in Fig. 5.

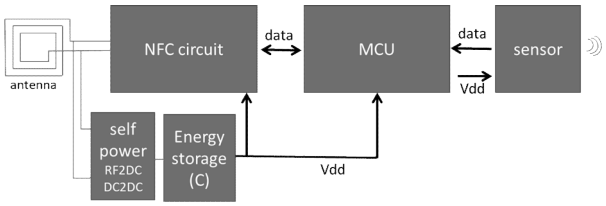


Figure 5. Considered battery-less smart tag.

The energy storage block is composed of two switches and storage capacitor as can be seen in Fig. 6. According to switches states, capacitor can charge (S1 closed, S2 opened) or discharge and supply circuits (S1 opened, S2 closed).

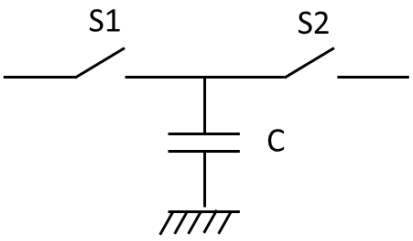


Figure 6. Energy storage block

This energy buffer offers new functionality: MCU and sensor can operate independently of NFC communications. We also target the energy use illustrated in Fig. 7. The electromagnetic field charges an energy tank, then this amount of energy is used along time by tag. Simulator will help us to check if it is possible.

Now the electronic system considered in this study has been described, the next section will provide the analysis of the signal propagation involved in such system. This will help to understand how the simulator can take into account the energy budget as the main constraint to validate a design.

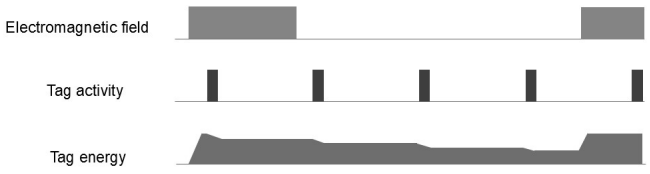


Figure 7. Targeted use of smart NFC Tag

B. Signal propagation

NFC systems use Low frequency (LF) from 125 KHz to 134 KHz, high Frequency (HF) at 13.56 MHz, or ultra-high frequency (UHF) at 860 or 960 MHz. We consider the 13.56 MHz frequency communication that is mostly used in NFC systems [13].

The emitter outputs a powerful signal in a coil (also called "antenna") at a specific -resonant- frequency. The electromagnetic field carries power and data with the help from an Amplitude Shift Keying (ASK) modulated signal [12].

Input power at receiver (tag) depends on signal strength at emitter, antennas gains, and distance between antennas. No simple exact equation exists since communication is near field and the Friis equation, which is used in classical long-range radiofrequency communications, is not valid. Indeed, Maxwell equations have to consider electric (E) and magnetic (H) fields. In our case, we use electrically small antennas. This implies that the near field limit distance r depends only on wavelength λ. This distance is given by the following equation [4], [7]:

$$r = \lambda/2\pi \quad (1)$$

As shown in Fig. 8, several kinds of radiofrequency signal propagation are observed according to the distance between emitter (represented as RFID reader and its antenna on the left) and the receiver (along the horizontal axis).

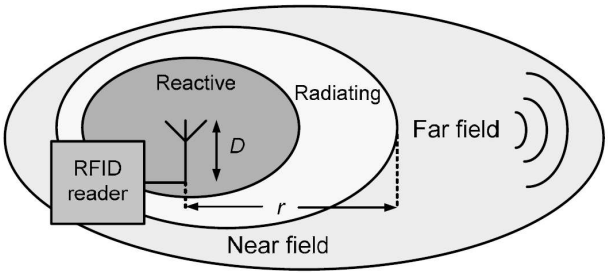


Figure 8. Radiofrequency signal propagation [14].

More precisely, the NFC system considered has frequency of 13.56 MHz and a distance between emitter and tag of few centimeters.

Thus, the tag is placed in reactive near field region leading to use simpler equations to model signal attenuation due to distance. Received power formulas are [12]:

$$P_{RX(E)} = \left(\frac{1}{k_1.d^2} - \frac{1}{k_2.d^4} + \frac{1}{k_3.d^6} \right) . P_{TX(E)} \quad (2)$$

$$P_{RX(H)} = \left(\frac{1}{k_4.d^2} + \frac{1}{k_5.d^4} \right) . P_{TX(H)} \quad (3)$$

Attenuation is also linked to pow of sixth, pow of four and pow of two of distance *d*; k1 to k5 are constants. Sometimes, simpler models are used, for example, [15] approximates power decay to be proportional to $1/d^6$.

III. CLS INTERFACE, MODELS AND SIMULATION

As it is shown in previous section, several electronic circuits compose the system. In order to provide a precise estimation of the energy consumption, it is necessary to model them separately. Models are high-level (electronic system level), they ar e written with C language.

Emitting power and antenna gain make it possible to calculate radiofrequency output power, in other terms the magnetic field H in mA/m. Frequency and distance between antennas lead to radiofrequency signal attenuation.

ContactLess Simulator (CLS) is organized as a set of configuration windows allowing the user to easily set up its system in order to launch the simulation. It has been developed in Visual C# in order to be easily portable on Microsoft 64-bit Windows operating systems. It is part of the Visual Studio Community, a free tool for academic research [11].

Graphical user interface is drawn in a horizontal way, from Emitter on the left towards Load (Electronic system) on the right. Fig. 9 shows the main windows of CLS where it is possible to recognize the global hardware architecture presented in Fig. 2. In this section, each part of CLS is described along with the model associated to each of the components of the electronic system.

A. Emitter model

The first menu in CLS (from the left) corresponds to the Field Generator (Emitter). It is modeled according to:

- Emitting power
- Frequency of the radiofrequency carrier
- Distance between emitter and tag antennas
- Emitter antenna gain

When the user enters in the Setup window of this part, a new window appears where all default meaningful values are prefilled, as shown in Fig. 10. Emitted electromagnetic field at emitter antenna and propagated electromagnetic field towards distance are calculated.

B. Tag antenna model

The emitter and tag antennas are PCB coil antennas in our prototype. Antenna gain is used to calculate propagation losses. According to above calculations, radiofrequency signal strength is known at tag antenna input. Tag antenna gain attenuation thus gives the signal after antenna. In CLS, Antenna setting, presented in Fig. 11, the tag antenna parameter is the gain (dBi). Magnetic field at tag antenna input is known from previous block. Tag antenna attenuation also decreases the signal.

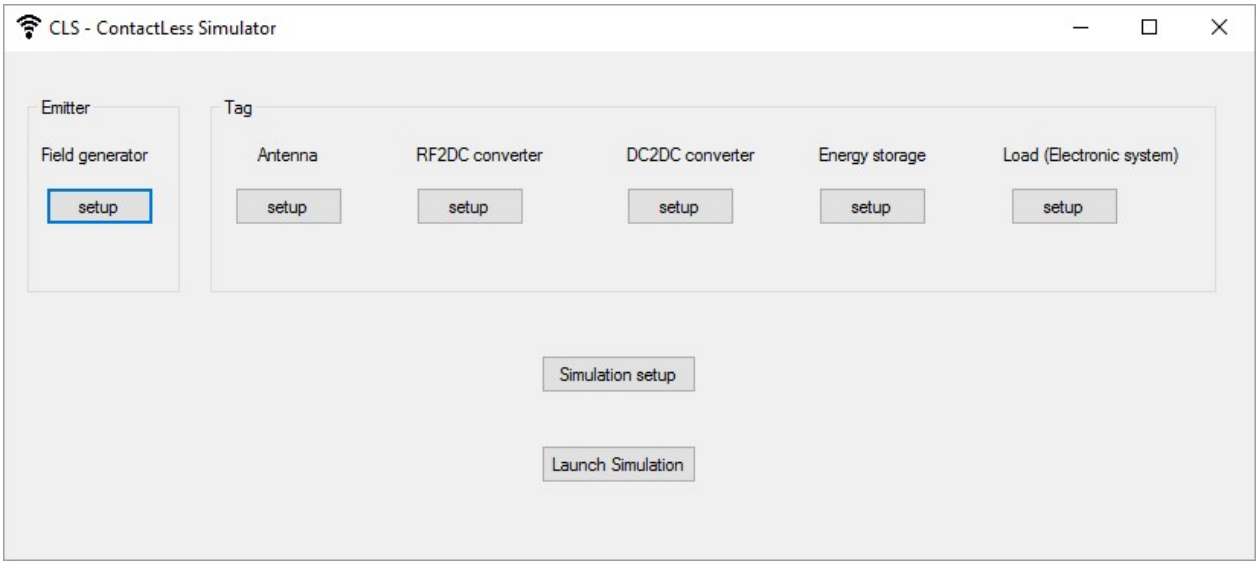


Figure 9. CLS Simulator graphical user interface

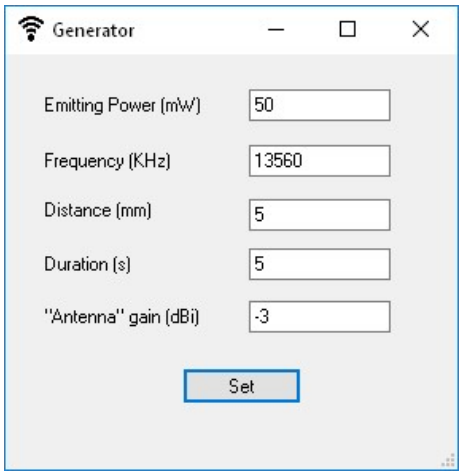


Figure 10. Electromagnetic field parameters at Emitter

Figure 11. Tag antenna parameters

C. NFC circuit model

NFC circuit has the task to demodulate the radiofrequency signal. Principle is to extract the low frequency information carried in the high frequency: the carrier. High frequency permits propagation in the medium (air, plastic packages, ...). Once the information is decoded, an answer can be send toward he emitter. Network-like communications occur in NFC. As this functionality is not the key point of this study, only the electrical consumption of this block is considered. To do this, this part is divided in two part in CLS: self-power and energy storage blocks.

D. Self-power and energy storage blocks

Several blocks in NFC circuit are considered. For a better understanding of these blocks, RF2DC and DC2DC are illustrated separately in Fig. 5. They are also modelled separately. RF2DC block receives the radiofrequency signal and converts it to a DC (voltage and current) signal. The aim is to create a power supply. Conversion goal is to extract the maximum electrical power. DC2DC block converts the RF2DC output, in order to create a usable voltage for electronic devices. Electrical power is given according to efficiency of blocks. As power is set, a good balance between voltage and current has to been chosen. Performances are based on ST-Microelectronics M24LR04E circuit characteristics. This communicating and harvesting circuit permits self-powering and provides an analog output to power other circuits in the system. Moreover, it permits data exchange between NFC circuit and microcontroller. If another circuit is to be used, global parameters, like efficiency, can be set.

1. RF2DC converter

The RF2DC part of the self-power block is shown in Fig 12. It calculates electrical power from the radiofrequency signal carrier. M24LR04E circuit has been modelled for this power conversion task. The datasheet of this circuit gives H field from radiofrequency strength and output power from H field.

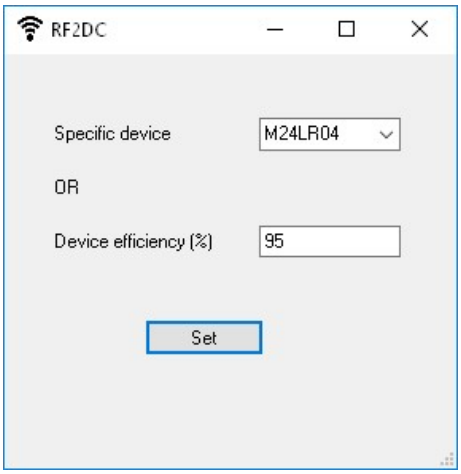


Figure 12. Radio to DC converter parameters

Then, the required current at output has to be known in order to calculate the output voltage Vout. According to the output current Isink: curves in datasheet give the output voltage the circuit can provide. For simpler use, a global parameter can be used for other circuits: the overall power efficiency (output versus input).

2. DC2DC converter

This block is presented in Fig. 13. Models have been separated because RF2DC and DC2DC blocks could be designed in separate circuits. Thus, in order to best distinguish individual performances in the system and to define parameters if they have to be designed, they are modelled in two separate blocks. DC2DC converter efficiency is expressed in percentage between output and input.

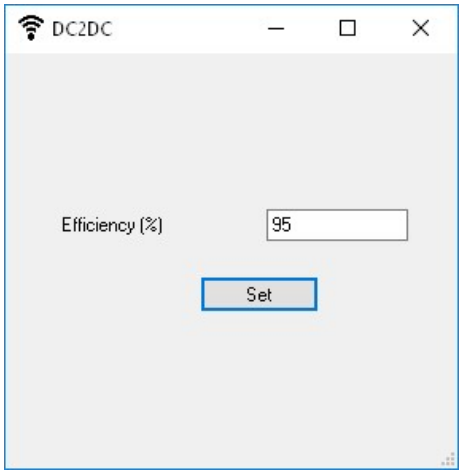


Figure 13. DC to DC converter parameters

3. Energy storage block

The energy storage parameters window presented in Fig. 14 are used to calculate the amount of energy that can be saved in an energy tank. We will later detail the novel switched capacitor architecture that is used. In Fig. 14, parameters are capacitance value (nF) and total leakages (fA) of switches and capacitor. The role of this module is to simulate the energy that can be stored (according to the power input from DC2DC bloc) and the energy that can be used (according to the supplied load and leakages). It will lead to an energy budget analysis.

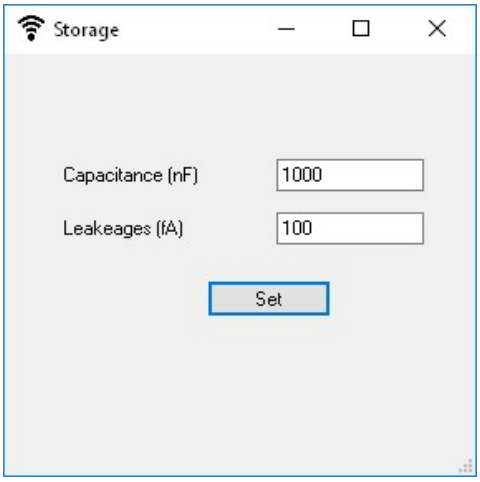


Figure 14. Load (electronic system) parameters

E. Microcontroller circuit model and sensor

For this release, microcontroller and sensor are simply modeled as an electrical load according to their activity. Microcontroller and sensor require a minimal voltage and consume a nominal current. In model, electrical power of microcontroller is calculated according to:

- Microcontroller brand and model,
- Oscillator type,
- Operating frequency.

At this stage, Microchip PIC18LF2525 is modeled. Table I shows all targeted microcontrollers that will be implemented in simulator.

TABLE I. TARGETTED MCUS MODELS

Microcontroller brand	Microcontroller model
ATMEL	ATMega-328
Microchip	PIC18LF2525
ST-Microelectronics	ST-8ML
ST-Microelectronics	STM-32
Texas Instruments	MSP-430

ATMEL, ST-Microelectronics and Texas Instruments models are under development and will be released soon. Sensor is a Maxim MAX6613 temperature sensor. Its analog output is connected to an ADC input of microcontroller. The MAX6613 is supplied with an output pin of microcontroller. Supply voltage is also the same for both components.

In this release, the load is a microcontroller unit (MCU) and a temperature sensor. When a microcontroller is chosen in the list-box, the "Configure MCU" button opens a new window. Oscillator type, operating frequency, desired supply voltage, active and sleep time of microcontroller are entered as shown in Fig. 15.

For this example, Microchip PIC18LF2525, oscillator type can be external RC (up to 4 MHz), external XTAL (crystal oscillator, up to 40 MHz), or internal oscillator. For internal oscillator, the example presented at the bottom of Fig. 15, user can select from the internal 8MHz source down to the 31KHz source.

Several frequencies are available according to frequency post-scaler in the microcontroller. Frequency (KHz) is the primary oscillator frequency that must match one possible configuration according to the oscillator type. Supply voltage of the microcontroller is then entered. All the parameters are taken from Microchip PIC18LF2525 datasheet; from current voltage versus frequency, current versus voltage and current versus frequency curves. Parameters for sensor are taken from Maxim MAX6613 datasheets.

Figure 15. Load (electronic system) parameters

F. *Simulation setup*

The simulation setup window (Fig. 16) permits to configure the simulation. The duration parameter will be used in future release in transient simulations. For the moment only static simulations are run. Design goal specifies which value is to maximize: voltage or current. Indeed, RF2DC and DC2DC blocks output a given power, and the couple voltage and current can vary. This option also configures the simulation in order to search the maximum current point or the maximum voltage point. The other parameter (for example, voltage if current is the design goal) is displayed as a result. Designer has to take it into account in design as a constraint. If the value of this other parameter is unreal, parameters concerning the hardware have to be changed, for example the microcontroller speed.

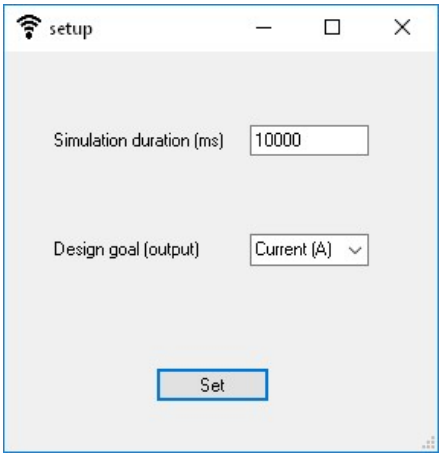


Figure 16. Simulation Setup window.

IV. TESTCASE AND RESULTS

To illustrate how the simulation behaves, a test case has been simulated. All parameters used in setup windows are summarized in Table II. Simulation time for a static analysis is configured to few milliseconds. Results are presented in Fig. 17. The left part shows results on tag side where the magnetic field H is calculated. It depends on emitter power, distance, and antennas gains. Then the harvesting part gives output of RF2DC and DC2DC parts in the M24LR04E.

TABLE II. PARAMETERS USED FOR TESTCASE SIMULATION

Emitting power	50 mW
Frequency of carrier for NFC communication	13560 MHz
Distance emitter / tag	5 mm
Duration of NFC communication	5 s
Antennas gain (emitter & tag)	-3 dBi
RF2DC & DC2DC converters	Equations from M24LR04E
Energy storage capacitance	1000 nA
Switches leakages (energy storage)	100 fA
Microcontroller brand/model	Microchip / PIC18LF2525
Microcontroller Oscillator / Operating voltage range	Internal oscillator, 31KHz / 2V to 5.5V
Sensor / Operating voltage range	Maxim MAX6613 / 1.8V to 5.5V
Duration of node sensing and MCU processing and storage	100 ms
Simulation setup	Design goal = current

As design goal of the example is the output current, the simulator calculates the nominal current to retrieve from harvesting part. Nominal current is fixed by the required current from the electrical load (MCU and sensor). This current is calculated from MCU and sensor parameters: PIC18LF2525 requires 15.05µA when running from internal oscillator at 31KHz. MAX6613 consumes 7.5µA. Simulator then calculates the harvesting possible voltage output for a sink current of 22.55µA. From datasheet curves in the model, simulation gives 2.67V. Fig. 17 shows that the required power for load is 67.64µW but the harvester can only provide 60.2µW. As a result, the required voltage 3V is not reached. Designer will have to deal with a 2.67V supply, or decrease load current consumption in order to increase supply voltage.

Energy calculations are also implemented in the simulator. It considers electrical power consumption and active time. Active time for the emitter correspond to the duration of the electromagnetic field. Result window thus displays 301µJ for 5s duration. Active time for the MCU is time while MCU is running (time running parameter in Fig. 15). Its energy is then 6.76µJ for 100ms. This result allows the designer to plan how many cycles the MCU could run with a single electromagnetic charge. It is 44 cycles for this example.

Battery-less system is also possible: for example, an application in industry where a sensor network is deployed in the aim of measuring a temperature on several machines. Each evening, a person will read the data on each sensor node with a 5 seconds NFC communication. While downloading data from node to phone (or tablet), the node is being charged for the next day: it is able to make a measurement every 33 minutes for the next 24 hours.

V. CONCLUSIONS AND FUTURE WORKS

In this article, a concrete energy analysis of novel electronics architecture and ContactLess Simulator was presented. CLS is a simulator dedicated to energy efficiency in battery-less sensor node. Its interface is based on setting windows to graphically configure a NFC system composed of an emitter and a smart tag. As targeted tags are battery-less (self-supplied), it comprises an energy harvesting module with a RF2DC and DC2DC converter, a microcontroller unit (MCU) and a sensor. Each hardware block is configured by a setup window form. Simulation can be tuned for one design goal: search maximal voltage or maximal current. This choice depends on designer priority. A launch button runs the simulation and displays a result window. Several electrical outputs are calculated: electromagnetic field at tag input, harvested power (voltage and current), harvested energy for a single contactless energy intake, required power (voltage and current) for the microcontroller and sensor, required energy for a main program loop. Result analysis on a realistic testcase shows that the harvested power is a bit weak (60.2µW) compared to the required power (67.64µW). According to the design goal, fixed to prioritize current, harvested voltage is 2.67v instead of 3v. Meanwhile, the sensing node is supplied with the required current (22.55µA). Energy calculation makes it possible to think about a better use of the energy. Indeed,

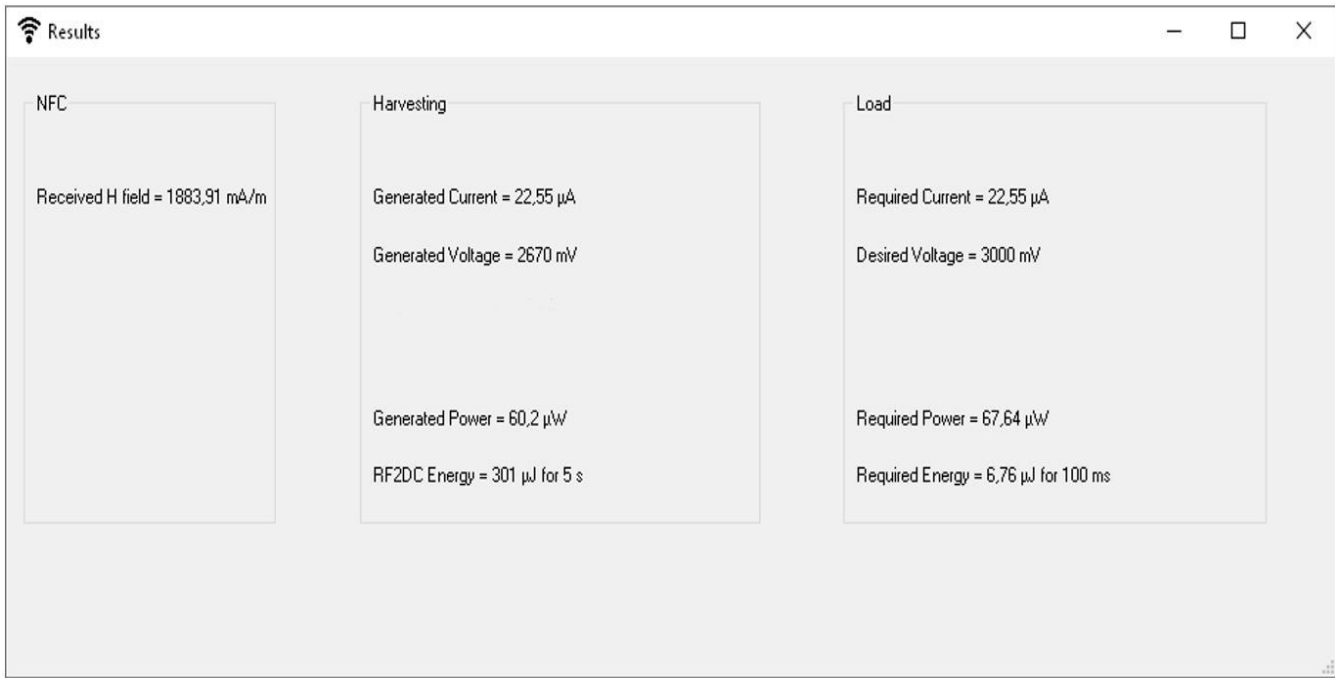


Figure 17. Result window

harvested power is weak but harvested energy (301µJ) is much bigger that required energy (6.76µJ). Feasibility is also proven: it is possible to charge the energy buffer during a 5 seconds NFC communication (while previous data are retrieved) then to use the battery-less system for 44 data recording (temperature measurements).

ContactLess Simulator can be improved to support transient analysis and to propose devices according to design constraints for example. Other microcontroller and sensor will be added and the global interface will be changed to be closer to Fig. 5. Another possible improvement will be to add curves in the results window in order to helps the designer to know understand where the sensor node can be modified to meet the constraint. Finally, it may be interesting to implement realtime static analysis to the simulator to provide an efficient support during the design phase and dedicate the simulation to transient analysis.

REFERENCES

[1] D. Navarro and G. Migliato-Marega, "Cls: Contactless simulator," in The 12th International Conference on Wireless and Mobile Communications, Barcelona, Spain, 2016, pp. 65–69.

[2] D.-K. Ahn, S.-G. Bae, and I.-C. Hwang, "A design of behavioral simulation platform for near field communication transceiver using matlab simulink," The Transactions of The Korean Institute of Electrical Engineers, vol. 59, no. 10, pp. 1917–1922, 2010.

[3] C. Angerer, B. Knerr, M. Holzer, A. Adalan, and M. Rupp, "Flexible simulation and prototyping for rfid designs," in Procceedings of the First International EURASIP Workshop on RFID Technology, 2007.

[4] Atmel, "Understanding the requirements of iso/iec 14443 for type b proximity contactless identification card. nov. 2005," Application Note, [Online, retrieved: 15th July 2017]. Available from: http://www.atmel.com/images/doc2056.pdf.

[5] G. Benigno, O. Briante, and G. Ruggeri, "A sun energy harvester model for the network simulator 3 (ns-3)," in 2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking - Workshops (SECON Workshops), June 2015, pp. 1–6.

[6] T. Cheng and L. Jin, "Analysis and simulation of rfid anti-collision algorithms," in The 9th International Conference on Advanced Communication Technology, vol. 1, Feb. 2007, pp. 697–701.

[7] A. B. Constantine et al., "Antenna theory: analysis and design," MICROSTRIP ANTENNAS, third edition, John wiley & sons, 2005.

[8] J. Deckmar and A.Perez-Boutavin, "Nfc," [Online, retrieved: 15th July 2017]. Available from: https://fr.mathworks.com/matlabcentral/fileexchange/34915-nfc.

[9] C. Floerkemeier and S. Sarma, "Rfidsim - a physical and logical layer simulation engine for passive rfid," IEEE Transactions on Automation Science and Engineering, vol. 6, no. 1, pp. 33–43, Jan. 2009.

[10] U. K. Madawala and D. J. Thrimawithana, "A bidirectional inductive power interface for electric vehicles in v2g systems," IEEE Transactions on Industrial Electronics, vol. 58, no. 10, pp. 4789–4796, Oct. 2011.

[11] Microsoft, "Visual studio community," [Online, retrieved: 15th July 2017]. Available from: https://www.visualstudio.com/enus/products/visual-studio-community-vs.aspx.

[12] P. V. Nikitin, K. V. S. Rao, and S. Lazar, "An overview of near field uhf rfid," in 2007 IEEE International Conference on RFID, March 2007, pp. 167–174.

[13] G. Proehl, "An introduction to near field communications," STMicroelectronics, [Online, retrieved: 15th July 2017]. Available from: http://www.st.com/content/st_com/en/applications/connectivity/nearfield-communication-nfc.html, 2013.

[14] S. Roundy, D. Steingart, L. Frechette, P. Wright, and J. Rabaey, Power Sources for Wireless Sensor Networks. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1–17. [Online]. Available: https://doi.org/10.1007/978-3-540-24606-0_1

[15] H. G. Schantz, "Near field propagation law a novel fundamental limit to antenna gain versus size," in 2005 IEEE Antennas and Propagation Society International Symposium, vol. 3A, July 2005, pp. 237–240 vol. 3A.

[16] R. Tseng, B. von Novak, S. Shevde, and K. A. Grajski, "Introduction to the alliance for wireless power loosely-coupled wireless power transfer system specification version 1.0," in 2013 IEEE Wireless Power Transfer (WPT), May 2013, pp. 79–83.

[17] J. Wang and D. Yang, "Design of a multi-protocol rfid tag simulation platform based on supply chain," in 2009 International Conference on Management and Service Science, Sept. 2009, pp. 1–4.

[18] W. S. Wang, W. Magnin, N. Wang, M. Hayes, B. O'Flynn, and C. O'Mathuna, "Bulk material based thermoelectric energy harvesting for wireless sensor applications," Journal of Physics: Conference Series, vol. 307, no. 1, p. 012030, 2011. [Online]. Available: http://stacks.iop.org/1742-6596/307/i=1/a=012030

Experimental Analysis and Resolution Proposal on Performance Degradation of TCP over IEEE 802.11n Wireless LAN Caused by Access Point Scanning

Toshihiko Kato, Kento Kobayashi, Sota Tasaki, Masataka Nomoto, Ryo Yamamoto, and Satoshi Ohzahata
Graduate School of Informatics and Engineering
University of Electro-Communications
Tokyo, Japan
kato@is.uec.ac.jp, kento_kobayashi@net.is.uec.ac.jp, t.souta@net.is.uec.ac.jp, noch@net.is.uec.ac.jp,
ryo_yamamoto@is.uec.ac.jp, ohzahata@is.uec.ac.jp

***Abstract*— In IEEE 802.11 wireless LAN, there is a problem that the access point scanning at stations which uses the power management function gives impacts on the performance of TCP communication. This paper is an extension of our previous paper that showed the result of experiments on this problem for uploading and downloading TCP data transfer over 802.11n wireless LAN. For the uploading transfer, we analyze the influence to TCP throughput focusing on the TCP small queues that limit the amount of data in wireless LAN's sending queue. We add some new results and discussions to the previous paper. As for the downloading transfer, we discuss the influence by the TCP congestion control algorithm at TCP senders and A-MPDU transmission rate at the access point. We also propose a method to stop access point scanning while data transfer is being performed, and show the results of the performance evaluation of the proposed method implemented on the Linux operating system.**

Keywords- WLAN; IEEE802.11n; Access Point Scanning; Power Management; TCP Small Queues; TCP Congestion Window Validation; NetworkManager; Linux Kernel Module.

I. INTRODUCTION

This paper is an extension of our previous paper [1] presented in an IARIA conference.

Nowadays, 802.11n [2] is one of most widely adopted IEEE wireless LAN (WLAN) standards. It establishes high speed data transfer using the higher data rate support (e.g., 300 Mbps), the frame aggregation in Aggregation MAC Protocol Data Unit (A-MPDU) and the Block Acknowledgment mechanism. On the other hand, TCP introduces some new functions to establish high performance over high speed WLANs. CoDel [3] and TCP small queues [4] that aim to resolve the Bufferbloat problem [5] are examples.

We reported some performance evaluation results on TCP behaviors over 802.11n [6][7]. While we were conducting those experiments, we encountered the situation that the TCP throughput decreases periodically. By analyzing the captured WLAN frame logs during the performance degradation, we confirmed that its reason is the periodical transmission of data frames without data (*Null data frames*) with the power management field set to 1 in the WLAN header. These frames are used by WLAN stations to inform access points that the stations are going to sleep and to ask the associated access point not to send data frames. It is pointed out that WLAN stations use Null data frames to scan another available access point periodically [8].

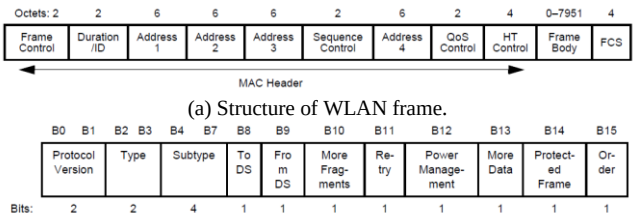
In our previous paper [1], we conducted detailed performance analysis and reported that the impacts of Null data frames on TCP data transfer change by the functions of TCP used in the communication. Specifically, the TCP sender behaviors and the throughput degradation depend on the direction of TCP data transfer (uploading from station to access point or downloading in the reverse direction), whether the TCP small queues are used or not, and what kind of congestion control algorithm is used. This paper is an extension of the previous paper, and shows the results of experimental analysis about the impacts on TCP throughput given by the access point scanning, by adding new results and discussions. This paper also proposes a method to stop access point scanning while data transfer is being performed. We implement the proposed method over *NetworkManager* software module running over the Linux operating system. This paper describes the implementation and performance evaluation of the proposed method.

The rest of this paper consists the following sections. Section II shows the technologies relevant to this paper. Section III explains the experimental settings. Sections IV and V give the detailed analysis of uploading TCP data transfer and downloading TCP data transfer together with access point scanning, respectively. Section VI proposes a method to protect the throughput degradation by the access point scanning. In the end, Section VII gives the conclusions of this paper.

II. RELEVANT TECHNOLOGIES

A. Power management function and Null data frames

As described above, IEEE 802.11 standards introduce the power management function. In the WLAN frame format depicted in Figure 1(a), bit 12 in the Frame Control field is the *Power Management field* (shown in Figure 1(b)). By setting this bit to 1, a station informs the associated access point that it is going to the *power save mode*, in which the station goes to sleep and wakes up only when the access point sends beacon frames. By setting the bit to 0, it informs the access point that it goes back to the *active mode*.



(a) Structure of WLAN frame.

(b) Structure of Frame Control field.

Figure 1. IEEE 802.11 WLAN frame format [2].

This function is used for several purposes. One example is the case that a station is actually going to sleep to save its power consumption for a while. In this case, a station wakes up only at the timing of receiving beacon frames from the access point. If the access point has data frames to deliver to the sleeping station, it indicates this fact in the Traffic Indication Map element in a beacon frame. In response to this information, the station requests the delivery of data frames by use of a PS-Poll frame.

Another example is the access point scanning. For example, a Lunix terminal executing the NetworkManager module searches periodically for an access point which provides stronger radio signal than the current access point with which the terminal is associated [9]. There are two schemes for the access point scanning; the passive scanning in which a station waits for a beacon frame from another access point, and the active scanning in which a station sends a probe request frame and waits for a probe response frame for the request. In either scheme, a station needs to ask the current access point to stop sending data frames to it. For this purpose, the station sends a frame with the Power Management field set to 1.

Null data frames are used by WLAN stations to inform access points of the shift to the power save mode or to the active mode [10]. This is a frame that contains no data (Frame Body in Fig. 1 (a)). An ordinary data frame has the type of '01' and the subtype of '0000' in the Frame Control field. That is, B3 and B2 in Fig. 1 (b) are 0 and 1, and B7 through B4 are all 0. On the other hand, Null data frame has the type of '01' and the subtype of '0100'. While an ordinary data frame has three Address fields, a Null data frame has only two Address fields; the transmitter is a station MAC address and the receiver is an access point MAC address. By using Null data frames with the Power Management field set to 0 or 1, stations can request the power management function for access points.

B. TCP small queues

In the Linux operating system with version 3.6 and later, a mechanism called TCP small queues [4] is installed in order to resolve the Bufferbloat problem. It keeps watching on the queues in Linux schedulers and device drivers in a sending terminal. If the amount of data stored in the queues is larger than the predefined queue limit, it suspends the TCP module until the amount of stored data becomes smaller than the limit. During this TCP suspension, the data which applications transmit is stored in the TCP send socket buffer, which may cause the application to be suspended if the TCP send socket buffer becomes full. After the TCP module is resumed, it processes the application data stored in the send socket buffer.

The default value of the predefined queue limit is 128 Kbyte, and it is adjustable by changing the following parameter;

/proc/sys/net/ipv4/tcp_limit_output_bytes.

This mechanism is different from the other mechanisms against the Bufferbloat problem, such as CoDel, in the point that no TCP segments are discarded intentionally to reduce TCP congestion window size.

C. Congestion window validation

The TCP congestion control uses the congestion window size (*cwnd*) maintained in TCP senders. TCP senders transmit TCP data segments under the limitation of cwnd and the window size advertised by TCP receivers. In general, cwnd is increased when TCP senders receive TCP acknowledgment (ACK) segments and is decreased when any data segments are retransmitted.

This mechanism is considered to work well under the assumption that the data transfer throughput is limited by cwnd. But it is possible that the throughput is controlled by an application in a TCP sender. In this case, the amount of data segments floating over network without being acknowledged (it is called *flight size*) might be smaller than cwnd. In an application limited case, however, cwnd also increases when the sender receives a new ACK segment, and the value of cwnd may be much larger than the current flight size. This means that the value of cwnd is invalid stage. If the TCP sender changes its status from application limited to cwnd limited suddenly, TCP segments corresponding to an invalid i.e. too large cwnd value will rush into a network.

In order to resolve such a problems, the congestion window validation (CWV) mechanism is proposed. RFC 2861 [11] proposes the following two rules. (1) A TCP sender halves the value of cwnd if no data segments are transmitted during a retransmission timeout period. (2) When a TCP sender does not send data segment more than cwnd during a round-trip time (RTT), then it decreases cwnd to

$$(cwnd + sent\ data\ size)/2$$

in the next RTT time frame.

RFC 7661 [12] revises the above rules and defines a new rule that, if the data size acknowledged during one RTT is smaller than half of cwnd, a TCP sender does not increase cwnd in the next RTT time frame. It defines the procedure in the case of congestion separately.

III. Experimental Settings

Figure 2 shows the configuration of the performance evaluation experiment we conducted. There are two stations conforming to 802.11n with 5GHz band and one access point connected to a server through 1Gbps Ethernet. One station called STA1 is associated with the access point, and communicates with the server through the access point. The other station called STA2 is used just to monitor WLAN frames exchanged between STA1 and the access point.

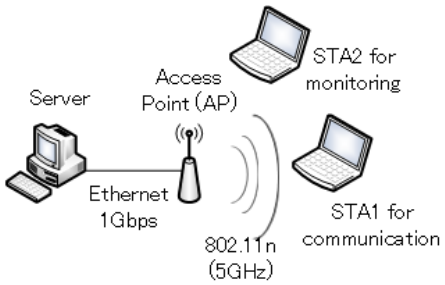


Figure 2. Network configuration in experiment.

TABLE I. SPECIFICATION OF NOTEBOOK AND ACCESS POINT.

NOTE	Manufacturer/Model	DELL Insilon 14
	Operating system	Ubuntu 14.04LTS (kernel 3.13) or Ubuntu 12.04LTS (kernel 3.2)
	WLAN driver	ath9k
AP	Manufacturer/Model	BUFFALO AirStation WZR-HP-AG300H
	WLAN chip	Atheros AR7161
	WLAN driver	ath9k

We use commercially available notebook PCs for the stations and the server. The access point used is also off-the-shelf product. The detailed specification of the notebook and the access point is shown in Table I. The access point is able to use the 40 MHz channel bandwidth and provides the MAC level data rate from 6.5 Mbps to 300 Mbps.

In the experiment, the data is generated by iperf [13] in the upload direction from STA1 to the server and the download direction from the server to STA1. The conditions for the experiment are the followings.

- Use or non-use of the TCP small queues, and
- use of CUBIC TCP [14] or TCP NewReno [15] as a congestion control mechanism.

When we use the TCP small queues, we installed Ubuntu 14.04 LTS in the notebook, and in the case not to use it, we installed Ubuntu 12.04 LTS.

During the data transmissions, the following detailed performance metrics are collected for the detailed analysis of the communication;

- the packet trace at the server, STA1 and STA2 , by use of *tcpdump*,
- the TCP throughput for every second, calculated from packet trace at TCP sender,
- the WLAN related metrics, such as the MAC level data rate, the number of A-MPDUs sent for a second and the number of MPDUs aggregated in an A-MPDU (an average during one second), from the device driver ath9k [16] at the access point and STA1, and
- the TCP congestion window size at the server, by use of *tcpprobe* [17] (an average during one second, calculated from the values obtained for every segment reception at the server).

IV. Analysis of Uploading TCP Data Transfer

In the experiments for uploading TCP data transfer, the results were different depending on whether the TCP small queues are used or not. On the other hand, the TCP congestion control algorithms did not affect the results so much. This section shows the results for uploading TCP data transfer focusing on the use or non-use of TCP small queues using CUBIC TCP.

A. Results when TCP small queues are used

Figure 3 shows the time variation of TCP throughput and cwnd, both of which are average value during one second, in the case that the TCP small queues are used in STA1. From this result, we can say that the throughput degradations occur periodically. Specifically, each *throughput degraded period* is around 10 sec. and such a period happens approximately once in 120 sec. In a *normal period*, the average TCP throughput is 136 Mbps, but it decreases to as much as 57 % in a throughput degraded period.

On the other hand, cwnd does not decrease even in a throughput degraded period, which means that there are no packet losses. Besides that, the increase of cwnd is depressed throughout the TCP communication. In this result, the value of cwnd is limited to around 200 packets.

Figure 4 shows the time variation of the MAC level data rate (average during one second). From this figure, it can be said that the data rate keeps high value. Figures 5 and 6 give

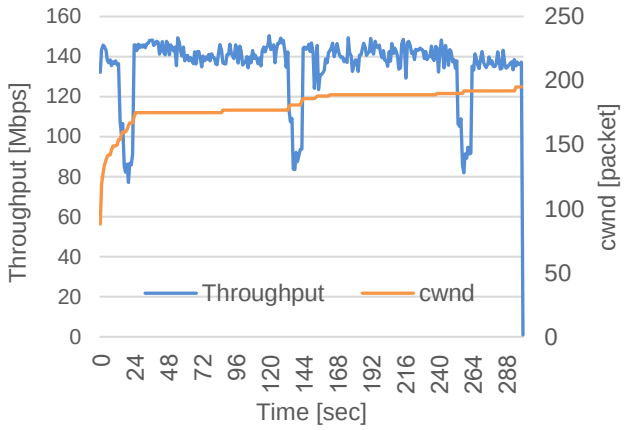


Figure 3. TCP throughput and cwnd vs. time in uploading data transfer with TCP small queues.

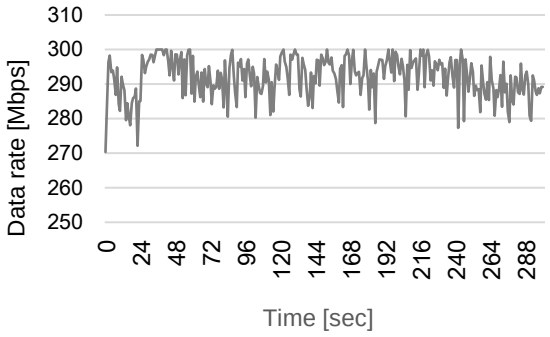


Figure 4. MAC level data rate vs. time in uploading data transfer with TCP small queues.

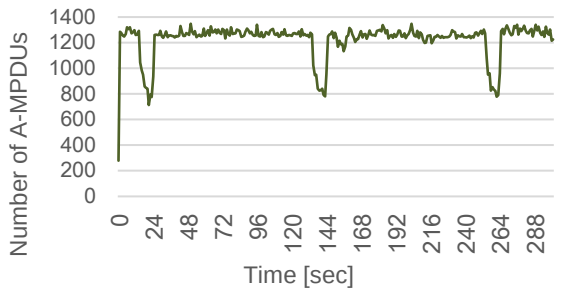


Figure 5. Number of A-MPDUs vs. time in uploading data transfer with TCP small queues.

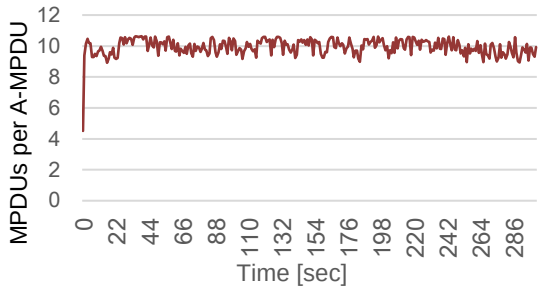


Figure 6. Number of A-MPDUs vs. time in uploading data transfer with TCP small queues.

the time variation of the number of A-MPDUs for one second and the number of MPDUs aggregated in an A-MPDU (an average during one second), respectively. Here, the number of MPDUs degrades during the throughput degraded period, while the number of MPDUs per A-MPDU keeps the same level. The decrease of the number of A-MPDUs is the reason for the periodic throughput degradation.

Figure 7 shows the packet trace, captured by STA2, of WLAN frames without data in the throughput degraded period starting at time 25 sec. This is the result of analysis by Wireshark and contains the number of frame, the time of packet capture measured from the TCP SYN segment, the source and destination MAC address, the protocol type (802.11), the frame length, and the information including name and parameters.

The frame whose number is 105,932, shown inverted to blue in the figure, is a Null data frame. For this frame, the Info column says that "Flags=. . .P. . .TC." This means that the Power Management field is set to 1 in this frame. The transmitter of this frame is STA1, whose MAC address is "LiteonTe_0b:ce:0c," and its receiver is the access point whose MAC address is "BuffaloI_27:2a:39." The sequence control ("SN" in the figure) is 1750 in this frame. In this packet trace, this Null data frame is not acknowledged by the access point. Instead, the same Null data frames are retransmitted eight times by the frames with numbers 105,933 through 105,953. They have the same sequence control (1750), and the Retry field in the Frame Control field is set to 1, as indicated "Flags=. . .PR. .TC" in the figure. From the fourth retransmission, a RTS frame is used before sending a Null data frame and the access point responds it by returning a CTS frame, which allows STA1 to send a Null data frame. But, from the fourth to the seventh retransmissions, the access point does not send any ACK frames. In the end, the access point sends an ACK frame at the eighth retransmission (see the frame with number 105,954). These frame exchanges takes 1.8 msec.

Then, the frame with number 105,955 is a beacon frame broadcasted by the access point. The duration between the ACK frame (No. 105,954) and this beacon frame is 90 msec in the figure.

24 msec after the beacon frame is transmitted, STA1

No.	Time	Source	Destination	Protocol	Length	Info
105931	25.814733	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	58	802.11 Block Ack, Flags=........C
105932	25.815296	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...P...TC
105933	25.815318	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105934	25.815323	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105935	25.815326	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105938	25.815343	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	58	802.11 Block Ack, Flags=........C
105939	25.817033	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	46	Request-to-send, Flags=...P....C
105940	25.817054		LiteonTe_0b:ce:0c	802.11	40	Clear-to-send, Flags=........C
105941	25.817066	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105942	25.817070	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	46	Request-to-send, Flags=...P....C
105943	25.817073		LiteonTe_0b:ce:0c	802.11	40	Clear-to-send, Flags=........C
105944	25.817077	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105945	25.817080	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	46	Request-to-send, Flags=...P....C
105946	25.817083		LiteonTe_0b:ce:0c	802.11	40	Clear-to-send, Flags=........C
105947	25.817088	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105948	25.817091	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	46	Request-to-send, Flags=...P....C
105949	25.817095		LiteonTe_0b:ce:0c	802.11	40	Clear-to-send, Flags=........C
105950	25.817099	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105951	25.817102	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	46	Request-to-send, Flags=...P....C
105952	25.817105		LiteonTe_0b:ce:0c	802.11	40	Clear-to-send, Flags=........C
105953	25.817111	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1750, FN=0, Flags=...PR..TC
105954	25.817116		LiteonTe_0b:ce:0c	802.11	40	Acknowledgement, Flags=........C
105955	25.907580	BuffaloI_27:2a:39	Broadcast	802.11	379	Beacon frame, SN=3924, FN=0, Flags=........C, BI=100, SSID=klab-n/a
105956	25.931649	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1751, FN=0, Flags=.......TC
105958	25.931677		LiteonTe_0b:ce:0c	802.11	40	Acknowledgement, Flags=........C
105959	25.931682	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1751, FN=0, Flags=....R..TC
105960	25.931686		LiteonTe_0b:ce:0c	802.11	40	Acknowledgement, Flags=........C
105961	25.931691	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1751, FN=0, Flags=....R..TC
105962	25.932891	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1751, FN=0, Flags=....R..TC
105963	25.932914	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	46	Request-to-send, Flags=........C
105964	25.932918		LiteonTe_0b:ce:0c	802.11	40	Clear-to-send, Flags=........C
105965	25.932922	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	57	Null function (No data), SN=1751, FN=0, Flags=....R..TC
105966	25.932927		LiteonTe_0b:ce:0c	802.11	40	Acknowledgement, Flags=........C
105967	25.932937	LiteonTe_0b:ce:0c	BuffaloI_27:2a:39	802.11	54	Null function (No data), SN=1752, FN=0, Flags=.......TC
105968	25.932940		LiteonTe_0b:ce:0c	802.11	40	Acknowledgement, Flags=........C

Figure 7. An example of packet trace during throughput degraded period focusing on WLAN frames without data

sends a Null data frame with the Power Management field set to 0, whose number and sequence control is 105,956 and 1751, respectively. Again, this Null data frame is retransmitted four times. In this case, although STA2 captures the corresponding ACK frames from the access point, STA1 retransmits it, repeatedly. After this frame is acknowledged by the ACK frame with number 105,966, STA1 transmit the next Null data frame whose sequence control 1752, and it is immediately acknowledged. These frame exchanges take 1.3 msec.

During this sequence, STA1 and the access point do not send any data frames. This paper refers to the time period as a *sleeping period*. After the last Null data frame with the Power Management field set to 0 is acknowledged, the data transfer from STA1 is restarted. But, several hundred milliseconds later, the similar communication sequence including the Null data frames and beacon frame occurs, and it goes to another sleeping period. This paper refers to the period when data frames are transmitted between sleeping periods as an *awoken period*. During a performance degraded period, there are around 26 pairs of sleeping period and awoken period.

The relationship among those periods are shown in Figure 8. In the evaluation result, the throughput degraded periods and the normal periods are repeated as shown at the top of this figure. The average duration for them is around 10 sec. and 110 sec., respectively. A throughput degraded period consists of sleeping periods and awoken periods. The average duration for them is 110 msec and 320 msec, respectively. As described in Figure 7 before, a sleeping period consists of a period sending Null data frames with the Power Management field set to 1, a period waiting for a beacon frame, and a period sending Null data frames with the Power Management field set to 0. The average durations for the individual periods are shown in Figure 8.

On the other hand, Figure 3 shows that the increase of cwnd is suppressed even if there are no packet losses. The reason is considered to be the collaboration of the TCP small queues and CWV. As described before, CWV intends to be used when a TCP communication is application limited. In the case the TCP small queues are used, however, it is possible that the data transfer stops when the buffered data in the sending queue for WLAN device exceeds the predefined queue limit, even if the flight size is smaller than cwnd. In this case, the MAC level data rate dominates the throughput instead of cwnd in the TCP level, and therefore, the control by CWV becomes effective. Actually, the source program of the TCP small queues implements a procedure such that, when the unacknowledged data amount is smaller than the current value of cwnd in the slow start phase, cwnd is not incremented even if a new ACK segment arrives. Note that this procedure itself is not conforming to RFC 2861 strictly but similar to RFC 7661, which was not standardized when the TCP small queues were introduced.

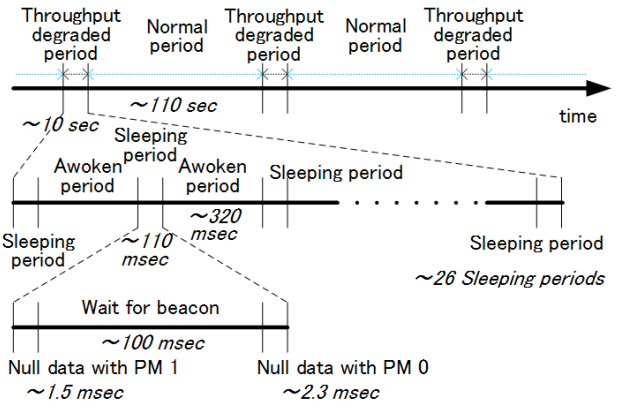


Figure 8. Detailed analysis of time periods.

B. Results when TCP small queues are not used

Figure 9 shows the time variation of TCP throughput and cwnd in the case that the TCP small queues are not used in STA1. In this case, the throughput degradations also occur periodically. In the normal periods, the average throughput is 174 Mbps, which is 38 Mbps higher than the case using the TCP small queues. This result seems to come from the fact that the cwnd value goes up to 970 packets. On the other hand, in the throughput degraded periods, the throughput decreases as low as 10 % of that in the normal period. The value of cwnd decreases largely in the throughput degraded periods.

Figure 10 shows the time variation of the MAC level data rate. In comparison with Figure 4, there are some drops of data rate at the timing of the throughput drop, but the drop is from 300 Mbps to around 280 Mbps, so it can be said that

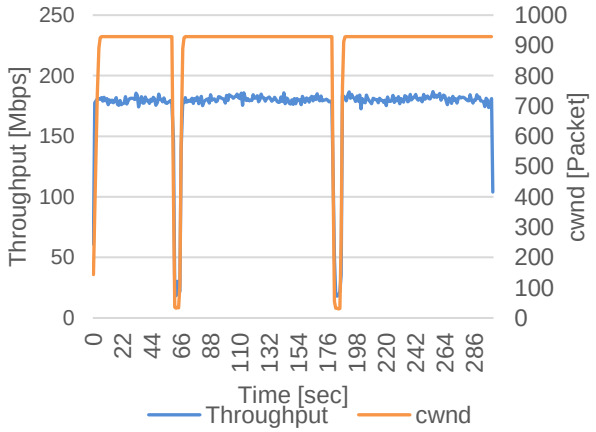


Figure 9. TCP throughput and cwnd vs. time in uploading data transfer without TCP small queues.

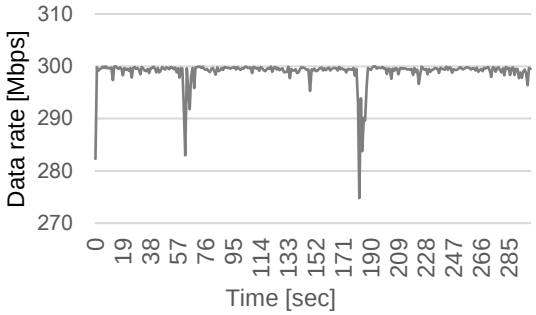


Figure 10. MAC level data rate vs. time in uploading data transfer without TCP small queues.

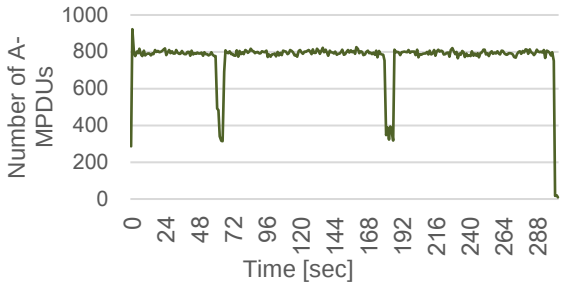


Figure 11. Number of A-MPDUs vs. time in uploading data transfer without TCP small queues.

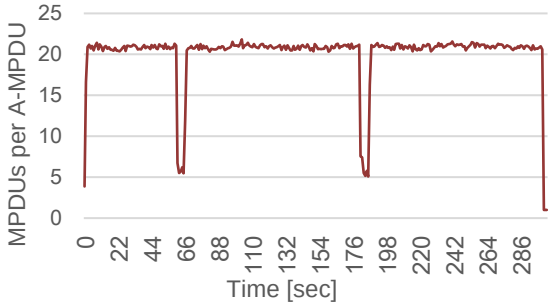


Figure 12. Number of A-MPDUs vs. time in uploading data transfer without TCP small queues.

the drop itself is not large. Figures 11 and 12 show the time variation of the number of A-MPDUs for one second and the number of MPDUs aggregated in an A-MPDU, respectively. In the case that the TCP small queues are used, only the number of MPDUs in an A-MPDU decreases in the throughput degraded periods. However, when the TCP small queues are not used, the number of MPDUs also decreases in the throughput degraded periods.

In this case, the reason for the periodic throughput degradation is also the periodic access point scanning using Null data frames with the power management function. In this case, however, there are some differences compared with the case of using the TCP small queues. At first, in the normal periods, the throughput and cwnd have larger values. On the other hand, in the throughput degraded periods, the drop of throughput is sharp, and cwnd as well as the number of MPDUs in one second drop sharply.

The reason is that there are some packet losses in the throughput degraded periods. The detailed discussions on the difference between the use and non-use of the TCP small queues are given in the following subsection.

C. Discussions

Figure 13 shows how the internal modules behave during a throughput degraded period, when the TCP small queues are used. During a sleeping period within a throughput degraded period, the WLAN module (WLAN interface hardware and its device driver) stops sending data requested from the IP module. As a result, several data are stored in the send queue maintained by the operating system and the driver. If the number of stored data exceeds the threshold, the TCP module stops reading data from the APP, which means application, module and they are kept in the send

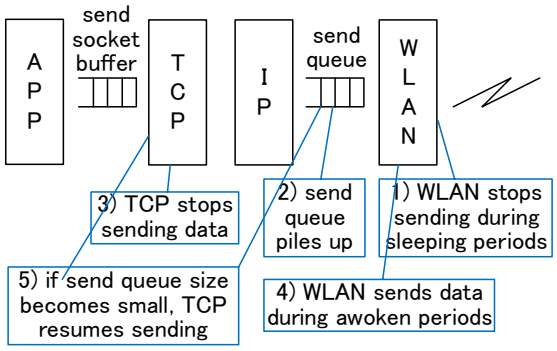


Figure 13. Behaviors when TCP small queues are used.

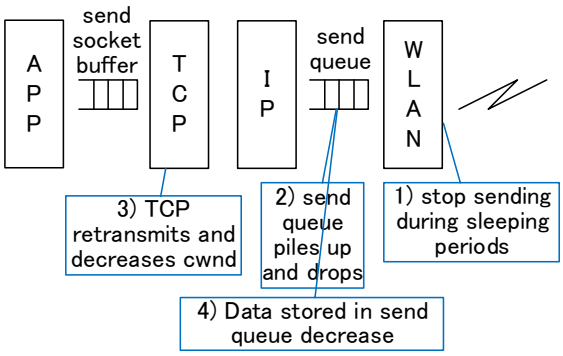


Figure 14. Behaviors when TCP small queues are not used.

socket buffer. This is the function of the TCP small queues. If the send socket buffer becomes full, the APP module stops sending data. In this situation, the WLAN module just keeps sending data stored in the send queue during the awoken periods, which is the same as the normal data transfer situation. If the size of data stored in the send queue is smaller the threshold, the TPC module resumes the sending. Therefore, only the TCP throughput and the number of A-MPDUs sent in one second decrease as shown in Figures 3 through 6.

Figure 14 shows the behaviors of the internal modules during a throughput degraded period, when the TCP small queues are not used. During a sleeping period within a throughput degraded period, the WLAN module stops sending data requested from the IP module. As a result, several data are stored in the send queue. This is the same as above. In this case, however, cwnd in the TCP module keeps increasing for every ACK segment while data segments are not lost. In the results in Figure 9, cwnd goes up to 970 packets as described above. During a sleeping period, the WLAN module stops sending data, but the TCP module keeps transmitting data, and so it is possible that the send queue overflows and some data segments are discarded. The TCP module retransmits lost data and decreases cwnd. As a result, the size of data stored in the send queue is also reduced. This means that the traffic load to the WLAN module is reduced and, in the awoken periods, the WLAN traffic is reduced. This corresponds to the drop of the TCP throughput, the number of A-MPDUs sent, and the number of MPDUs aggregated in an A-MPDU, as shown in Figures 9 through 12.

V. Analysis of Downloading TCP Data Transfer

In the experiments for downloading TCP data transfer, the results were not different depending on the use or non-use of TCP small queues. This is because neither the TCP small queues nor CWV are implemented at the access point. Instead, the results slightly depended on the TCP congestion control algorithms in the server. This section shows these results.

A. Results when CUBIC TCP is used

Figure 15 shows the time variation of TCP throughput and cwnd in the case that CUBIC TCP is used at the server. From this figure, we can say that the periodic throughput degradation also occurs at the downloading TCP data transfer. By analyzing the monitoring results of WLAN frames captured by STA2, we confirmed that there are periodic exchanges of Null data frames and beacon frames between STA1 and the access point, which is similar with the sequence in the uploading TCP data transfer. So, in the case of downloading TCP data transfer, the access point scanning by WLAN stations reduces the throughput.

As shown in the figure, cwnd at the server takes the value between 350 and 500 packets. The drops of cwnd indicate that packet losses occur frequently. The increase of cwnd takes a cubic curve of time, which is characteristic for CUBIC TCP.

Figure 16 shows the time variation of the MAC level data rate. Mostly the MAC data rate of 270 Mbps is kept. There is one drop, but the rate is still as high as 240 Mbps. It can be said that the MAC level data rate maintains a high level.

In contrary to the upload results, the throughput in a throughput degraded period drops sharply, although the cwnd value does not decrease largely during this period. In addition, the throughput just after a throughput degraded period is rather low. In order to investigate those results, we checked the number of A-MPDUs sent in one second by the access point, and the number of MPDUs contained in one A-MPDU. The results are given in Figures 17 and 18. These figures show that both A-MPDUs and MPDUs per A-MPDU decrease largely in throughput degraded periods. This result accounts for the throughput reduction. Besides that, the number of MPDUs aggregated in an A-MPDU is low just after a throughput degraded period. This is considered the reason for low throughput in this time frame.

B. Results when TCP NewReno is used

Figure 19 shows the time variation of TCP throughput and cwnd in the case that TCP NewReno is used at the server. Figure 20 shows the time variation of the MAC level data rate. Figures 21 and 22 show the time variation of the number of A-MPDUs sent in one second by the access point, and the number of MPDUs contained in one A-MPDU, respectively.

From Figure 19, it is confirmed that the throughput is degraded sharply in every 120 sec. From the monitoring results of WLAN frames captured by STA2, we also confirmed the periodic exchanges of Null data frames and beacon frames between STA1 and the access point. This is an access point scanning by STA1 and the reason for the

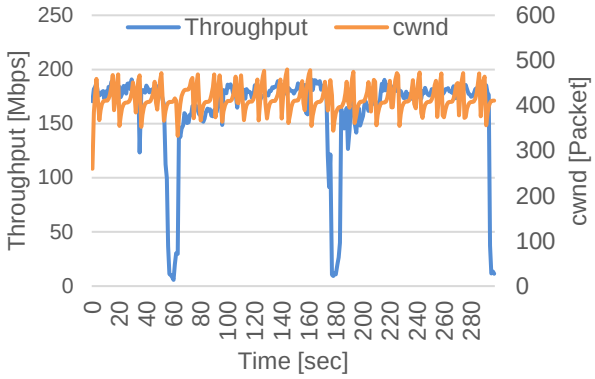


Figure 15. TCP throughput and cwnd vs. time in downloading data transfer using CUBIC TCP at server.

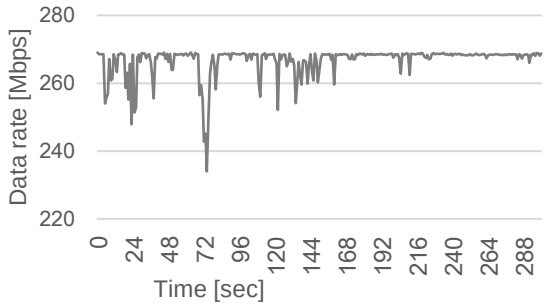


Figure 16. MAC level data rate vs. time in downloading data transfer using CUBIC TCP at server.

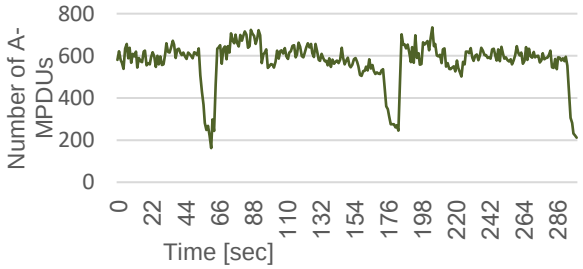


Figure 17. Number of A-MPTU vs. time in downloading data transfer using CUBIC TCP at server.

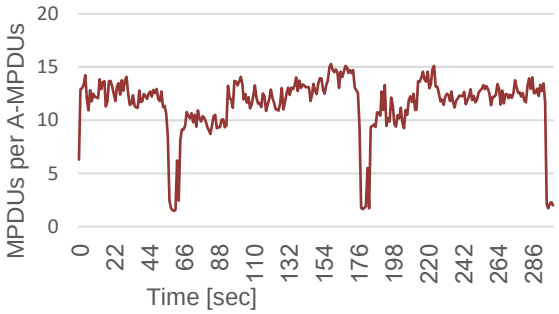


Figure 18. Number of MPTUs in A-MPDU vs. time in downloading data transfer using CUBIC TCP at server.

throughput reduction. This figure also shows that cwnd at the server takes the value between 300 and 500 packets, and that cwnd drops frequently, similarly with the case of CUBIC TCP. The increase of cwnd takes a linear curve along with time, which is characteristic for TCP NewReno.

From Figure 20, it can be said that, although there some decreases, the MAC level data rate keeps high value such as 270 Mbps.

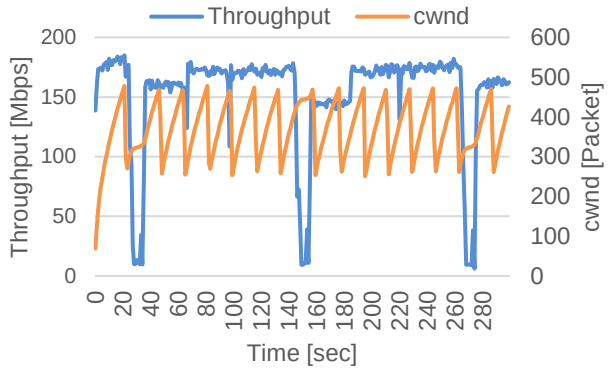


Figure 19. TCP throughput and cwnd vs. time in downloading data transfer using TCP NewReno at server.

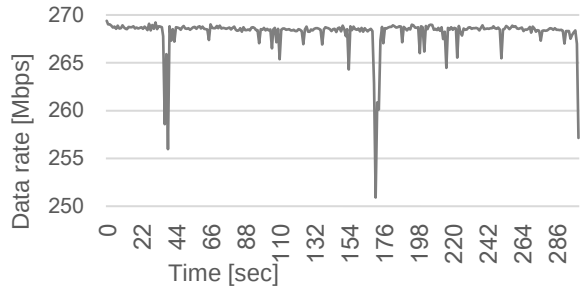


Figure 20. MAC level data rate vs. time in downloading data transfer using TCP NewReno at server.

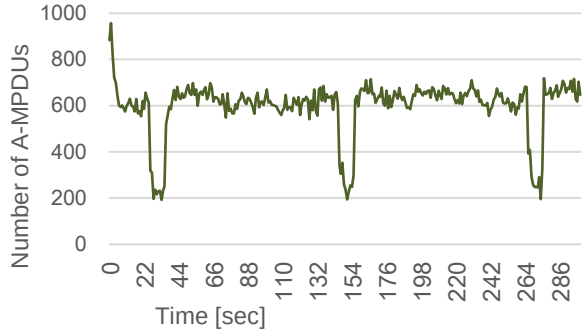


Figure 21. Number of A-MPTU vs. time in downloading data transfer using TCP NewReno at server.

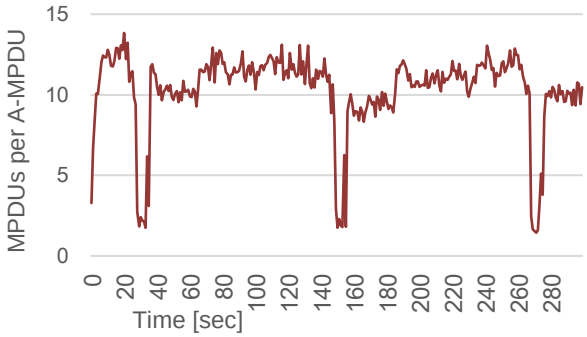


Figure 22. Number of MPTUs in A-MPDU vs. time in downloading data transfer using TCP NewReno at server.

While the throughput in a throughput degraded period drops sharply, the cwnd value does not decrease largely. The throughput just after a throughput degraded period is rather low. These are similar with the case of CUBIC TCP. As Figures 21 and 22 show, the numbers of transmitted A-MPDUs and MPDUs per A-MPDU decrease largely in throughput degraded periods. Besides that, the number of MPDUs aggregated in an A-MPDU is low just after a throughput degraded period. This will be the reason for low throughput in this time frame. These results are similar with the case of CUBIC TCP.

VI. Method to Stop Access Point Scanning

In this section, we present a method to avoid the throughput degradation by stopping access point scanning while data transfer is being performed.

A. Implementation

As we mentioned above, the access point scanning is realized by the NetworkManager software module in the Linux operating system [18]. It is executed as a daemon to support network configuration and operation. The software of NetworkManager is maintained within the Linux package management system that maintains binary files of various software, configuration files, and the information about their dependencies. Since the Ubuntu distribution we use in this paper depends on the Debian package management system, based on a tool called `dpkg` with the `apt` (advanced packaging tool) system.

Figure 23 shows a Linux script that allows a modified NetworkManager source program to be installed within the Debian package management system [19]. In the first step, the binary NeworkManager is installed using an `apt-get install` command. Then, some software necessary for package building is installed in the second step. After that, the source program of NetworkManager is downloaded under the subdirectory `net-man` using an `apt-get source` command. By this step, several c/h files are expanded under the directory `network-manager-0.9.4.0`. The number `0.9.4.0` indicates the version of NetworkManager, which corresponds to Ubuntu 14.04 LTS. The source files are expanded in the `src` directory under `network-manager-0.9.4.0`. At this timing, the downloaded software are modified as described below. This is step 4. Step 5 is preparing the package building, and the `debuild` command builds the NetworkManager package. By these steps, there are several `deb` files generated. In the end, a `dpkg` command generates an executable file for NetworkManager.

The access point scanning is realized by the C language functions described in Figure 24.

- In the function `schedule_scan()`, the function `request_wireless_scan()` is called periodically by use of `g_timeout_add_seconds()`.
- The function `g_timeout_add_seconds()` is a function to make another function called at regular intervals, which is defined in the framework of GTK+ (the GIMP Toolkit) [20].

```
# step 1: install binary NetworkManager
sudo apt-get install network-manager
sudo apt-get update

# step 2: prepare package building
sudo apt-get install dpkg-dev devscripts fakeroot

# step 3: get NetworkManager source
mkdir src
cd src
apt-get source network-manager

# step 4: modify NetworkManager source

# step 5: install build package
sudo apt-get build-dep network-manager

# step 6: build NetworkManager
cd network-manager-0.9.4.0
debuild -uc -us -b

# step 7: install modified NetworkManager
sudo dpkg -i ../*.deb
```

Figure 23. Linux script to build NetworkManager from source program.

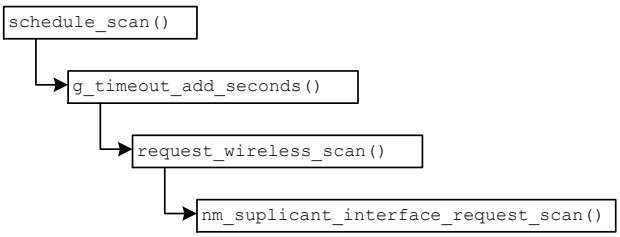


Figure 24. Functions for access point scanning.

- The function `request_wireless_scan()` calls `nm_suplicant_interface_request_scan()`, which performs access point scanning actually.

Based on this consideration, we modified the function `request_wireless_scan()` in the way that, if some data transfer has been performed since the last scan request, the actual scanning is skipped.

B. Performance evaluation

We evaluate the TCP throughput of the proposed method. The evaluation is done in the same configuration described in Section III. We use the uploading data transfer with the TCP small queues and CUBIC TCP. Figures 25 and 26 show the time variation of TCP throughput, which is an average during one second, without and with the proposed method implemented in a station, respectively. As shown in Figure 25, the TCP throughput is degraded around every 120 second, when the proposed method is not implemented. This is similar with the result in Figure 3. On the other hand, Figure 26 shows that, when the proposed method is implemented in a station, these periodic large drops of TCP throughput do not occur. These results indicate that the proposed method can resolve the throughput degradation problem caused by the access point scanning.

VII. CONCLUSIONS

This paper discussed the results on performance evaluation on the periodic TCP throughput degradation in IEEE 802.11n WLAN. The degradation is invoked by the periodic access point scanning using Null data frames with the Power Management field set to on and off. We showed that the throughput degradation is different depending on whether the TCP small queues are used or not in an uploading TCP data transfer, and what type of TCP congestion control algorithms are used in a downloading TCP data transfer. In the uploading data transfer, the TCP small queues and the congestion window validation suppress both of the increase of congestion window size in a normal period and its decrease in a throughput degradation period. So, the throughput degradation invoked by the access point scanning is smaller than the case when the TCP small queues are not used. In the downloading data transfer, the congestion control algorithms give some impacts on the time variation of congestion window size. However, the actual reason for the throughput degradation is the decrease in the transmission rate of A-MPDUs and the number of MPDUs in one A-MPDU, which are observed at an access point. This paper also proposed a method to resolve the performance degradation caused by access point scanning by stopping scanning while data transfer is being done. The proposed method was implemented in the NetworkManager software module running over the Linux operating system, and the performance evaluation showed that the performance degradation is resolved by the proposed method.

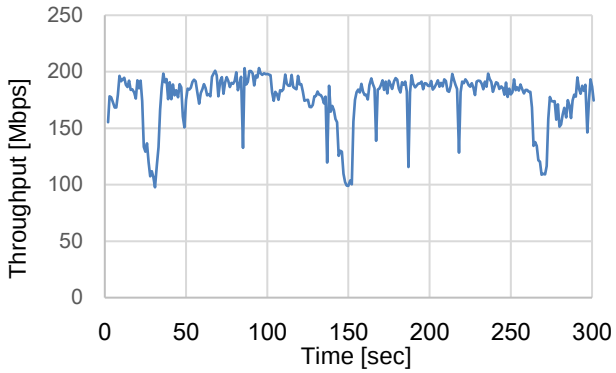


Figure 25. TCP throughput vs. time in uploading data transfer without proposed method.

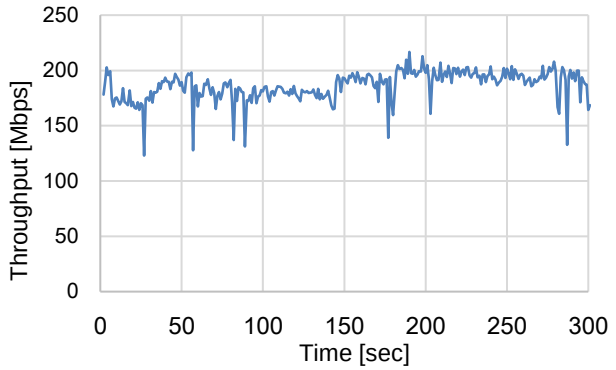


Figure 26. TCP throughput vs. time in uploading data transfer with proposed method.

REFERENCES

[1] K. Kobayashi, Y. Hashimoto, M. Nomoto, R. Yamamoto, S. Ohzahata, and T. Kato, "Experimental Analysis on Access Point Scanning Impacts on TCP Throughput over IEEE 802.11n Wireless LAN," Proc. ICWMC 2016, pp. 115-120. Nov. 2016.

[2] IEEE Standard for Information technology: Local and metropolitan area networks Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2012.

[3] K. Nichols and V. Jacobson, "Controlling Queue Delay," ACM Queue, Networks, vol. 10, no. 5, pp. 1-15, May 2012.

[4] Eric Dumazet, "[PATCHv2 net-next] tcp: TCP Small Queues," http://article.gmane.org/gmane.network.routing.codel/68, 2012, retrieved Feb. 2018.

[5] J. Gettys and K. Nichols, "Bufferbloat: Dark Buffers in the Internet," ACM Queue, Virtualization, vol. 9, no. 11, pp. 1-15, Nov. 2011.

[6] M. Nomoto, T. Kato, C. Wu, and S. Ohzahata, "Resolving Bufferbloat Problem in 802.11n WLAN by Weakening MAC Loss Recovery for TCP Stream," Proc. PDCN 2014, pp. 293-300, Feb. 2014.

[7] Y. Hashimoto, M. Nomoto, C. Wu, S. Ohzahata, and T. Kato, "Experimental Analysis on Performance Anomaly for Download Data Transfer at IEEE 802.11n Wireless LAN," Proc. ICN 2016, pp. 22-27, Feb. 2016.

[8] My80211.com, "802.11: Null Data Frames," http://www.my80211.com/home/2009/12/5/80211-null-data-frames.html, Dec. 2009, retrieved Feb. 2018.

[9] ath9k-devel@lists.ath9k.org, "disable dynamic power save in AR9280," http://comments.gmane.org/gmane.linux.drivers.ath9k.devel/5199, Jan. 2011, retrieved Feb. 2018.

[10] W. Gu, Z. Yang, D. Xuan, and W. Jia, "Null Data Frame: A Double-Edged Sword in IEEE 802.11 WLANs," IEEE Trans. Parallel & Distributed Systems, vol. 21, no. 7, pp. 897-910, Jul. 2010.

[11] N, Handley, J. Padhye, and S. Floyd, "TCP Congestion Window Validation," IETF RFC 2861, Jun. 2000.

[12] G. Fairhurst, A. Sathiaseelan, and R. Secchi, "Updating TCP to Support Rate-Limited Traffic," IETF RFC 7661, Oct. 2015.

[13] iperf, http://iperf.sourceforge.net/, retrieved Feb. 2018.

[14] S. Ha, I. Rhee, and L. Xu, "CUBIC: A New TCP-Friendly High-Speed TCP Variant," ACM SIGOPS Operating Systems Review, vol. 42, no. 5, pp. 64-74, July 2008.

[15] S. Floyd, T. Henderson, and A. Gurtov, "The NewReno Modification to TCP's Fast Recovery Algorithm," IETF RFC 3728, April 2004.

[16] ath9k Linux Wireless, https://wireless.wiki.kernel.org/en/users/drivers/ath9k, retrieved Feb. 2018.

[17] Linux foundation: tcpprobe, http://www.linuxfoundation.org/collaborate/workgroups/networking/tcpprobe, retrieved Feb. 2018.

[18] ubuntu documentation, "NetworkManager," https://help.ubuntu.com/community/NetworkManager, retrieved Feb. 2018.

[19] Devian/Wiki/, "Building Turorial," https://wiki.debian.org/BuildingTutorial, retrieved Feb. 2018.

[20] The GTK+ Project, "What is GTK+, and how can I use it?" https://www.gtk.org/, retrieved Feb. 2018.

Flexible Spatial Light Modulator Based Coupling Platform for Photonic Integrated Processors

Cátia Pinho[1,2], George S. D. Gordon[4], Berta Neto[1], Tiago M. Morgado[1], Francisco Rodrigues[1,3], Ana Tavares[1,3], Mário Lima[1,2], Timothy D. Wilkinson[4], António Teixeira[1,2]

[1] Instituto de Telecomunicações (IT), University of Aveiro, Aveiro, 3810-193, Portugal
[2] Department of Electronics, Telecommunications and Informatics (DETI), University of Aveiro, Portugal
[3] PICadvanced, University of Aveiro, Incubator, PCI – Creative Science Park Via do Conhecimento, Ílhavo, Portugal
[4] Electrical Division, Engineering Department, University of Cambridge, 9, JJ Thomson Avenue, Cambridge, UK
e-mail: catiap@ua.pt, gsdg2@cam.ac.uk, bneto@av.it.pt, tmcm@ua.pt, francisco@picadvanced.com, ana@picadvanced.com, mlima@ua.pt, tdw13@cam.ac.uk, teixeira@ua.pt

Abstract — **Enhanced Photonic Integrated Circuits (PIC) are required for the current demand of flexibility and reconfigurability in telecommunications networks. However, the technical and functional requirements of the PIC demand a thorough characterization and testing to provide an accurate prediction of the PIC performance. In the characterization and testing context, the use of Spatial Light Modulator (SLM) can be beneficial. SLM is a diffractive device to reconstruct images from Computer Generated Holograms (CGH) that allows to modulate the wavefront of a light beam. This capability can be explored to feed/receive optical signals to the PIC, i.e., as a flexible SLM based coupling platform. The feasibility of this approach was tested with the generation of a multiplexing/demultiplexing CGH to be applied into an optical chip for data compression based on Haar wavelet transform. Simulation results for building blocks as well as the all-optical network for the optical data compression chip are presented, supporting their theoretical feasibility. A new concept to use a SLM as a flexible coupling platform to complement PIC characterization process is proposed and a Haar transform data compression PIC described.**

Keywords - photonic integrated circuits (PIC); integrated optics; spatial light modulator (SLM); computer generated holography (CGH); all-optical devices; Haar transform.

I. INTRODUCTION

In the recent years, we have witnessed a significant increase in the data traffic, which the traditional copper based electronic media fail to carry [1]–[3]. Furthermore, the increasing demand for higher image/video storage capacity and data transmission rates led to the search of new bandwidth optimization solutions. Integrated photonics appears as a promising technology to achieve this outcome. Photonic Integrated Circuits (PIC) are the equivalent of Electronic Integrated Circuits (EIC) in the optical domain. As an alternative to transistors and other electronic components, PIC contain optical elements, such as modulators, detectors, attenuators, multiplexers, optical amplifiers and lasers. PIC advantages can be attributed to their lower power consumption, smaller volume and weight, higher thermal and mechanical stability, and the easier assembly of numerous and complex systems. In summary, PIC-based optical communication systems offer an efficient and cost-effective solution to data transmission driving to a significant boost in the segment [2]. An annual growth rate of 25.2% in the PIC market during the period of 2015 to 2022 is foreseen [3]. There is also an increasing demand for PIC driven by innovative applications in biophotonics [2].

PIC can be characterized as a multiport device composed of an integrated system of optical elements embedded onto a single chip using a waveguide architecture [4]. The testing of optical components is more difficult than on electrical components and for an accurate prediction of the PIC performance, an extensive characterization/testing is required [5]. Moreover, optical component testing is difficult and time-consuming, e.g., due to the tight 3D alignment tolerances for accurate coupling of light [5].

Given increasing demand for data transmission and storage, data compression emerges as an important field of study with different available techniques explored to release additional bandwidth. Specifically, for faster image processing, compression methods are fundamental tools to decrease redundant data. Different compression transformation techniques can be used, with the wavelet-based transforms as the most promising ones due to their simplicity and fast computation [6]. All-optical network design appears as a prominent solution for the application of such compression methods. By applying this architecture into a PIC, image compression can be attained with lower cost, less power consumption and high data rate due to an all-optical processing implementation [7]. Among the wavelet-based methods, Haar transform (HT) offers a good approach for image processing and pattern recognition due to its simple design, fast computation power and efficiency, being easily implemented by optical planar interferometry [4] [6] [7]. The HT implementation can be achieved with a two level network of asymmetric coupler devices [4].

The capability of a Spatial Light Modulator (SLM) to dynamically reconfigure the optical wavefronts makes it an attractive technology to excite cores or modes of optical waveguides [8] [9], as it allows the arbitrary addition or

removal of channels by the software and it is anticipated that it can achieve some basic channel equalization. This feature can then be explored to feed/receive optical signal from PIC [1].

SLM is an electronically programmable device that modulates light using an array of reconfigurable pixels [10]. This device can control incident light in amplitude-only, phase-only or a combination of phase-amplitude [10] [11]. One of the most commonly used modulation mechanisms is the electro-optical SLM containing liquid crystals as the modulation material [11] [12]. The liquid crystal spatial light modulators have a microdisplay that is used to collect and modulate the incident light, in a transmissive (liquid crystal display – LCD) or reflective (Liquid Crystal on Silicon – LCoS) form. Another distinguishing characteristic of these modulators is the alignment of the liquid crystal molecules, which is typically either parallel, vertical, or with twisted formation. In combination with appropriate polarizing optics, this determines which properties of the incident light beam can be altered, i.e., phase, amplitude or a combination of the two [11] [12].

Nonetheless, common hologram generation methods cannot arbitrarily modulate the amplitude and phase of a beam simultaneously [13] [14]. It is not then possible to simply address the inverse Fourier transform of the desired pattern into the far-field and replicate the resulting distribution of amplitude and phase directly on the SLM [13]. Thus, it is necessary to apply optimization algorithms to calculate the best hologram possible within the constraints of the device [13].

The SLM based on nematic LCoS technology is an electrically addressed reflection type phase-only SLM in which the liquid crystal is controlled by a direct and accurate voltage and can modulate the wave front of a light beam [11] [15], as shown in Figure 1.

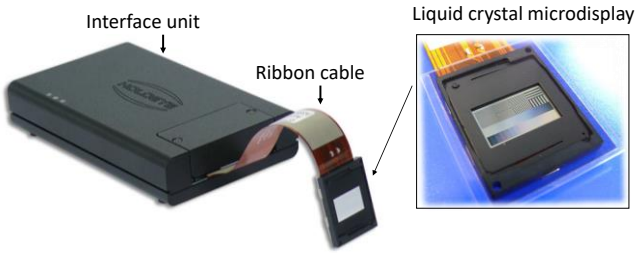


Figure 1. LCOS SLM Pluto phase modulator from Holoeye © 2018 Holoeye Photonics AG.

LCoS SLM is used as a diffractive device to reconstruct images from Computer Generated Holography (CGH) [16]. Appropriate holograms can be generated using a range of different optimization techniques, e.g., linear Fourier transform (i.e., linear phase mask) [17][18], Iterative Fourier Transform Algorithm (IFTA) [19] [20], Gerchberg-Saxton algorithm [21] and simulated annealing [22]. The use of a SLM as a diffractive device to reconstruct images from CGH allows to modulate the wavefront of a light beam.

In this study, we proposed the use of the SLM technology as a flexible coupling platform for feeding photonic integrated processors, i.e., to feed/receive optical signal from a PIC [1]. Furthermore, it can be used as a parallel implementation of the HT image compression algorithm. Preliminary results were obtained to produce an expected CGH to be applied into an optical chip for data compression based on Haar wavelet transform [1].

The paper is organized in four sections. Section II describes the methodology applied for the design of the HT two level network, building blocks, and PIC; the generation of the CGH; and the setup for the flexible SLM coupling platform. Subsection II-A presents the all-optical system architecture for data compression based in HT; Subsection II-B presents the design of the PIC for data compression, addressing the asymmetric coupler and chip design; Subsection II-C presents the generation and optimization of the CGH. Sections III and IV present the obtained results and its discussion, respectively. Section V concludes the study.

II. METHODOLOGY

The methodology is divided into three subsections: (A) the design of an all-optical system architecture for data compression based in Haar wavelet transform; (B) the algorithms used for the generation and optimization of the CGH; and (C) the implementation of the SLM setup to acquire the CGH.

A. *All-optical system architecture for data compression based in HT*

A digital image can be seen as a group of pixels, where neighboring pixels are correlated and usually redundant. Through the decreasing of this redundancy (by compression techniques) the transmission speed and the bandwidth of the system can be optimized. Transforms based on orthogonal functions are the most frequently used in signal compression techniques. The orthogonality is an important property for multi-resolution analysis, where the original signal can be split into low and high frequency components without duplicating information. These functions only require subtractions and additions for their forward and inverse transforms. Examples of these transforms are the Discrete Fourier Transform (DFT), the Discrete Cosine Transform (DCT), and the Discrete Wavelet Transforms (DWT) [23]. DWT have the advantage of representing a fundamental tool for local spectral decomposition and nonstationary signal analysis, used in the JPEG2000 standard as wavelet-based compression algorithms [24]. DWT represent an image as a sum of wavelet functions, with different location and scale [25], i.e., High-pass (detail) and Low-pass (approximate) coefficients. Low-Pass (LP) and High-Pass (HP) filters are applied to the input data with a two level signal decomposition architecture, as depicted in Figure 2.

The Haar wavelet transforms [7] [26] [27] (an example of multiresolution analysis) were chosen due to their simplicity and fast computation.

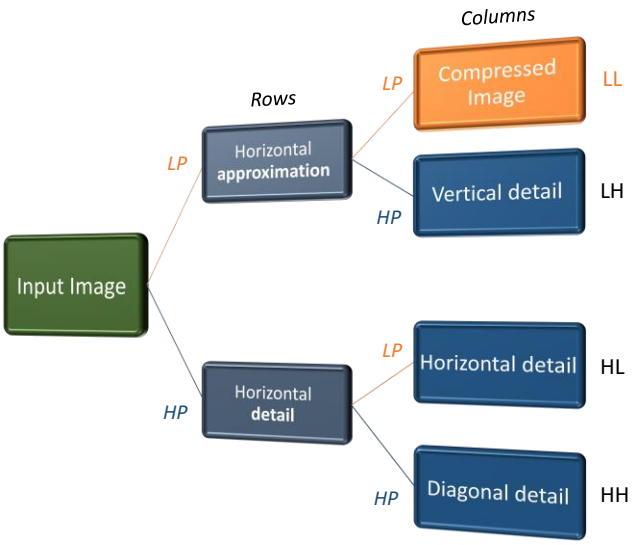


Figure 2. Two level band decomposition using multi-resolution analysis based on wavelet transform. Low-Pass (LP) and High-Pass (HP) filters are applied two times to obtain the 1D transform (L and H components) and the 2D transform (LL, LH, HL and HH components).

The sub-band decomposition achieved through the wavelet transform enables the compression directly on a specific portion of the spectrum, through spatial frequency characterization.

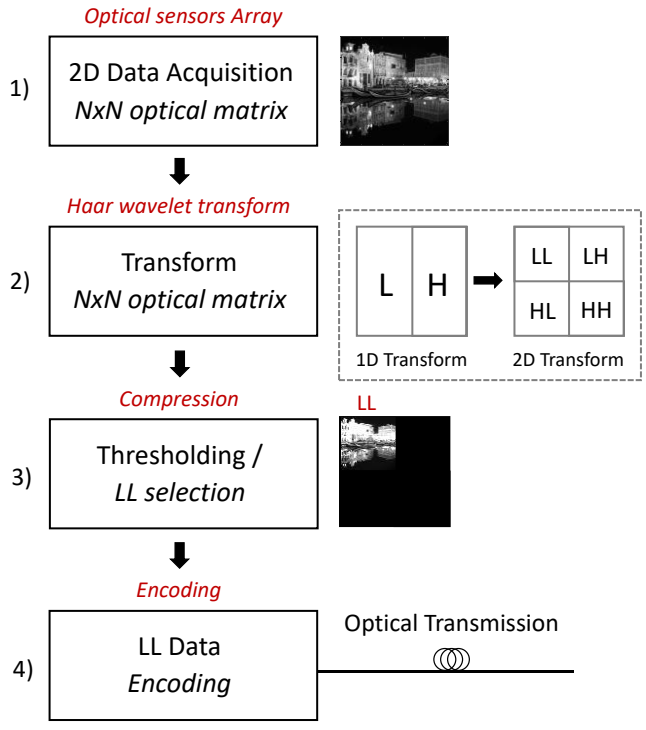


Figure 3. All-optical scheme of system building blocks for Haar wavelet transform processing and compression. 2D transform process schematic describes Low-pass (L) and High-pass (H) filtering through sub-band decomposition [7].

The all-optical system architecture for data compression based in the HT can be divided in four main building blocks: i) optical sensors array; ii) Haar wavelet transform; iii) compression; and iv) data encoding section.

The scheme for all-optical image acquisition, processing, and transmission is depicted in Figure 3.

The first building block entails the acquisition level with optical sensors for light detection and two dimensional (2D) data sampling. The HT is implemented in the second building block, to extract the image properties by exploiting the energy compaction features of the wavelet decomposition.

The HT block (second building block) includes Low-pass (L) and High-pass (H) filters associated with the Haar wavelet, applied over one dimension (1D) at time. The filtering operation can be simplified as the calculation of the average between two neighbors' pixels values (LP) or the difference between them (HP). Equation (1) presents the Haar transform scattering matrix for a generic 1D input (a_i coefficients), i.e., pixel line or column. LP and HP filters are applied two times to obtain the 1D transform (L and H component) and the 2D transform with the four LL, LH, HL and HH components, see Figure 2 and Figure 3.

The coefficients on the left side of (1) are the scaling c_{ij} and detail d_{ij} coefficients (where i refers to the transform level and j to the coefficient index) obtained from the LP and HP filtering, respectively, for each pixel pair, which corresponds to the 1D first level of the Haar discrete wavelet transform. In a 2D matrix input ($N \times N$) this operation is performed twice, i.e., horizontally and vertically, for each transformation level, to guarantee that image intensity variations are evaluated along the two dimensions.

$$\begin{bmatrix} \vdots \\ c_{10} \\ d_{10} \\ c_{11} \\ d_{11} \\ c_{12} \\ d_{12} \\ \vdots \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cdots & & \cdots & & \cdots \\ & 1\ 1\ 0 & & 0\ 0\ 0 & \\ & 1-1\ 0 & & 0\ 0\ 0 & \\ & 0\ 0\ 1 & & 1\ 0\ 0 & \\ \vdots & & & & \vdots \\ & 0\ 0\ 1 & & -1\ 0\ 0 & \\ & 0\ 0\ 0 & & 0\ 1\ 1 & \\ & 0\ 0\ 0 & & 0\ 1-1 & \\ \cdots & & \cdots & & \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ \vdots \end{bmatrix} \quad (1)$$

The same filtering operation is performed in the LL sub-band for the next level of the transform, whereas the other sub-bands (i.e., LH, HL and HH) can be stored, transmitted or discarded, being the transform coefficients related to higher-frequency components.

The third building block carries out the compression and extracts the desirable information from the 2D transform, e.g., LL component. The all-optical system ends with the encoding building block where the data stream is delivered through the optical channel [7].

The optical device chosen to implement the HT was a 3 dB asymmetric coupler, also known as a magic-T, depicted in Figure 4. The asymmetric coupler is characterized by having different waveguides widths, which can present a wide range of coupling ratios and low value of excess loss (0.7 dB),

including input and output single-mode fiber coupling losses [28].

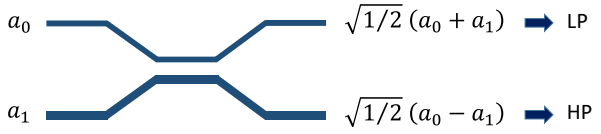


Figure 4. Scheme of a 3 dB asymmetric optical coupler.

To perform the HT operations the asymmetric coupler must be designed in order to perform a 50% coupling ratio.

B. PIC design for data compression

A data compression chip based on Haar wavelet transform was designed in accordance with the rules and using building blocks available from "Application Specific Photonic Integrated Circuit" (ASPIC) foundries [29], as well as proprietary building blocks created and simulated by the authors [4].

The chip was fabricated through a Multi-Project Wafer (MPW) offered by the consortium "Joint European Platform for Indium Phosphide based Photonic Integration of Components and Circuits" (JePPIX) [30].

This platform allows the development of low-cost ASPIC using generic foundry model and it supplies design kits for MPW. The fabrication process was achieved under the program "Photonic Advanced Research and Development for Integrated Generic Manufacturing" (PARADIGM) [31], developed to allow Universities to access to foundry processes. This program reduces the costs of the design, development and manufacture by establishing library-based design combined with technology process flows and design tools.

1) Asymmetric adiabatic coupler

An asymmetric adiabatic coupler in Indium Phosphide (InP) platform, based on adiabatic coupling arrangement was designed using the medium-index-contrast waveguide E600 structure, provided from Fraunhofer Gesellschaft Heinrich Hertz Institute (FhG-HHI) design manual structures [32].

Due to non-disclosure agreement (NDA) of Oclaro and HHI generic foundry processes, further details about the waveguide structure (e.g., structure dimensions and refractive indexes) cannot be provided. The wavelength supported by the developed structure is infrared C-band.

To achieve the phase and coupling ratios necessary for the asymmetric coupler requirements, extensive simulations and fine tuning of all design parameters were performed to attain the right profiles and outputs. Design and propagation analysis was conducted under the Beam Propagation Method (BPM) in OptoDesigner, a tool provided by Phoenix Software [33] [34]. The generic design of the developed InP asymmetric coupler is depicted in Figure 5.

A set of several sections of different sizes was applied in the waveguides design to guarantee the expected coupler behavior.

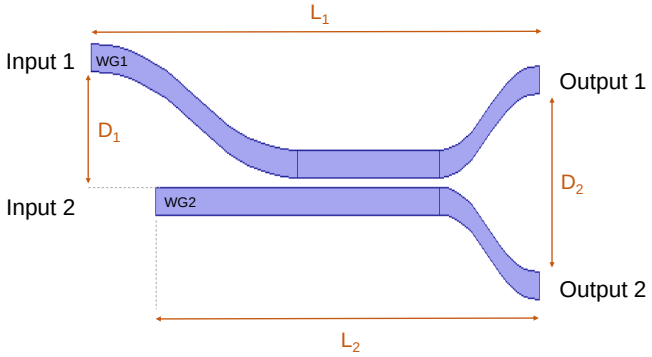


Figure 5. Diagram of the InP asymmetric adiabatic coupler composed by several sections of different sizes. The scheme diagram is not in scale.

A summary of the general dimensions of the coupler is presented in Table I.

TABLE I. GENERAL DIMENSION S OF THE ASYMMETRIC COUPLER

	Coupler dimensions	**(μm)**
D_1	Distance between input WG	40
D_2	Distance between output WG	70
L_1	Length of WG_1	2815
L_2	Length of WG_2	2264

WG: Waveguides. WG_1: Top waveguide from the coupler (waveguide 1). WG_2: Bottom waveguide from the coupler (waveguide 2).

The waveguides were composed by a set of different sections, such as straight, taper, and bend elements [35], as depicted in Figure 6.

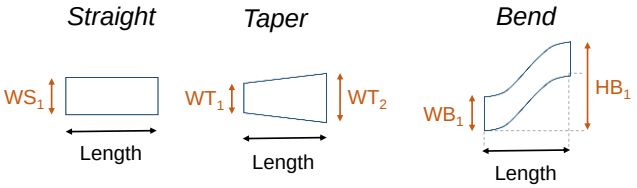


Figure 6. Diagram of three general elements that compose the different sections of the asymmetric coupler waveguide. The *taper* element is also applied in the mirror form, i.e., input as WT_2 and output as WT_1. The scheme diagram is not in scale.

The general dimensions of the elements provided in Figure 6 are presented in Table II.

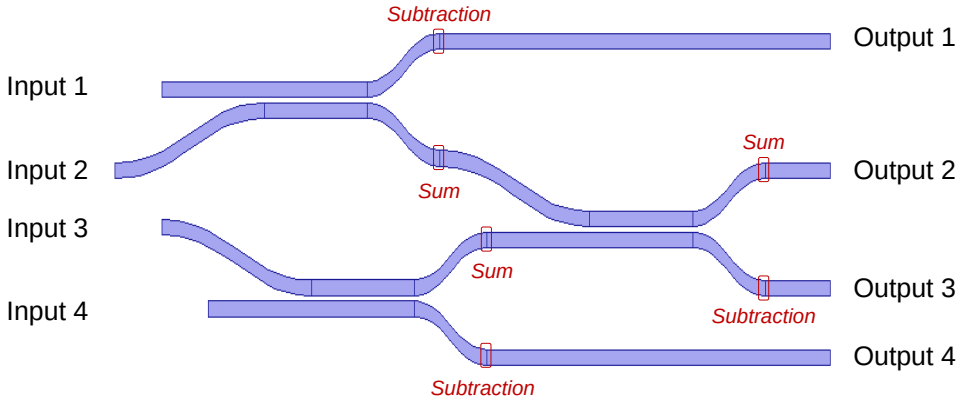


Figure 7. Diagram of the two level network composed by three InP asymmetric adiabatic couplers to perform the expect operations of the Haar wavelet transform.

TABLE II. GENERAL DIMENSIONS OF THE ELEMENTS THAT COMPOSE THE DIFFERENT SECTIONS OF THE ASYMMETRIC COUPLER

	Waveguide elements dimensions	**(μm)**
WS_1	Width of the *straight* element	1.15
WT_1	Input width of the *taper* element for WG_1	1.30
	Input width of the *taper* element for WG_2	1.00
WT_2	Output width of the *taper* element for WG_1 and WG_2	1.15
WB_1	Width of the *bend* element	1.15
HB_1	Height of the *bend* element for WG_1	5.00
	Height of the *bend* element for WG_2	3.60

WG: Waveguides. WG_1: Top waveguide from the coupler.
WG_2: Bottom waveguide from the coupler.

To perform the Haar wavelet transform a two levels network with three asymmetric couplers was designed, as depicted in Figure 7.

2) Chip design

An InP data compression chip to address the Haar wavelet transform was designed [4]. The optical chip is composed by four Distributed Feedback (DFB) lasers (L1-L4), three asymmetric couplers (C1-C3), six PIN photodiodes for network monitoring, two spot size converters, six multimode interferometers (MMI) 1×2 and one MMI 2×2. The e PIC includes one coupler network for compression and another one for decompression. The compression network is composed by the three asymmetric adiabatic couplers, arranged in a two level network, as depicted in Figure 7.

The inputs of the compression network are fed by four DFB lasers.

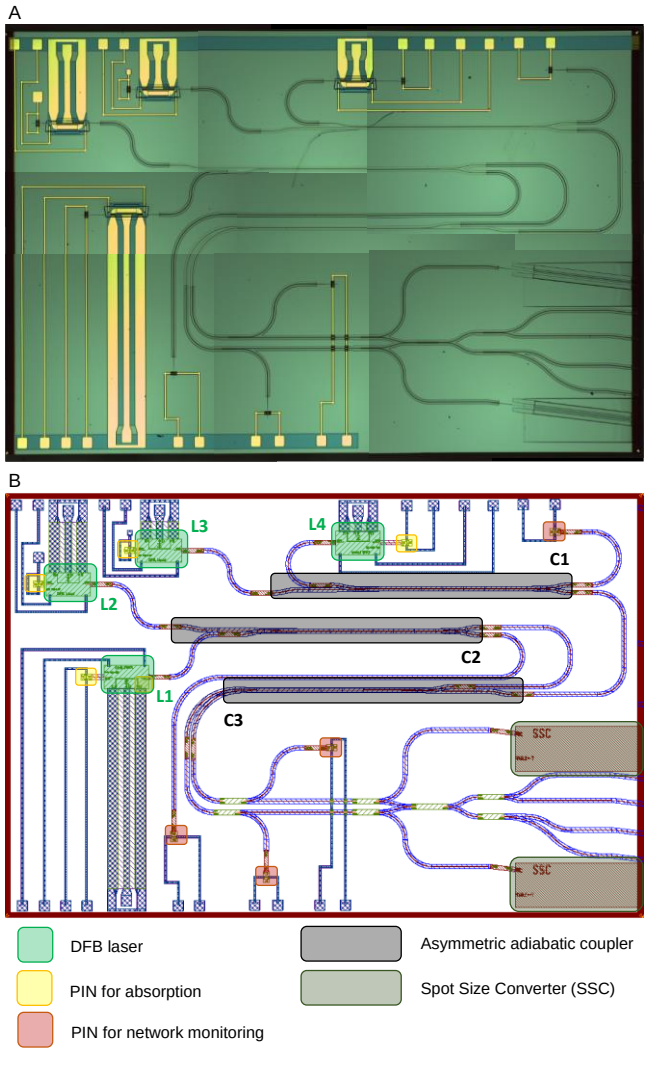


Figure 8. (A): Microscope image of the optical chip (with objective of 5×). (B): Design architecture of optical chip for data compression based on Haar wavelet transform.

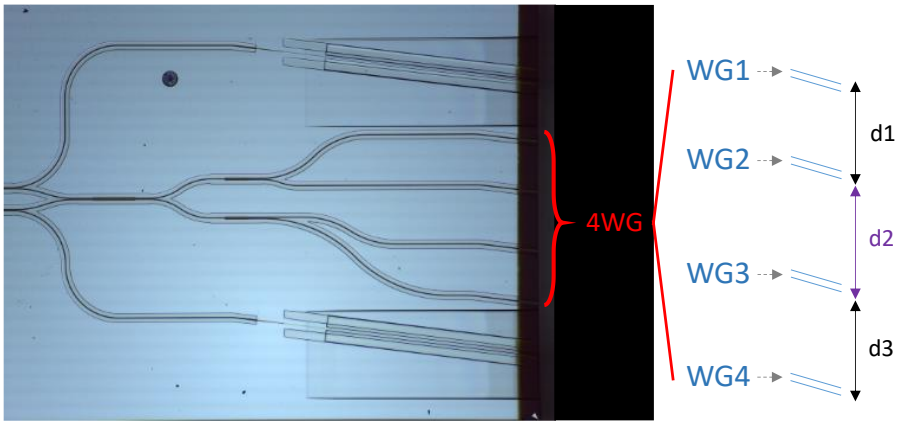


Figure 9. Measurements of the distance between the four waveguides (WG) at the end of the two level compression network.

The outputs are connected to two spot size converters (providing optical output signal) and PIN photodiodes (providing electrical output signal), see Figure 8.

The decompression network is composed by four MMI 1×2 and one MMI 1×1. Four optical outputs are provided, as depicted in the bottom right corner of Figure 8–B. The complete circuit architecture is presented in Figure 8.

The HT operations include Low-pass (L) and High-pass (H) filters applied over one dimension at a time. This filtering operation corresponds to the calculation of the average between two neighbors' pixels values (LP) or the difference between them (HP) [7]. The HT is implemented with a two level network composed by three asymmetric adiabatic couplers (2×2), reproducing the required operations, i.e., the average (sum) and the difference (subtraction) between the optical input pair [4].

The 2D HT can be decomposed in four sub-bands, LL, LH, HL and HH [7]. The LL gives the data compressed. In the chip these four sub-bands can be extrapolated from the four output waveguides (WG) at the end of the three asymmetric couplers network, as depicted in Figure 9.

The measurements of the distance between the four WG at the end of the three asymmetric coupler network are $d_1 = 241.3\ \mu m$, $d_2 = 278.6\ \mu m$, and $d_3 = 248.0\ \mu m$, see Figure 9. Measurements were performed with a Leica microscope (DM 750M; 1CC50 HD) and an objective of 20× (HI Plan EPI, 20×/0.40) [36].

BPM simulations from OptoDesigner of the asymmetric adiabatic coupler and the two level network are provided in *Results* subsection *A*.

C. Generation of the CGH

The CGH is a phase mask or diffractive optical element that can be displayed on an SLM [17].

The information to be transformed (in the Fourier domain) is introduced into the optical system by the SLM, with a phase mask that is appropriate to the input function of interest [37].

The following calculations applied for the generation of the CGH were based in the Fourier optical principles presented in [37].

The CGH was obtained with a linear phase mask calculated in the frequency domain (2), where c_x and c_y are the horizontal and vertical tilt parameters, respectively; and f_x and f_y are the components of the spatial frequency vector corresponding to the image to be generated in the X and Y axis, respectively.

$$M(f_x, f_y) = -2\pi(c_x f_x + c_y f_y) \quad (2)$$

The mask transfer function to be sent to the SLM, is given by $H_{mask} = M\left(f_x, f_y\right) mod 2\pi$, ensuring that the phase values are set in the range of $[-\pi, \pi]$.

A collimated Gaussian beam with transverse profile S_{in} is imaged onto the SLM via a lens. Using the Fraunhofer approximation this produces the Fourier transform at the SLM plane, $fft(S_{in})$. Next, this illumination profile is multiplied with the phase mask, $e^{iH_{mask}}$.

Finally, the result is Fourier transformed by a second lens through an inverse Fourier transform to give S_{out}, the field at the input plane of the PIC.

An estimation of the output signal is given by (3).

$$S_{out} = ifft(H(fft(S_{in})) \quad (3)$$

$$S_{in} = \exp\left(-\left(2\frac{x - x_0}{w_x\, log(\sqrt{2})}\right)^2 - \left(2\frac{y - y_0}{w_y\, log(\sqrt{2})}\right)^2\right) \quad (4)$$

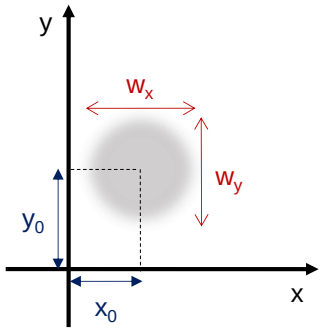


Figure 10. Diagram in Cartesian coordinate system describing the parameters (x_0, y_0) and (w_x, w_y) used for the estimation of the Input beam S_{in}.

S_{in} describes the signal of the input beam (4), where (x_0, y_0) provides the horizontal and vertical position and (w_x, w_y) the width and the height of the beam, respectively, as depicted in Figure 10.

1) *Optimization of the CGH*

To obtain a hologram that replicates the output of the four WG of the optical chip (see Figure 9) was computed a composite hologram of four beams by approximate phase-only superimposition of four independent holograms generated by (2). The correspondent linear transformations in the Fourier domain presented in (5), (6) were applied [1].

$$H = \angle(e^{iH_1} + e^{iH_2} + e^{iH_3} + e^{iH_4}) \quad (5)$$

$$H_1 = \exp\left(i2\pi\left(c_{x1}f_x + c_{y1}f_y\right)\right) \quad (6)$$

A phase-only SLM does not allow to simply address the inverse Fourier of the desired pattern into the far-field and replicate the resulting distribution of amplitude and phase directly on the SLM [13], thus it is challenging to spatially modulate the light with the expected resolution and accuracy.

To overcome this difficulty, an iterative algorithm to obtain the desired hologram with an error factor $\delta \leq 10\%$ was implemented. This threshold was set to avoid an infinite loop in the optimization algorithm, while ensuring an accuracy $\geq 90\%$ in the output result.

The main steps of the algorithm can be described as:

i) generate a 1st linear phase mask to produce the expected initial field based on (5);
ii) initially set the four values a_{1-4} to 1, from $H = \angle(a_1 e^{iH_1} + a_2 e^{iH_2} + a_3 e^{iH_3} + a_4 e^{iH_4})$;
iii) acquire the replay field form the hologram generated by SLM (I_{SLM}) with a camera and feed this data to the algorithm;
iv) calculate the difference between the hologram generated and the initial field expected, defined as error factor: $\delta = |I_{SLM} - I_1| \leq 0.1$;
v) if the condition $\delta \leq 0.1$ is not satisfied repeat steps (ii-iv) by iteratively adjusting the values of a_{1-4} to compensate the error factor.

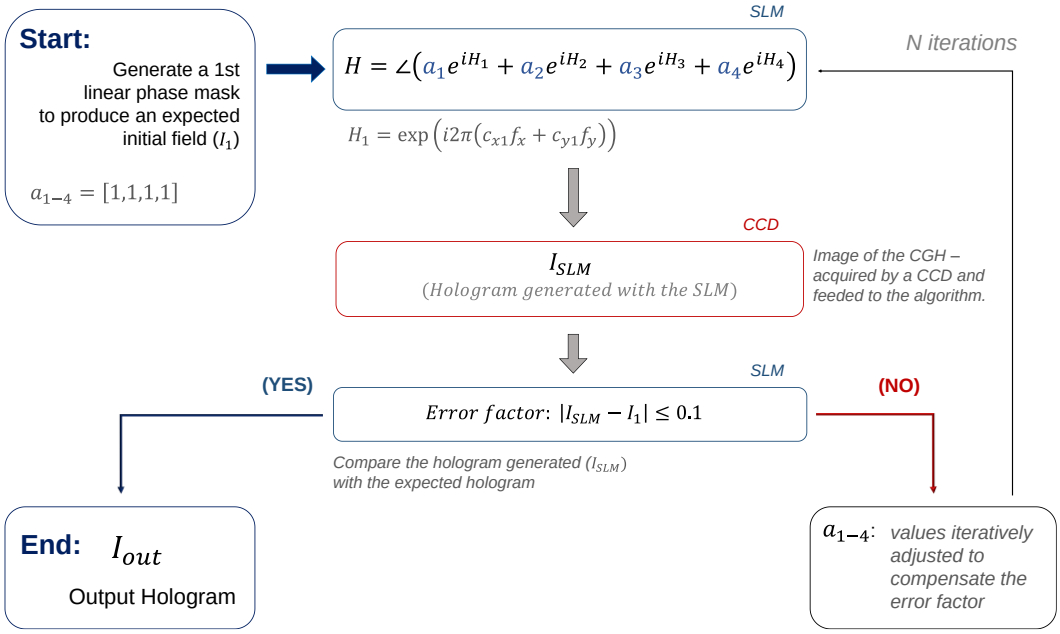


Figure 11. Block diagram of the algorithm applied for the optimization of the CGH.

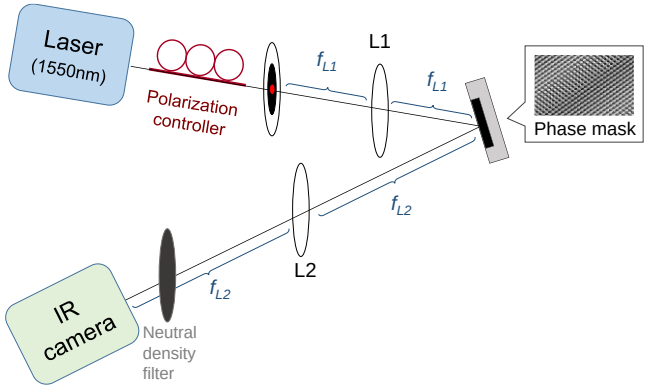
The algorithm developed in Matlab® [38] was able to control both SLM and camera hardware. The block diagram of the algorithm is presented in Figure 11.

The error factor (δ) quantifies the deviation of the generated hologram when compared with the expected output of the optical chip, i.e., the dimensions of the four WG.

D. Setup to generate the CGH

A reflective LCoS phase only SLM, model PLUTO-TELCO-012, with a wavelength range of 1400-1700 nm, an active area of 15.36 mm × 8.64 mm, a pixel pitch of 8.0 µm, a fill factor of 92% and reflectivity of 80% [11] was used to display the hologram.

To remove the phase distortion and have the full Fourier transform scaled by the factor of the focal length (f) the optical system was design based in the $4f$ system configuration, i.e., four distances of length f separating the input plane from the output plane [37]. This forms the basis of a low distortion optical system.



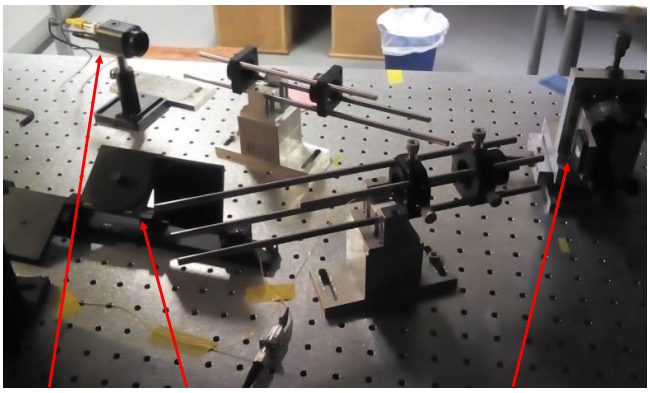


Figure 12. Top figure: Scheme of the hologram reconstruction system, using an infrared (IR) laser of 1550nm, a polarization controller, lens L1, a LCoS-SLM, lens L2 and a IR camera. Bottom figure: Photography of the setup presented.

The setup was composed of: a laser (1550 nm wavelength); a polarization controller; two lenses (AC254-050-C-ML, AR coating 1050-1620 nm) L1 and L2 with a focal length of 75 mm and 250 mm, respectively; a Near-Infrared (IR) (1460-1600 nm) camera (sensing area: 6.4×4.8 mm, resolution: 752×582, pixel size: 8.6×8.3 µm) to capture the hologram produced; and a neutral density filter to avoid saturation in the camera acquisition, see Figure 12.

III. RESULTS

The results section is divided in two subsections: (A) 2D BMP simulation results for the asymmetric adiabatic coupler; and (B) experimental CGH results.

A. BPM simulations

Light propagation simulations of the InP asymmetric adiabatic coupler with input signal in the: i) upper waveguide (WG_1); ii) lower waveguide (WG_2); and iii) same input signal in both WG_1 and WG_2; are presented in Figure 13 and Figure 14.

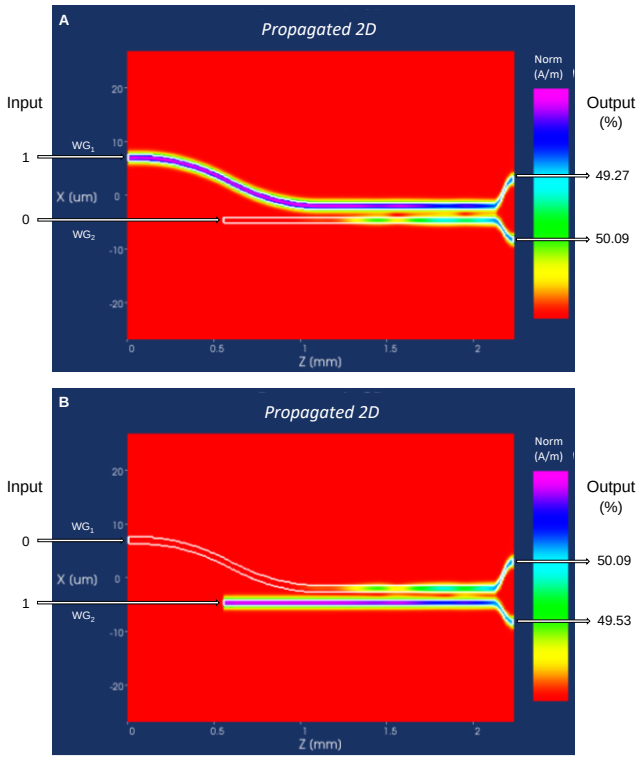


Figure 13. Power propagation in the asymmetric adiabatic coupler when fed with signal on: (A) upper waveguide (WG_1) and (B) lower waveguide (WG_2). Output power values are presented in percentage.

The simulation results demonstrate that the coupler is behaving as expected.

As depicted in Figure 13-A and Figure 13-B, the behavior as a 50% splitter is observed, when only one of the input waveguides carries an optical signal.

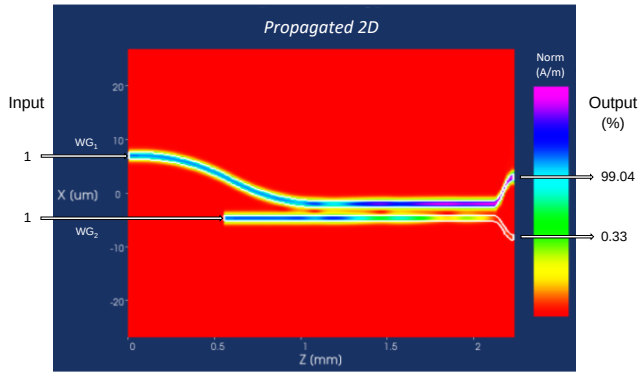


Figure 14. Power propagation in the asymmetric adiabatic coupler when the same input signal is provided on both WG_1 and WG_2. Output power values are presented in percentage.

When both of the input waveguides carry an optical signal, sum and subtraction are achieved at the output waveguides. As can be seen by the duplication of power in the WG_1 (99% of the output signal) and the absence of power in the WG_2 (0.3%), see Figure 14.

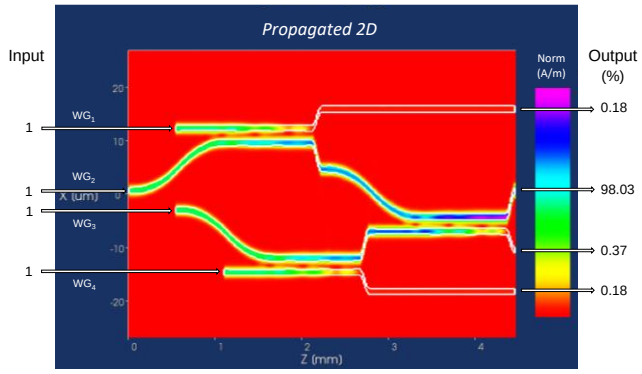


Figure 15. Power propagation for the two level network composed by three asymmetric adiabatic couplers. Output power values are presented in percentage.

The power propagation result for the two level network composed by three InP asymmetric adiabatic couplers is presented in Figure 15. As expected, the HT operations are carried out correctly, which can be confirmed by the power at the four output waveguide ports, i.e., sum at the output WG_2 (98% of the overlap output signal).

B. *Experimental CGH results*

A hologram was generated so as to produce four beam profiles in the first order of diffraction when displayed on the SLM.

Figure 16 presents the image acquired from the replay field of the hologram generated with the initial (I_1) and optimized (I_{out}) CGH.

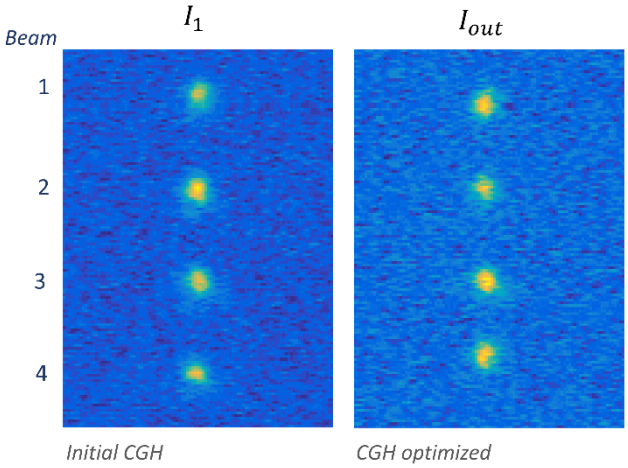


Figure 16. Replay field of the hologram acquired by the IR camera using an: i) initial hologram (left figure), and ii) optimized hologram (right figure).

The analysis of the obtained replay field images can be described by the following steps:

(1) calculate the intensity integration of the image matrix, i.e. sum of all elements along each line of the image matrix, depicted as S_{raw};
(2) application of the Savitzky-Golay (SG) filter to smooth the intensity integration signal obtained in step (1), depicted as S_{SG};
(3) implementation of a first order Gaussian fit curve to the filtered signal, depicted as *Gauss fit*;
(4) extraction of Gaussian parameters to calculate the distances between the four beams (obtained from the CGH) and compare with the expected results (d_1, d_2 and d_3 from the optical chip).

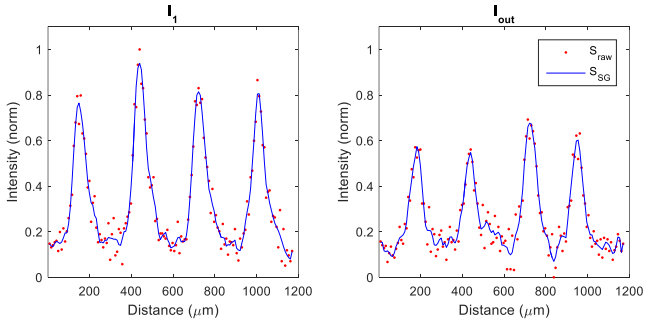


Figure 17. Integrated intensity from the replay field image S_{raw} (red dots), and correspondent smoothing with Savitzky-Golay (SG) filter S_{SG} (blue line). Left: Initial CGH; Right: Optimized CGH.

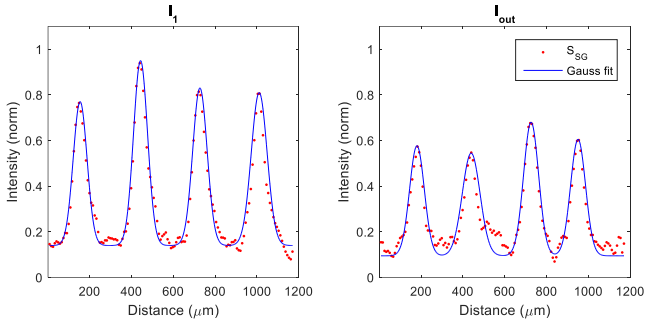


Figure 18. Gaussian fit (*Gauss fit* – blue line) of smoothed integrated intensity signal from the replay field image (S_{SG} – red dots). Left: Initial CGH; Right: Optimized CGH.

The signal smoothing of the intensity integration was obtained with the Savitzky-Golay filter, which can be characterized by a generalized moving average with filter coefficients determined by an unweighted linear least-squares regression and a polynomial model of specified degree [38]. The parameters applied in the filter were a polynomial order 9 and a window length 19.

Results after steps (1) and (2) are depicted in Figure 17, and Gaussian curve fitting application are presented in Figure 18.

The distance between the four beams was calculated from the center position of each beam profile, given by the Gaussian fit coefficient corresponding to the position of the center of the peak. The coefficients were obtained with 95% confidence bounds.

The deviation values (δ) of the generated hologram (i.e., initial I_1 and optimized I_{out} holograms) when compared with the expected output of the optical chip (i.e., d_1, d_2 and d_3 from Figure 9) are presented in Table III.

TABLE III. ERROR FACTOR (δ) VALUES FOR d_1, d_2, AND d_3

	Initial CGH (%)	**Optimized CGH** (%)
δ_{d1}	19.76	7.48
δ_{d2}	1.96	2.90
δ_{d3}	14.31	9.44

An error factor $\delta \leq 20\%$ was obtained for the initial CGH and $\delta \leq 9\%$ for the final optimized CGH.

Power measures of the beams were performed through the integration intensity profiles, i.e., the integral of the Gaussian fit. Table IV presents the integration of the intensity profiles for each beam when applying the initial an optimized CGH. Correspondent mean and standard deviation values of the beam profile for both cases are also provided.

TABLE IV. INTEGRATION OF THE INTENSITY PROFILES FOR THE FOUR BEAMS

Beam	**Initial CGH** (*u.a.*)	**Optimized CGH** (*u.a.*)
1	6.30	5.12
2	8.21	5.78
3	7.18	6.37
4	7.69	5.51
Mean	7.35 ± 0.81	5.69 ± 0.52
Std (%)	11.17	9.14

Integration of the normalized Gaussian fits presented in Figure 18. The four beams are numbered from 1 to 4 from top to down, as depicted in Figure 16. Std: Standard deviation.

A beam mean power loss of 1.1 dB between the initial and the optimized CGH was observed. Nonetheless, a better beam equalization was achieved in the optimized CGH with a standard deviation reduction of 2%.

IV. DISCUSSION

The design of the asymmetric adiabatic coupler and the all-optical network implemented to perform the Haar wavelet transform in InP were demonstrate to operate according to predictions as confirmed by the BPM simulations, supporting their feasibility for compression purposes.

An improvement in the generated hologram is achieved with CGH optimization, i.e., a major reduction of 11% (difference between initial and optimized) in the error factor (δ) was obtained. Nevertheless, the loss of 1.1 dB identified on the mean beam power for the optimized CGH an improved equalization between the beams was observed, with a 2% reduction in the standard deviation.

Algorithm improvements will be addressed to mitigate the power discrepancies between the four beams and optical artefacts associated with the diffraction of light not yet completely eliminated, which can cause a reduction of signal expected at the four output WG of the optical chip.

An alternative approach to correct some of this artefacts would be the implementation of the Gerchberg-Saxton [21] or simulated annealing [22] algorithms, nonetheless due to the power-loss (up to 9dB [13]) associated with these approaches they were not addressed in this study.

The phase mask that replicates the expected output of the optical chip can be used to multiplex/demultiplex the obtained result. Furthermore, a phase mask, which addresses the HT operations can also be applied to invert the compression induced by the HT (optically implemented in the chip with the three asymmetric couplers network).

The use of the SLM coupling platform will allow to provide a proof of concept of the PIC operation.

V. CONCLUSION

An extensive PIC characterization and testing is essential to provide an accurate prediction of its performance. In this study, we proposed a new concept to use the SLM as a flexible platform for feeding photonic integrated processors in order to complement the PIC characterization process. The capacity of the SLM to dynamically reconfigure light allows to feed and/or receive information to the PIC and can be used as a parallel implementation of the HT image compression algorithm. This data can be used to provide a proof of concept of the operation performed by the optical chip, e.g., 2D HT. The design of building blocks for the HT implementation as well as the all-optical network were proposed and simulated, demonstrating their viability for compression. A first result was obtained, i.e., a phase mask that can be used to receive the output of an optical chip for data compression based in the HT.

Further developments will be conducted to provide a more robust SLM based flexible coupling platform, e.g., by improving the optical system components and the implemented phase masks.

ACKNOWLEDGMENT

This work is funded by Fundação para a Ciência e a Tecnologia (FCT) under through national funds under the scholarships PD/BD/105858/2014 and SFRH/BPD/119188/2016; and the project COMPRESS - All-optical data compression – PTDC/EEI-TEL/7163/2014. Additional support was provided by the European Regional Development Fund (FEDER), through the Regional Operational Program of Centre (CENTRO 2020) of the Portugal 2020 framework [Project HeatIT with Nr. 017942 (CENTRO-01-0247-FEDER-017942)], and the COST action CA16220 European Network for High Performance Integrated Microwave Photonics (EUIMWP). The authors acknowledge PICadvanced and Patricia Lopes for their collaboration.

REFERENCES

[1] C. Pinho *et al.*, "Flexible Platform for Feeding Photonic Integrated Processors," in *The Thirteenth Advanced International Conference on Telecommunications (AICT 2017)*, 2017, pp. 1–4.

[2] Grand View Research, "Photonic Integrated Circuit (IC) Market Size Report," 2016.

[3] Credence Research, "Photonic Integrated Circuits Market," 2016.

[4] C. Pinho *et al.*, "Design and Characterization of an Optical Chip for Data Compression based on Haar Wavelet Transform," in *OFC 2017 - Optical Fiber Communication Conference*, 2017, vol. Part F40-O, p. Th2A.9.

[5] M. Smit *et al.*, "An introduction to InP-based generic integration technology," *Semicond. Sci. Technol.*, vol. 29, no. 8, p. 083001, 2014.

[6] V. Ashok, T. Balakumaran, C. Gowrishankar, I. L. A. Vennila, and A. N. Kumar, "The Fast Haar Wavelet Transform for Signal & Image Processing," *Int. J. Comput. Sci. Inf. Secur.*, vol. 7, no. 1, pp. 126–130, 2010.

[7] G. Parca, P. Teixeira, and A. Teixeira, "All-optical image processing and compression based on Haar wavelet transform," *Appl. Opt.*, vol. 52, no. 12, pp. 2932–2939, 2013.

[8] J. Carpenter, S. Leon-saval, B. J. Eggleton, and J. Schröder, "Spatial Light Modulators for Sub-Systems and Characterization in SDM," in *2014 OptoElectronics and Communication Conference and Australian Conference on Optical Fibre Technology*, 2014, pp. 23–24.

[9] H. J. Lee, H. S. Moon, S.-K. Choi, and H. S. Park, "Multi-core fiber interferometer using spatial light modulators for measurement of the inter-core group index differences," *Opt. Express*, vol. 23, no. 10, p. 12555, May 2015.

[10] Meadowlark Optics, "XY Spatial Light Modulator," 2015. [Online]. Available: https://www.meadowlark.com/xy-spatial-light-modulator-p-119. [Accessed: 22-Nov-2017].

[11] Holoeye, "Spatial Light Modulators," *Holoeye Photonics AG*, 2013. [Online]. Available: http://holoeye.com/spatial-light-modulators/. [Accessed: 22-Nov-2017].

[12] Department of Physics, "An Introduction to Spatial Light Modulators," *Stony Brook University*, 2013. [Online]. Available: http://laser.physics.sunysb.edu/~melia/SLM_intro.html. [Accessed: 22-Nov-2017].

[13] J. Carpenter, "Holographic Mode Division Multiplexing in Optical Fibres," University of Cambridge, 2012.

[14] G. Lazarev, A. Hermerschmidt, and S. Kr, "LCOS Spatial Light Modulators : Trends and Applications," *Opt. Imaging Metrol. Adv. Technol.*, pp. 1–29, 2012.

[15] Hamamatsu, "Phase spatial light modulator LCOS-SLM," in *Handbook LCOS-SLM*, 2012, pp. 1–14.

[16] M. Kovachev *et al.*, "Reconstruction of Computer Generated Holograms by Spatial Light Modulators," *Multimedia Content Representation, Classification and Security*, vol. 4105. Springer Berlin Heidelberg, pp. 706–713, 2006.

[17] C. Pinho, A. Shahpari, I. Alimi, M. Lima, and A. Teixeira, "Optical transforms and CGH for SDM systems," in *18th International Conference on Transparent Optical Networks (ICTON 2016)*, 2016, pp. 1–4.

[18] L. B. Lesem, P. M. Hirsch, and J. A. Jordan, "The Kinoform: A New Wavefront Reconstruction Device," *IBM J. Res. Dev.*, vol. 13, no. 2, pp. 150–155, Mar. 1969.

[19] Y. Torii, L. Balladares-Ocana, and J. Martinez-Castro, "An Iterative Fourier Transform Algorithm for digital hologram generation using phase-only information and its implementation in a fixed-point digital signal processor," *Optik (Stuttg).*, vol. 124, no. 22, pp. 5416–5421, 2013.

[20] O. Ripoll, V. Kettunen, and H. P. Herzig, "Review of iterative Fourier-transform algorithms for beam shaping applications," *Opt. Eng.*, vol. 43, no. 11, pp. 2549–2556, 2004.

[21] R. Gerchberg, W. O. Saxton, B. R. W. Gerchberg, and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," *Optik (Stuttg).*, vol. 35, no. 2, pp. 237–246, 1972.

[22] J. Carpenter and T. D. Wilkinson, "Graphics processing unit–accelerated holography by simulated annealing," *Opt. Eng.*, vol. 49, no. 9, pp. 095801-7, 2010.

[23] K. Deb, M. S. Al-Seraj, M. M. Hoque, and M. I. H. Sarkar, "Combined DWT-DCT based digital image watermarking technique for copyright protection," in *7th International Conference on Electrical and Computer Engineering*, 2012, pp. 458–461.

[24] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, 2000.

[25] K. S. Thyagarajan, *Still Image and Video Compression with MATLAB*. Hoboken, NJ, USA, NJ, USA: John Wiley & Sons, Inc., 2011.

[26] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing*. 2014.

[27] J. Kovacevic, V. K. Goyal, and M. Vetterli, *Fourier and Wavelet Signal Processing*, no. October. 2013.

[28] A. Takagi, K. Jinguji, and M. Kawachi, "Design and fabrication of broad-band silica-based optical waveguide couplers with asymmetric structure," *IEEE J. Quantum Electron.*, vol. 28, no. 4, pp. 848–855, Apr. 1992.

[29] G. Gilardi and M. K. Smit, "Generic InP-Based Integration Technology : Present and Prospects," *Progress In Electromagnetics Research*. 2014.

[30] JePPIX, "Joint European Platform for InP-based Photonic Integrated Components and Circuits." [Online]. Available: http://www.jeppix.eu.

[31] PARADIGM, "Photonic Advanced Research and Development for Integrated Generic Manufacturing," *May 29, 2014, PARADIGM and EuroPIC consortia*. [Online]. Available: http://www.paradigm.jeppix.eu/.

[32] JePPIX, "PARADIGM/EuroPIC Design Manual." PARADIGM and EuroPIC consortia, p. 232, 2014.

[33] Phoenix Software, "OptoDesigner 5 - The ultimate Photonic Chip Design environment," 2016. [Online]. Available: http://www.phoenixbv.com/index.php. [Accessed: 07-Feb-2018].

[34] PhoeniX Software, "User Manual OptoDesigner - version 5.0.7." p. 1825, 2015.

[35] PhoeniX BV, "User Manual OptoDesigner PhoeniX Software, version 5.1.4." p. 2913, 2017.

[36] Leica Microsystems, "Leica Application Suite," *Http://Www.Leica-Microsystems.Com/*, 2015. [Online]. Available: http://www.leica-microsystems.com/products/microscope-software/life-sciences/las-easy-and-efficient/. [Accessed: 04-Sep-2016].

[37] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. Stanford: McGraw-Hill Companies, 1996.

[38] The MathWorks, "MATLAB - The language of technical computing." 2015.

The Impact of Regulatory Frameworks on Competition and Penetration of Telecommunication Markets

Analysis of the European and Asian Broadband Markets

Erik Massarczyk, Peter Winzer
Faculty of Design – Computer Science – Media
RheinMain University of Applied Sciences
Wiesbaden, Germany
Email: erik.massarczyk@hs-rm.de, peter.winzer@hs-rm.de

***Abstract*—Based on the rising numbers of broadband Internet users and the resulting higher importance of broadband infrastructures, previous analyses often focused on the relation between competitive market behaviors and the development of customer broadband penetration rates. Additionally, some prognoses also consider the relation between the development of market concentration and customer prices. In both methods, researchers have started to implement some different regulatory variables, which measure the difference between the competition within an infrastructure and between different infrastructures. Here, there is either a simple binary variable (regulation implied: yes or no) or the variable expresses how many connection lines (in relation to the overall market) are affected by regulatory intervention. The target of this and further research will be to expand the current status of knowledge. Besides the analysis of the influence of regulatory frameworks as single (binary) variable on the development of market concentrations, penetration rates and customer prices, two further approaches will be discussed. In the first step, the regulatory variable is changed so that the duration of the implemented regulation is included in this variable. Then, in a second step, regression analyses will examine the relationship between (a) the market and regulatory variables and (b) the broadband connection speed development variables. Chiefly, this paper gives insights about the telecommunication market developments depending on the degree of regulation.**

***Keywords*—broadband development; market concentration; regulatory frameworks; Hirschmann-Herfindahl-Index; Linda-Index; broadband penetration rates.**

I. INTRODUCTION

Previous researches indicate the rising importance of the worldwide Internet and the increased usage of Internet services within broadband infrastructures in daily business and private life [1]. Here, especially the availability of Internet services like (a) cloud computing, (b) video on demand, (c) online telephony, (d) social media networks and (e) email services describes the importance of the Internet nowadays. The availability/implementation of broadband infrastructures and high broadband connection speeds are becoming increasingly important as location factors to guarantee the accessibility of Internet services [2]-[5]. The increased usage and rising importance of broadband infrastructures/connection speeds underline the significance of future communication/data transport for entertainment and work. Therefore, broadband access lines are one of the main indicators for economic growth [2].

In the world and particularly in the following considered European and Asian broadband markets, different standards for the provision of broadband infrastructures subsist [4], which are beside other factors responsible for the various broadband developments in the past years. On this account, in each regional/national telecommunication market different regulatory obligations and technical standards for broadband infrastructures can be observed, which result in different market situations and broadband penetrations [3]-[5]. These differences result by the following reasons: (a) customer broadband demand, (b) prices for broadband services, (c) quality and technologies providing broadband infrastructures (availability of wires and ducts), (d) implementation costs, (e) competition policy and regulatory obligations, (f) competition, and (g) demography and culture [3]-[6].

Most publications on this topic focus on the analysis of the relationship between: (a) regulatory and governmental frameworks, (b) competition, (c) broadband diffusion and adoption, (d) coverage and (e) penetration [7][8]. Furthermore, various papers deal with considerations regarding (a) the relations between implementation costs and customer prices, (b) operators and different broadband infrastructures, and (c) demand and supply of broadband Internet services [9]-[11]. Yet, the development of broadband does not only depend on the customer adoption and diffusion of broadband infrastructures. Broadband developments include all services and benefits, which are targeted to strengthen the following factors: (a) higher broadband coverage and penetration, (b) higher broadband connection speeds, (c) more services, (d) higher technical standard for infrastructures, and (e) measures to create acceptable prices for customers and to induce customer broadband demand. The following relations have been rarely considered in terms of their influence by the competition: (a) the influence of competition (market concentration) on the development of broadband access speeds, and (b) the influence of competition on the development of customer prices for broadband services. In addition, the extent to which the following regulatory provisions influence market factors will be examined: (a) the impact of regulatory obligations and regulatory frameworks on the market concentrations in broadband networks, (b) the impact of regulatory frameworks on the development of broadband penetration rates, and (c)

the influence of regulatory behavior on the customer prices are not considered in detail until the current point of time. As mentioned, the other impacts are not considered. In the past research approaches, regulatory obligations and frameworks are only estimated as binary variables or by number of regulatory used. It examined the influence of different types of competition on the development of broadband penetration rates.

This study will firstly examine the impact of market concentrations on the fixed-line broadband development to prove if the achieved data could present the past research. Next to the consideration of market concentrations, we have also measured the disparity in the broadband markets and have estimated this disparity in the relation to the named broadband developments. Based on this measurement, we have analyzed the different types of regulatory obligations for fixed-line broadband markets and their influence on competition. In the further steps, we have focused on the influences of the aforementioned factors with the focus being on the implementation of regulatory obligations. For the evaluation, we have collected secondary data of fixed-line broadband markets in Europe and Asia to conduct a combined cross-sectional and longitudinal panel data analysis with pooled ordinary least square regressions. The chosen time range of said data will include the years between 2004 and 2015 in order to reflect on the reasons for the different country-specific broadband developments, levels of competition/market concentration and regulatory behaviors over time. Apart from the different regression models, the intensity of competition will be – in a first step – measured through the usage of different economic concentration models. Following this approach, we will discuss how the regulatory frameworks can be examined. The major aim of this study is the analysis and the determination of the country specific different broadband developments.

The paper proceeds as follows: based on the introduction, Section II presents the literature review and the hypotheses. Section III includes the research methodology. Section IV gives the first results of the investigations regarding the development of competition and market penetration. In Section V, we discuss first insides of the regression analysis. After all, we conclude the paper in Section VI.

II. LITERATURE REVIEW

A. Influence of Competition on Market Developments

Due to the various influence factors described, the term of broadband development includes: the development of coverage and penetration of the existing broadband infrastructures, the expansion of new and upgrade of old infrastructures, the customer prices for broadband services and the quality of the broadband networks (broadband connection speeds).

Based on liberalizations of the fixed-line broadband markets in developed and emerging countries, various network operators and service providers compete in the provision of broadband Internet accesses and services. In order to get a quick return for their investments, operators often focus on broadband developments in regions with potentially large customer base and, high population densities and low implementation costs [10][12], which count as economic efficient areas [11]. This approach reduces significantly the incentives for investments, implementations and upgrades of the existing broadband infrastructures in rural regions with lower population density.

However, when competitors get access to the broadband infrastructure of the incumbent or when the competitors have their own broadband access infrastructure (cable or fiber), the customer prices for broadband services, the broadband diffusion and provision respectively are influenced. Especially in cases of providing access for new entrants and controlled prices, regulatory decisions and behaviors by the governmental authorities could strongly influence the existing market situations.

The opening of existing broadband infrastructures creates an intense price competition, which strengthens the broadband adoption by customers [7][8].

In case of competitive situations in broadband markets, the prices for broadband services decrease and the broadband diffusion and provision increase heavily [7][8]. The competition of different network operators and service providers exert a positive influence on customer adoption of broadband access networks and can be named as one of the key drivers to reach high broadband penetration rates [8]. Despite the fact that competition induces more broadband adoption, Gruber and Koutroumpis (2012) have figured out that competition within an infrastructure would be quite more effective for the customer broadband adoption than the competition between different broadband infrastructures [8]. Although the competition between different broadband infrastructures can stimulate stronger market behaviors because of the increased rivalry for market shares, the implementation of further broadband infrastructures is also quite more expensive than the wholesale of an existing broadband infrastructure.

To sum up the previous findings, the first hypothesis will examine the relationship between broadband diffusion and the development of market concentrations.

H1: A stronger competition (higher intensity of competition) leads to higher broadband penetration rates.

In consideration of the available data and the status of the work in progress, we focus on the further considerations of the influences between competition, regulatory frameworks and penetration rates. The estimations of customer broadband prices and broadband connection speeds will be addressed, but not finally concluded.

As mentioned with the first hypothesis, we assume that an existing competitive market structure in broadband markets could positively affect broadband penetration. As additionally addressed by Distaso et al. [7], Gruber and Koutroumpis [8], a higher intensity of competition leads to lower prices for the lease of an unbundled line and lower customer prices for broadband services in general. The

opening of the existing broadband infrastructures by regulatory obligations tackles directly the current market structure, the new market players have to encounter a price competition [9][10][13]. An increased competition with reduced prices could lead to a better acceptance of (broadband) accesses by the customers, because prices are most important driving indicators for the customers´ decision [9][10][13]. Following Distaso et al. and Gruber, there is a negative relationship between customer prices and the adoption of broadband accesses [7][14]. Moreover, Katz & Berry [10] also mention that a weak competition with high market concentrations would induce higher customer broadband prices.

On this account, the market entry is quite difficult for new market entrants, because a business on the boarder of price competition leads to lower sales on the base of constant costs for the use of the infrastructure of the incumbent [6]. Consequently, the incentives to get into the market, to compete with existing market players, and to get only low revenues are quite weak.

However, currently the prices are above the level of marginal costs and thus, new market entrants have the possibility to achieve revenues to stay successfully in the market. The main target is here: *Do customer prices have an impact on broadband developments in regional markets?*

Nonetheless, a former monopolist with an existing infrastructure has the advantage that he gains revenues and do not have the same investment costs like the entrants, because the network usually is already (mostly) depreciated. Therefore, the incumbent in the cases can (a) gain more usable resources, (b) react more flexible on market behavior and (c) longer survive in a price competition.

In competitive market situations, provider decrease their prices to reach a broader customer base [13]. Therefore, the broadband adoption can be positively influenced and will increase over time. The influence of competition on customer revenues may cause problems if the network operators have difficulty to provide the financial resources for new investments in broadband infrastructures. Furthermore, companies try: (a) to differentiate their products and (b) to invest in the broadband infrastructure to get into a better market position than competitors [10]. Generally, it can be ascertained that prices for broadband services and the adoption of accesses are negatively related [7][14]. However, the prices also depend on the customer's willingness to pay and the demand for broadband services. Since customers are price sensitive and their behavior is very price elastic, a declining price induces a higher willingness to adopt and use broadband access [15]. Based on the presented literature background, the following two hypotheses could be developed.

H2: A stronger competition leads to lower monthly customer prices for broadband access.

H3: Lower customer prices relates positively to higher broadband penetration rates.

The hypotheses H2 and H3 figure out how the behavior of the market players regarding the market shares and customer prices influence the broadband development and the adoption of broadband accesses.

However, previous studies do not consider the relationship between (a) competition, broadband and broadband penetration, and (b) available broadband connection speeds [9]. So far, researchers have often considered the relationships between competition and broadband penetration rates and between competition and customer prices. Additionally, some studies have focused on the influence of broadband prices on the development of penetration rates.

Normally, the assumption would be that more competition leads to faster broadband connection speeds, lower prices and higher penetration rates. If this expectation turns out to be true, it can be concluded that in broadband markets with higher concentrations usually strong monopolists and/or oligopolists try to hold and increase their market shares instead of investing into new infrastructures and push further broadband developments. Based on the missing pressure (potential market entry of a new competitor), the incumbent has no incentive to develop a new or better broadband infrastructure.

Only if the (former) monopolist fears a competitor's market entry or the incumbent is forced to grant the network access for new market entrants, it will have an incentive to upgrade the current infrastructure in order to improve the quality of its broadband networks and services [11][12]. Here, with a strong competition it can be assumed that on the one hand, the providers try to find new ways to win customers from competitors, and therefore, they have incentives to invest in new infrastructures [29][30]. On the other hand, if competition is not that strong, the (former) monopolist may (by owning the sole broadband infrastructure) be able to generate better margins and to invest in a better quality of his own infrastructure.

As mentioned before, the focus of this study will be the analysis of the relationships between regulatory behaviors and the market developments in specific countries. Nonetheless, some insides about the relations of broadband connection speeds will be given too.

Since the dependency between (a) market conditions and (b) customer pricing and broadband connection speeds has not been considered so far, it is considered that competition is a key driver for the development of broadband infrastructure and broadband services. It can be expected that a competitive broadband market structure leads to higher connection speeds, since competitors invest financial resources in new infrastructures and equipment in order to differentiate from existing market players and to get in a better market position in comparison to the incumbent.

H4: In regional telecommunication markets with a higher level of broadband competition the average connection speeds are higher.

H5: Lower customer prices for broadband access lead to a faster development of broadband connection speeds.

Actually, the customer willingness to pay does not increase heavily in terms of a rising broadband connection speed [6]. Furthermore, the customer willingness to pay determines the demand for broadband accesses. Because customers are only willing to pay higher prices if there is a substantial improvement of quality and availability of the broadband services [11][16][17].

New entrants usually have to pay access charges if they are willing to use existing infrastructures of other operators [6].

Despite the high significance of the customer part in this topic, the focus will be further in the analysis of the regulatory impact on market developments.

B. *Influence of Regulatory Frameworks on Market Developments*

Following the introduction of the presented competitive considerations, the relationships of the regulatory frameworks on the development of (a) market concentrations, (b) customer prices, (c) penetration rates, and (d) broadband connection speeds need to be analyzed too.

It can be almost confirmed that the huge range of governmental initiatives, involvements and regulatory instruments lead to different market conditions in the considered countries [18]-[24].

For example, the European Union forced the member states to liberalize the fixed telecommunication markets and to open up the past monopolistic state-owned infrastructures between 1985 and 1998. Liberalization should normally strengthen the forces of the market [20][21]. If the market forces are not strong enough to develop the telecommunication markets, the political and regulatory authorities have to intervene [18][19]. Based on the vast range of governmental initiatives and regulatory instruments, it is normally intended that the market regulates itself [18]-[21].

Kiesewetter et al. [22], and Waverman and Koutroumpis [23] found out that regulations (especially access regulations) directly influence the market concentration in broadband markets. Regulations are able to force the incumbent to open the networks for competitors [20]. Which means, the existing market structures and especially the market position of the incumbent can be influenced by the implementation of regulations. In this situation, the regulations shall remove burdens and constraints and may overcome the lack of competitive behavior [8][20][24]. A possible change of market structures allows new entrants to enter the market. However, regulations could only determine the competition within a network. Here, the access regulations are differentiated between regulations for intra-platform competition and service-based competition. The competition between operators with different broadband infrastructures is normally not targeted by regulations [6].

Nonetheless, regulations usually prescribe existing and dominating network operators to open their networks for new entrants [24]. On the one hand, the mandatory access allows competitors to join the broadband markets with only few investments in the provision of broadband services and without any investment in sunk costs assets [6]. On the other hand, the regulatory authority can improve the competitive situation and help to overcome possible competitive deficits [8][24]. Hence, the acceleration of competition should induce a stronger competition with a higher rate of broadband adoptions [7].

H6: Regulatory behavior and mandatory access regulations will positively enhance competitive market behaviors.

Furthermore, regulations also depend on the market power of the incumbent and existing network operators, because they try to avoid or overcome regulations with own behaviors or investments. Incumbents (and big operators) generally would not allow that a new provider could use their networks (without making investments for own broadband infrastructures) [24]. In this case, the regulatory authority has to pay attention that an incumbent (or other big/dominant operator) would not able to offer higher prices to hold the market position, to hinder further market developments and to foreclose other companies to join the market. The regulatory opening of infrastructures for entrants should (a) remove burdens and constraints, (b) allows the creation of retail competition, and (c) ensures that the incumbent cannot foreclosure (new) competitors [20][24].

On the other hand, access regulations for existing infrastructures allow operators to hold their power with their infrastructures and they are able to overview the competitors [7]. One intension of the regulatory authority could be that entrants get a market access and later, when they have had gained enough financial resources, they will invest in own infrastructures (so called “Ladder on Investment”). But, in several countries a couple of enterprises are quite comfortable with the access on an existing network. Consequently, regulations can open the market for further competition, but an infrastructure competition does not necessarily result.

Based on the literature regulatory measures influence competitive market behaviors. Furthermore, it would be necessary to analyze how these regulatory measures affect the development of the coverage and adoption of broadband infrastructures.

Besides the opening of accesses for existing broadband infrastructures by regulations, the offering of grants and subsidies could be regulatory or governmental interventions too. These funds should stimulate operators to invest in further broadband infrastructures and to enhance the quality of existing broadband infrastructures and services. Furthermore, the subsidies should support the operators in their investments to overcome possible investment gaps and to make investments quite reasonable [20][25].

Supporting the previous explanations, Gruber and Koutroumpis [8], and Wallsten [26] mention the fact that the implementation of regulations (especially unbundling) stimulate higher broadband penetration rates. However, Briglauer and Gugler [6] found that only few regulatory decisions influence broadband penetration rates directly. Possibly, regulations can also negatively influence the development of broadband penetration rates [4].

Nevertheless, the assumption here is that governmental interventions want to enhance the broadband penetration and therefore, the following hypothesis indicates a positive relationship between regulatory behaviors and broadband penetration.

H7: Regulatory behavior and mandatory access regulations will positively relate to broadband penetration.

Furthermore, regulatory authorities are able to set price regulations. Therefore, they have to check if the incumbent is trying to misuse his market power to set higher prices than a market with competitive structures. If the incumbent cannot force higher prices, the gained revenues, financial resources and the incentives for further broadband investments will decrease. Also, the new entrants are not willing to invest high amounts, because they would not be able to set higher prices as the incumbent [8][24].

As introduced, regulations are able to offer the opportunity for new market entrants to enter the broadband market. However obviously, the entrants have to pay charges for the usage of existing infrastructures. These fees represent additional costs for the competitors and tend to secure the (dominant) position of (incumbent) network operators [6][7]. Therefore, regulators need to ensure that network access charges are close to marginal costs.

Nonetheless, the entrance of the new competitors normally lead to a stronger competitive market situation which results in lower customer prices [6]. Due to the induction of competition by regulatory measures, it can be hypothesized that regulatory obligations reduces customer prices.

H8: Regulatory behavior and mandatory access regulations reduce to customer prices.

Regulations are able to change previous market structures, especially in the case of non-transitory barriers and a non-existent competition. Due to the implementation of regulations, incumbent could be limited to set higher prices, which normally lead to decreasing revenues. Falling revenues discourage the operators to invest in future infrastructures. Due to the high implementation costs (which are mainly sunk costs), operators and governmental authorities have to take into account the high investment risks [6]. Due to high investment requirements in network infrastructures, entrants have high market entry barriers. In contrast, the resale of broadband services (based on existing broadband infrastructure) is a relatively risk-free alternative for making profits [8][24]. Mandatory access regulations reduce incentives to invest in infrastructure; furthermore, strict cost-based/ex-ante regulatory approaches are suspect and hinder further broadband developments [4][24][25]. However, the providing operator (usually the incumbent) has to be compensated for release of broadband capacities [24].

The literature review does not provide a clear picture of how broadband developments could be supported by regulatory intervention with regulatory commitments and decisions. On the one hand, regulatory interventions reduce investment incentives of companies and operators to develop broadband. On the other hand, the regulatory authorities enable the possibility to enter the market and to offer several funds to support possible new market players that they could develop their own broadband services [6].

H9: Regulatory behavior and mandatory access regulations will positively affect stronger broadband developments and higher broadband connection speeds.

III. METHODOLOGY

As the previous explanations indicate, we will analyze relationships between broadband developments, the respective market concentrations and broadband market regulations in particularly Western European and Southeast Asian markets.

The focus lies on countries of the European Union 28 (EU28) and the Association of Southeast Asian Nations (ASEAN), as well as additional countries such as Switzerland, Norway, Japan, China, Hong Kong, and the Rep. of Korea. The reason why said regions of the world were selected are as follows: (1) EU28 and ASEAN are regions with (a) multiple countries, (b) a comparable number of inhabitants, and (c) national territories. (2) Like the EU28, the ASEAN system is also developing to get in the position of a central commission for economic, social, regulatory and juridical resolutions. The comparison of the countries of the two systems and the additional ones will be presented for period between the years 2004 and 2015. To limit the scope, some of the countries (Laos, Cambodia, Myanmar and Indonesia) in the named two analyses are excluded from the analysis due to the lack of data.

The evaluation of the competitive intensities follows different concentration models, Hirschmann-Herfindahl-Index (HHI) and Linda-Index (LI), which measure the intensity and disparity of the operators in the specific national broadband markets' competition and compare the market shares of the operators [27]-[30].

The HHI, as one of the most popular models to evaluate market concentrations, will be used to measure the intensity of competition based on key figures. In the economic theory, the HHI is signified as the total concentration measure, which analyzes the share of sales in comparison to the total market volume [27][28]. Here, the HHI will be measured with the customer share of one provider in relation to the whole number of the customers in the market. The collected market shares illustrate the number of customers of each of the biggest three providers in relation to the total number of customers in the specific national broadband market [27][28].

The possibly non-observance and non-implementation of all network operators base on the issue that there are some countries with only three network operators, which provide broadband accesses. To generate a comparable base over all countries, we choose to consider only the three biggest network operators for all countries.

The HHI describes the weighted average of concentration and squares the collected market shares (see (1), S describes

the market share of each specific network operator, i describes the considered operator) [28]-[30].

$$HHI = \sum_{i=1}^{m} S_i^2 \times 10.000 = \sum_{i=1}^{m} (100 \times S_i)^2 \quad (1)$$

The HHI follows the subsequent classification in Table I [28]-[30].

TABLE I. BOUNDARIES FOR THE ASSESSMENT OF THE HHI

HHI < 1.000	non concentrated market
HHI = 1.000 - 1.800	moderately concentrated
HHI > 1.800	highly concentrated

The LI does not reach the same usage and awareness level but the results show how much the market varies from perfect competition (LI-value of 1). Generally, the LI is used to examine the disparity between the biggest and following companies. Therefore, the disparity measures an existence of market dominance and describes if the inequalities between the operators lead to significant changes in the competitive behavior [28]. The LI value is based on a two times calculation and presents a double average index (see (2) and (3), CR stands for the Concentration Ratio, which is the single sum of the market shares of the considered number of network operators, i describes the considered operator) [26], which differentiates companies that are relevant to the market because of their size from less relevant companies. In general, the index compares the average market shares of the dominant enterprises in relation to the market shares of the insignificant enterprises. As mentioned before, only the three biggest operators are included in the further competition examinations.

$$V_{i,m} = \frac{\frac{CR_i}{i}}{\frac{CR_m - CR_i}{m-i}} \quad (2)$$

$$L_m = \frac{1}{m-1} \times \sum_{i=1}^{m-1} V_{i,m} \quad (3)$$

Both indicators are good measures to estimate the current existing market power situations in respective broadband markets.

Furthermore, we will only examine the developments in the fixed-line broadband markets, in which smaller network operators is of secondary importance for the competition situation.

As introduced in Section I, more and more people use the Internet and especially the mobile Internet, which is not considered in this study. Nonetheless, the mobile Internet needs also the connection with cable-bound infrastructures (mostly fiber) to deliver the high broadband connection speeds per mobile transmission. Based on this connection and for the reason that the most of the considered countries have already strong implemented fixed-line broadband infrastructures, the treatment of cable-bound broadband infrastructures is further important and present.

For the cross-sectional and longitudinal panel data analysis of the described relationships, we have collected secondary data from: (a) the regulatory authorities of the considered countries, (b) the International Telecommunication Union (ITU), (c) the Organization for Economic Cooperation and Development (OECD), (d) the European Union, (e) the World Bank, (f) telecommunication authorities and ministries, (g) telecommunication providers and suppliers, and (h) national institutions and governments. Due to the different sources, the elicitation of the data can vary. To take all sources into account, average values from all available data are used. Moreover, we test the data validity and reliability with exploratory factor analysis and Cronbach's Alpha to verify the trust in the collected secondary data [32]-[34]. As mentioned in the introduction of this section, for some countries, the data do not exist and therefore, these countries are not considered in detail.

Nevertheless, some discrepancies between the collected data and the anticipated time trend of the data cannot be excluded. It should also be pointed out that the time in the presented model is an important variable with a high influence.

The longitudinal analysis, which spans a time range from 2004 to 2015, will also cover some cross-sectional elements to conduct comparisons between the various countries in consideration. The needed data is composed of the network operators' market shares, introduction of regulations and regulatory frameworks for broadband markets, broadband penetration rates, broadband connection speeds, customer prices and some basic economic facts like Gross Domestic Product (GDP), exchange rates, price parities, households and population densities. The hypotheses will be analyzed and estimated using various econometric and panel data techniques. Generally, each hypothesis will be tested by a pooled ordinary least square regression to figure out if the results are significantly able to present the named relationships. For each hypothesis, we define the following regression equations, which can be seen in Table II. The different stated equations indicate that we try to differentiate the analysis of the equations and the effects between the dependent and independent variables. Furthermore, the application of the different equations for each hypothesis would be necessary to deal with possible autocorrelation and endogeneity effects. The approaches will be utilized to get a broader understanding of the collected data and the possible relationships.

Regarding the consideration of the regulatory measures, the implementation of regulations normally started with the conduct of the liberalization processes, which were applied, e.g., in the European Union in 1998 (and later). The regulations, which are able to limit the market behaviors by obligations, are needed to be investigated. Definitely, current market structures and concentrations are the result of the regulatory decisions of the past. This indicates that further regulatory decisions and regimentations will build up the future market structures [4][6][35].

Nevertheless, it is necessary to define how the regulatory measures will be examined in this study. From the literature, we already know that some of the researches have chosen a simple binary coding if a regulation is implemented or not [7][8][14]. Other approaches have focused on the estimation of regulatory measures regarding on the share of the lines, which are used through regulations, in comparison to the total number of sold lines [4][6][36]. From our point of view, both approaches have still weak points in the consideration.

The binary coding gives a good introduction to deal with, because the directly impact of a regulation on market behaviors can be analyzed. However, if the period between the moment of regulation implementation and the current moment becomes longer, the impact of the sole regulation introduction gets weaker, because the existing operators can estimate quite better the competitors, because the existing operators are better able to assess the competitors who use regulated access to lines. Therefore, the impact of a regulatory action (through the above-mentioned "learning effects") is expected to decrease significantly over time. The regulated network operators could still estimate very well the market and will use their market power.

The second approach, which covers the implementation of the variables, includes the share of the total lines which are reached by regulated obligations. Here, the problem arises that existing network operators can also stipulate access conditions for the access on broadband infrastructures with new entrants/competitors without any regulations. As a matter of principle, it should be noted in the quantitative assessment of regulated access lines that, of course, on a contractual basis (for example, unbundling or bitstream) are possible. Due to these potential problems, we arrange a double analysis approach.

Generally, the binary coding if a regulation is implemented or not is a good introduction, because this approach gives an overview about the several national markets, which regulations work in the market. As we have carried out, the single implementation does not illustrate possibly correct the influence of the regulations over time. Therefore, we consider how long the regulations are implemented in the markets and so, we are able to map, how the regulations work over time. Here, we use as variable for each regulation, the years since the regulation is implemented. However, there is only the problem that in specific broadband markets (e.g., in the Netherlands and Romania) the regulations were implemented in the past and later the regulatory authorities have decided to completely deregulate these markets. For this situation, it is quite difficult to measure the past impact of implemented regulations and therefore, we choose to present the results of the impact of the duration of already implemented regulations.

In this situation, the regulations have no impact in the market anymore. Due to the assumption that regulations affect the market developments over time, the previously implemented regulations continue to have an effect in the market behaviors of the operators. However, when the market is fully deregulated then the variables will be estimated with a zero. Here, we have to acknowledge that this treatment could lead to slightly discrepancies in the analysis. Contrary, we have to do this estimation in this way, because other reflected approaches do not lead to a better result.

Currently, the most studies just differentiate between infrastructure competition (by access regulations) and service-based competition [4][6][7]. Here, the researches often imply the regulations of unbundling and resale as measures. With the considerations of fiber lines, more and more studies also implement the kinds of regulations of fiber and bitstream [4][6][36]. However, line sharing and the virtual unbundled local access (VULA) are not considered in detail. Here, we want to consider all of the different kinds of access regulations, the previous considered and especially the currently unconsidered ones.

Based on the different approach in the assessment of the regulations and the inclusion of different kinds of regulations, our approach will deepen the insight of the impact of regulatory frameworks on the market developments over time.

Due to the status of the work in progress, the analysis of the influences of the regulatory frameworks on the broadband market developments is not completed. Here, we do the analysis of variances (ANOVA) to figure out, how the implementation and non-implementation of regulations affect in average (a) the competition between the network operators, (b) the adoption of broadband accesses, (c) customer prices, and (d) broadband connection speeds.

IV. DESCRIPTIVE RESULTS

A. Competition Analysis

In order to analyze the relationship between competition, broadband connection speeds, customer broadband penetration rates and prices, the intensity of competition (HHI) and the disparity (LI) between the market players will be examined.

For the analysis of the broadband market concentrations, the considered values of the HHI will be separated into the three parts: (1) HHI below the value of 2,000 (low concentration), (2) HHI between the values of 2,000 and 4,000 (moderate concentration), and (3) HHI above the value of 4,000 (high concentration), based on [27]-[30].

Ideally, the fixed-line broadband markets should have stable HHI market concentration values, which do not exceed 1,800 overt time.

Apart from Japan (divided consideration of NTT East and West), all countries with low HHI-values below 2,000 are European countries situated in the continent's Northern or Eastern parts (Lithuania, Denmark, Sweden, UK) (see Figures 1, 3, and 4). These countries are also in the Global top ten of highest average broadband connection speeds [37]-[41].

In general, most fixed-line broadband markets of the EU28 and ASEAN now reach HHI-values between 2,000 and 4,000 and are moderately concentrated.

TABLE II. REGRESSION EQUATIONS

H1: a) $PE_t = \alpha + \beta_1 CI_t + \beta_2 TI_t + \beta_3 PD_t + \beta_4 GDPC_t + \beta_5 HH_t + \beta_6 DM_t + \varepsilon$ b) $TPE_t = \alpha + \beta_1 CI_t + \beta_2 PD_t + \beta_3 GDPC_t + \beta_4 HH_t + \beta_5 DM_t + \varepsilon$ c) $PECHR_t = \alpha + \beta_1 CI_t + \beta_2 TI_t + \beta_3 PD_t + \beta_4 GDPC_t + \beta_5 HH_t + \beta_6 DM_t + \varepsilon$	PE – value of the broadband penetration TPE – trend based value of the broadband penetration PECHR – yearly change rate of broadband penetration CI – values of the competition index (HHI, LI) TCI – trend based values of the competition index SF – monthly subscription fee PD – population density BS – broadband connection speeds TF – termination fees (regulated) GDPC – Gross Domestic Product per Capita RI – regulatory index DM – years of membership in EU28 or ASEAN TI – time variable – capture the influence of time HH – number of households β – changing variable term ε – error term α – constant t – year of consideration
H2: $SF_t = \alpha + \beta_1 CI_t + \beta_2 GDPC_t + \beta_3 TF_t + \beta_4 TI_t + \beta_5 DM_t + \varepsilon$	
H3: a) $PE_t = \alpha + \beta_1 CI_t + \beta_2 TI_t + \beta_3 PD_t + \beta_4 GDPC_t + \beta_5 HH_t + \beta_6 DM_t + \beta_7 SF_t + \varepsilon$ b) $TPE_t = \alpha + \beta_1 CI_t + \beta_2 PD_t + \beta_3 GDPC_t + \beta_4 HH_t + \beta_5 DM_t + \beta_7 SF_t + \varepsilon$ c) $PECHR_t = \alpha + \beta_1 CI_t + \beta_2 TI_t + \beta_3 PD_t + \beta_4 GDPC_t + \beta_5 HH_t + \beta_6 DM_t + \beta_7 SF_t + \varepsilon$	
H4: $BS_t = \alpha + \beta_1 CI_t + \beta_2 PE_t + \beta_3 TI_t + \beta_4 DM_t + \varepsilon$	
H5: $BS_t = \alpha + \beta_1 SF_t + \beta_2 GDPC_t + \beta_3 TI_t + \beta_4 DM_t + \varepsilon$	
H6: a) $CI_t = \alpha + \beta_1 RI_t + \beta_2 TI_t + \beta_3 DM_t + \varepsilon$ b) $TCI_t = \alpha + \beta_1 RI_t + \beta_2 DM_t + \varepsilon$	
H7: a) $PE_t = \alpha + \beta_1 CI_t + \beta_2 TI_t + \beta_3 PD_t + \beta_4 GDPC_t + \beta_5 HH_t + \beta_6 DM_t + \beta_7 RI_t + \varepsilon$ b) $TPE_t = \alpha + \beta_1 CI_t + \beta_2 PD_t + \beta_3 GDPC_t + \beta_4 HH_t + \beta_5 DM_t + \beta_6 RI_t + \varepsilon$ c) $PECHR_t = \alpha + \beta_1 CI_t + \beta_2 TI_t + \beta_3 PD_t + \beta_4 GDPC_t + \beta_5 HH_t + \beta_6 DM_t + \beta_7 RI_t + \varepsilon$	
H8: $SF_t = \alpha + \beta_1 RI_t + \beta_2 GDPC_t + \beta_3 TF_t + \beta_4 TI_t + \beta_5 DM_t + \varepsilon$	
H9: $BS_t = \alpha + \beta_1 RI_t + \beta_2 PE_t + \beta_3 TI_t + \beta_4 DM_t + \varepsilon$	

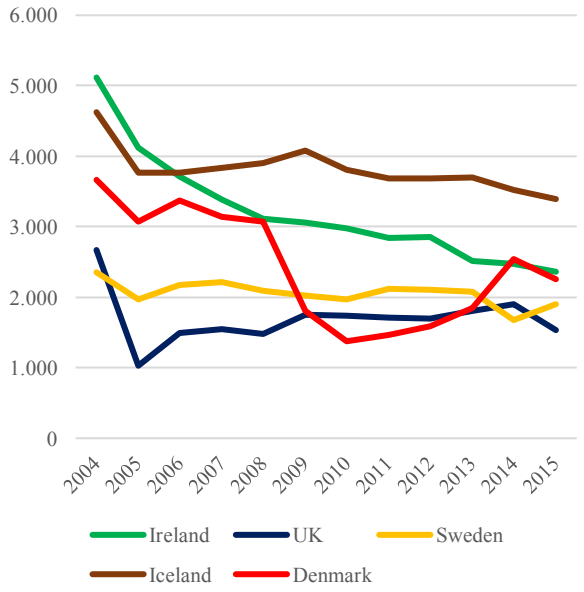


Figure 1. Market concentration of the three biggest fixed broadband network providers in Northern Europe from 2004 to 2015 (x-axis: years; y-axis: HHI values)

When considering the named period, it can be concluded that market concentrations in most countries have decreased from HHI-values above 4,000 (high concentrated) to moderate concentrated market structures.

This development presents diminished market forces and the change of strong monopolistic into rising competitive market structures. Generally, the considered broadband markets are moderately concentrated (e.g., Ireland, Germany, Portugal, South Korea) (see Figures 1, 2, 3, and 4). Nevertheless, some countries (Croatia, India, Philippines) still have HHI-values above 4,000, which implies that the biggest operators were able to hold their market powers and avoid strong competitive structures (see Figures 1, 3, and 4).

Generally, the moderate or high market concentrations in the broadband markets suggest that national regulatory authorities should review the current market behaviors of the existing network operators. To create better competitive and network access opportunities, regulatory authorities could introduce access regulations, which secure possible market entries by competitors.

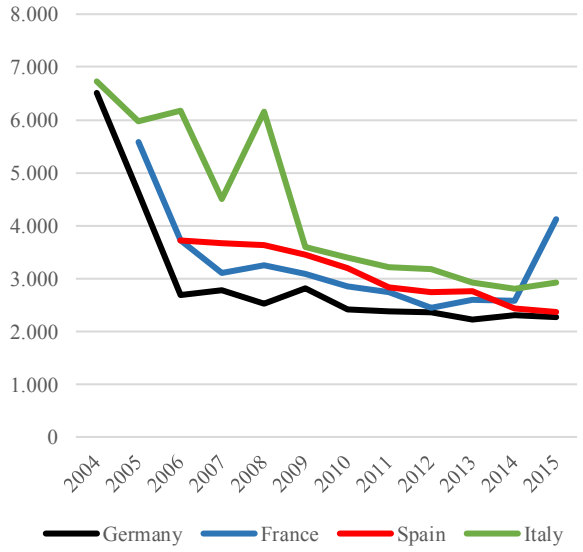


Figure 2. Market concentration of the three biggest fixed broadband network providers in the biggest four Western European countries (except UK) from 2004 to 2015 (x-axis: years; y-axis: HHI values)

Nevertheless, there are two main developments. (1) During the last ten years, the intensity of competition in the most considered broadband markets increased and the previous monopolistic structures could be diminished.

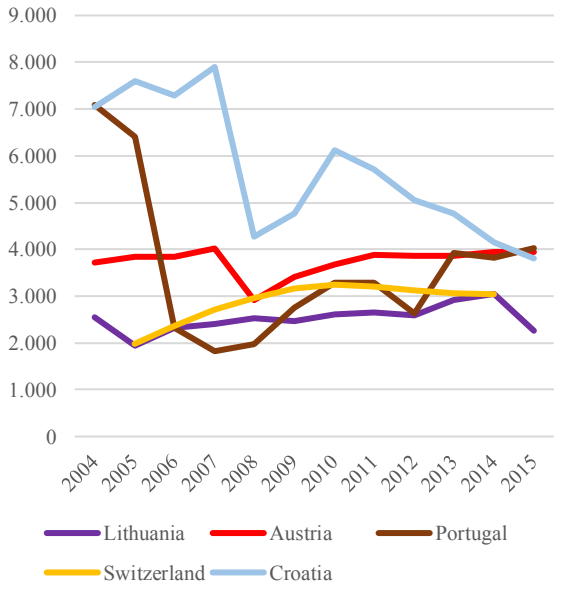


Figure 3. Market concentration of the three biggest fixed broadband network providers of further European countries from 2004 to 2015 (x-axis: years; y-axis: HHI values)

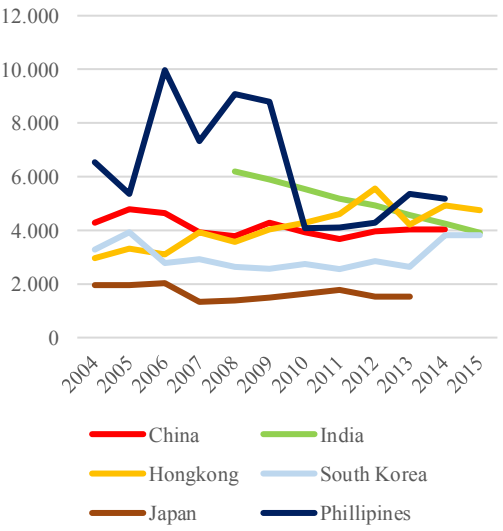


Figure 4. Market concentration of the three biggest fixed broadband network providers of Asian countries from 2004 to 2015 (x-axis: years; y-axis: HHI values)

(2) In the developed countries, the reduction of the power of the monopolistic incumbent is stronger than in the developing countries and the developed countries also have stronger competitive broadband market structures.

The used Linda-Index describes the disparity between the biggest three operators. In general, higher market concentrations translate into higher disparities between the operators. The disparity can be measured in two different ways. On the one hand, the LI examines the discrepancy between the biggest and second biggest companies in the market and on the other hand, the LI can evaluate the discrepancy between the biggest, the second biggest and third biggest companies in the considered market. Based on the evaluation of the three biggest operators in the broadband markets, we will consider the second option with the inclusion of the second and third biggest companies.

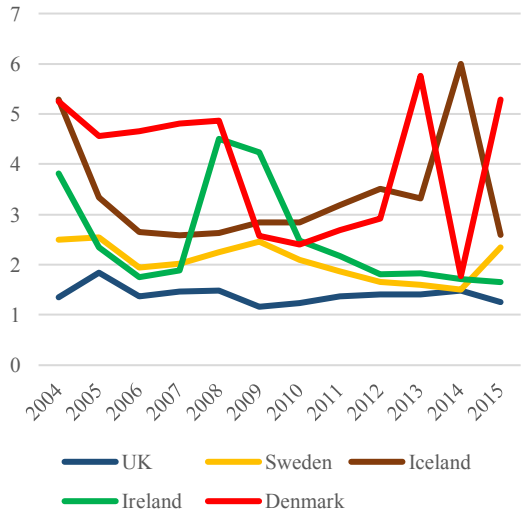


Figure 5. Market concentration of the three biggest fixed broadband network providers in Northern Europe from 2004 to 2015 (x-axis: years; y-axis: LI values)

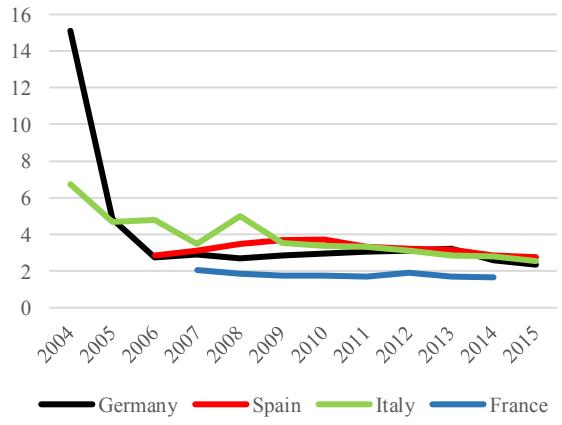


Figure 6. Market concentration of the three biggest fixed broadband network providers in the biggest four Western European countries (except UK) from 2004 to 2015 (x-axis: years; y-axis: LI values)

The consideration of the European and Asian fixed-line broadband markets yields LI-values between 2 and 5 for the most countries (see Figures 5, 6, and 8), which indicates that discrepancies between the operators still exist. Nevertheless, the declining trend of the LI-values shows that in most countries the differences between the incumbents and the new market entrants decrease (e.g., Germany, Italy, Slovenia, see Figures 6 and 7). In the future, these broadband markets could

reach a nearly equal distributed market power. However, the results also show that the disparities between the network operators in some markets increase (e.g., Austria, Switzerland, see Figure 7).

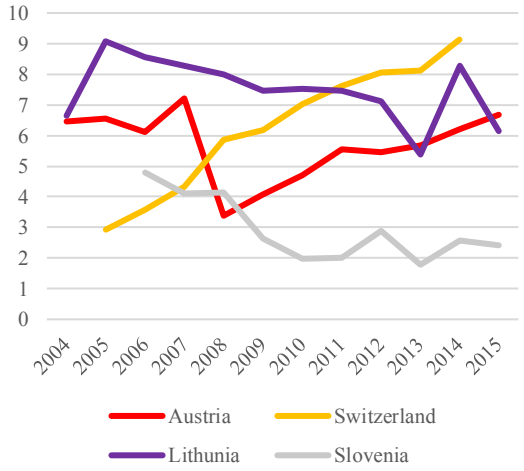


Figure 7. Market concentration of the three biggest fixed broadband network providers of further European countries from 2004 to 2015 (x-axis: years; y-axis: LI values)

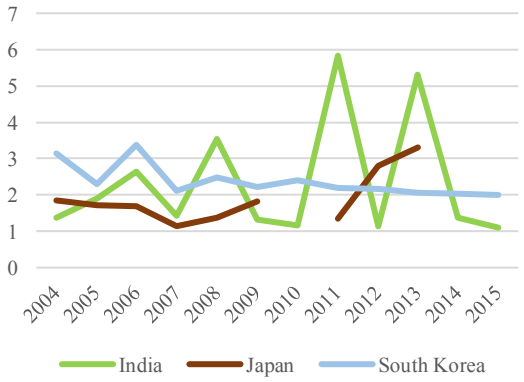


Figure 8. Market concentration of the three biggest fixed broadband network providers of Asian countries from 2004 to 2015 (x-axis: years; y-axis: LI values)

Only in the British market the LI-value is close to 1 and indicates a nearly equal distributed broadband market (between the different market operators, see Figure 5). Combining this result with the fact that the British market has the oldest history of liberalization, it can be concluded that longer open access market could lead to more equally distributed market shares. This issue needs verification by hypothesis testing and we will include this in the evaluations. Furthermore, a couple of countries show nearly the same LI-values over the whole-time frame (e.g., France, South Korea, see Figures 6 and 8). The reasons why, on the one hand, the disparities are very stable and, on the other hand, they vary, will be investigated in the future.

The variations between European and Asian markets are quite low, but nonetheless the LI-values of a couple of countries present higher values. These discrepancies are not sufficiently to draw conclusions from since the results of the LI-values also vary too strongly among network operators in a couple of countries. In general, the disparity (difference in market power and influence) between the incumbent and the competitors cannot be taken as reason for the different broadband connection speeds and developments. It can be just estimated that a more equal distribution of market power could lead to higher broadband connection speeds.

B. Penetration Analysis

Following the statements to the relations of the intensity of competition, we will also consider how the broadband penetration rates have been developed between 2004 to 2015. For the individual national markets, the data were used to provide appropriate average values. Based on these average values, a (unweighted) average was calculated and agglomerated in the overall considerations. [41]-[47] (and several national regulatory authorities).

In Figure 9, the agglomerated average broadband penetration over the considered countries in our model demonstrates that within the time period from 2004 to 2015 the penetration rate increased from 8 to 28 lines/inhabitant. In relative numbers, the score of the penetration in 2015 is 3.5 times higher than the score of the year 2004. The yearly growth of the broadband penetration level per inhabitant for the considered countries is 12.06%. Based on the considerations on the European and Asian countries, the reached broadband penetration rates differs quite heavily. On the one hand, some developed countries like Switzerland, Denmark, the Netherlands and South Korea reach broadband penetration rates per inhabitant between 40% and 50%, where Switzerland is the world leader with over 50% [49]-[51]. On the other hand, for example, the emerging economies of India (1%) and the Philippines (5%) have very low broadband penetration per capita, although it is not entirely due to the poorer economic data compared to the developed economies. Rather, the development of broadband penetration depends on several factors. Some of them, like competitive behavior, regulations and social economic basics (GDPC), we have already introduced in Section II. However, not all factors relevant to the evolution of broadband penetration rates can be fully captured. Also, cultural values and network effects influences the growth of broadband penetration. Furthermore, in the following regression analyses we also find a time effect. The time effect describes the fact that, in addition to the innovators, other market participants also consider the technology to be useful and adapt over time.

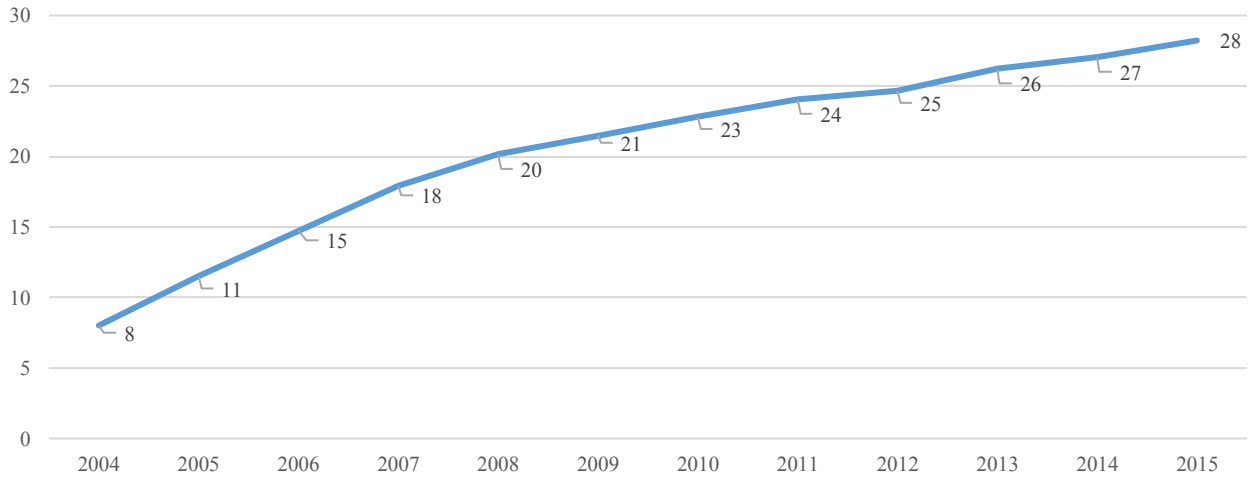


Figure 9. Average development of the broadband penetration of the considered broadband markets from 2004 to 2015 (x-axis: years; y-axis: access lines per 100 inhabitants)

Over time more and more people see the usefulness of this technology and will adopt this technology (in our case the broadband access). Normally, this behavior follows the distribution curve of Rogers [48]. We conclude here that more and more will adopt broadband accesses because they recognize their usefulness. This development we will cover with the so named time effect.

In the following, we will compare the results of the broadband penetration values of the considered countries in our model. Generally, in all countries the rising adoption of broadband accesses per inhabitants can be comprehended. However, in the consideration of Figures 10 to 13, the regional differences need to be addressed. All of the Scandinavian countries and the UK widely reach broadband penetration rates over 30% and therefore, all of them are quite above the calculated mean, which we have already visualized in Figure 9. Therefore, the Scandinavian countries and the UK can be seen as pioneers in broadband penetration. As mentioned above, Denmark is one of the leading providers, with broadband penetration of almost 42% per inhabitant (in 2015). Figure 10 shows that the United Kingdom and Iceland, with a broadband penetration rate of 37% in 2015, are close. In Figure 10, it can be followed that the big European countries like Germany, France, Spain and Italy reach broadband penetration rates between 25% and 40%. In average, these four countries present a quite good status of broadband penetration. Germany (37%) and France (40%) have nearly the same broadband coverage status as seen by the above-mentioned broadband frontrunners Denmark, South Korea and the Netherlands. However, in Spain (27%) and Italy (23%), broadband penetration is below average. For further economic development, it would be necessary for these countries to improve their broadband coverage.

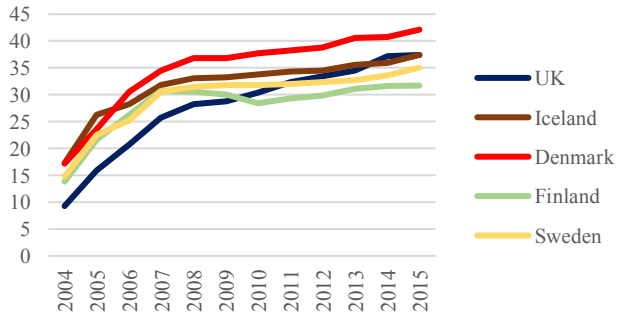


Figure 10. Average development of the broadband penetration rates of the Northern European countries of 2004 to 2015 (x-axis: years y-axis: penetration values per 100 inhabitants)

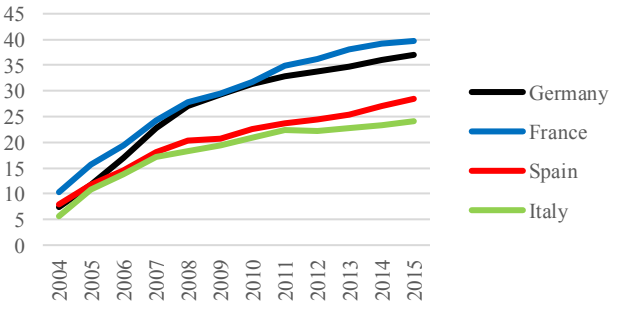


Figure 11. Average development of the broadband penetration rates of the "Big Four" European countries of 2004 to 2015 (x-axis: years y-axis: penetration values per 100 inhabitants)

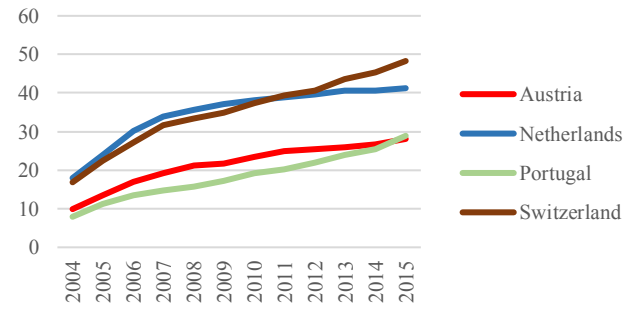


Figure 12. Average development of the broadband penetration rates of the Western European countries of 2004 to 2015 (x-axis: years y-axis: penetration values per 100 inhabitants)

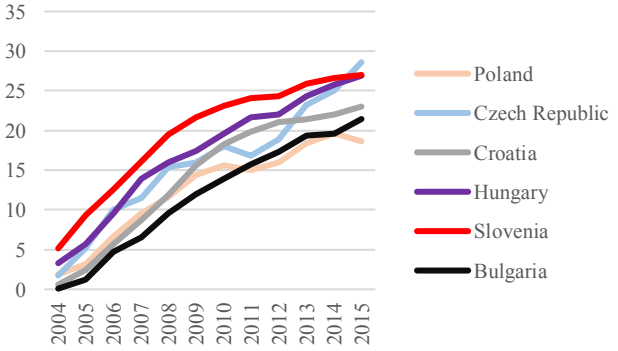


Figure 13. Average development of the broadband penetration rates of the Eastern European countries of 2004 to 2015 (x-axis: years y-axis: penetration values per 100 inhabitants)

Figure 12 illustrates that Switzerland and the Netherlands reach both broadband penetration rates over 40%. From the market review, it can be noted that both countries are counted as the world leaders in broadband penetration per inhabitants [49]-[51]. In 2015, Austria and Portugal, with 28% broadband penetration, have average broadband penetration in the range considered.

The Western and Northern European countries reach at least medium broadband penetration rates and some of them are the broadband penetration forerunners.

Regarding the situation in the Eastern European and the considered Asian countries, the broadband provision and adoption are quite divers. Figure 13 demonstrates that most of the Eastern European countries present a huge growth in the broadband penetration rates. However, only the Czech Republic reach a broadband penetration level quite above the average, which we have presented in Figure 9. Poland is one of the few considered European countries in the analysis, which does not reach a 20% fixed broadband penetration level. In average, the Eastern European countries vary in their broadband penetration between 21% and 28%, which present a lower level of broadband penetration in comparison to the Western and Northern European countries. Poland, with a penetration rate of only 19% (in 2015), is one of the broadband "laggards". Due to the different economic and socioeconomic developments, it can be assumed that the Eastern European countries do not reach the same broadband coverage as the Western and Northern European countries.

Despite that, all of the considered European countries are estimated as developed countries, the differences regarding the broadband penetration are the results from the past developments. It is necessary to mention that the Western and Northern European countries firstly reach the status of developed countries. The Eastern European countries reach this status in later stage of time. The fundamentals of the broadband infrastructure and broadband provision base on the developments and the implementation of the telecommunication networks in the past. Due to Eastern European countries begin on a later stage of the implementation of broadband infrastructures, the differences between the considered European countries can be comprehended.

The development of the broadband penetration rates and broadband networks in the Asian countries depends on the status of the whole country. The developed economies of Japan, South Korea, Singapore and Hong Kong have fairly high broadband penetration rates of between 25% and 40% (see Figure 14), with South Korea leading the pack with a penetration of 40% (in 2015). Overall, these four countries have a broadband penetration similar to Western and Northern European countries.

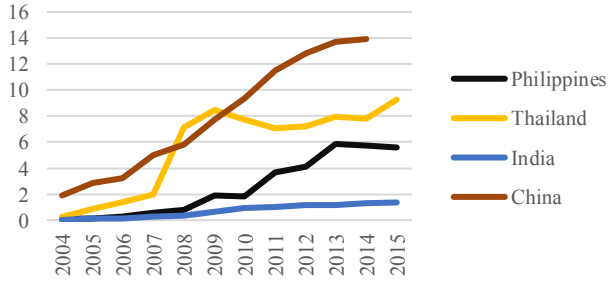


Figure 14. Average development of the broadband penetration rates of the emerging Asian countries of 2004 to 2015 (x-axis: years y-axis: penetration values per 100 inhabitants)

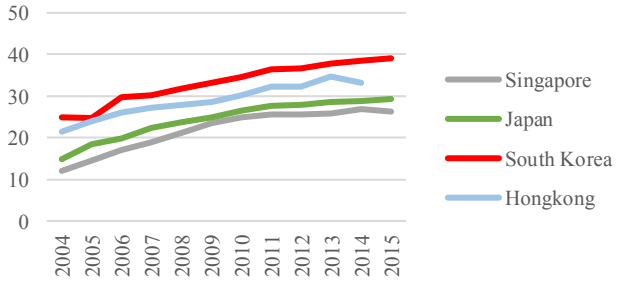


Figure 15. Average development of the broadband penetration rates of the developed Asian countries of 2004 to 2015 (x-axis: years y-axis: penetration values per 100 inhabitants)

However, the emerging countries in Asia do not reach a comparable level of fixed broadband penetration (see Figure 15). The considered countries Thailand, China and India fluctuate between 2% and 15% in terms of their broadband penetration. From the literature, it is already known that the level of broadband penetration depends on the economic situation of the country. We assume that if the named emerging countries would increase in their economic situation, they will also achieve a better broadband provision and better broadband penetration rates. For example, a very positive economic development has been observed in China over the past 10 years. In the same period, broadband penetration in China increased from 2% to nearly 15% (per capita). Looking at the other emerging economies in East and Southeast Asia, none of the countries considered reach the same development as China. At the end of the said 10-year period, only clearly low penetration rates were observed in India (1%), the Philippines (5%) and Thailand (8%). In Thailand, broadband penetration was in part even declining. As mentioned above, most of the recently developed and emerging countries do not have the same status of fixed broadband networks, and therefore supply is lacking. However, most countries are addressing this problem by introducing large and faster mobile broadband networks.

C. Price Analysis

Finally, we conclude the descriptive section with the consideration of the development of customer prices. Here, the data we gain mostly from the ITU [41] and OECD (Broadband Portal) [43]. In Figure 16, the development of the prices symbolizes that the monthly subscription fees decreased from an average fee of 27 Euro per month to nearly 18 Euro (including value added taxes) per month. This means in 2015, the level of the monthly subscription fees was only two thirds in comparison to the level of 2004.

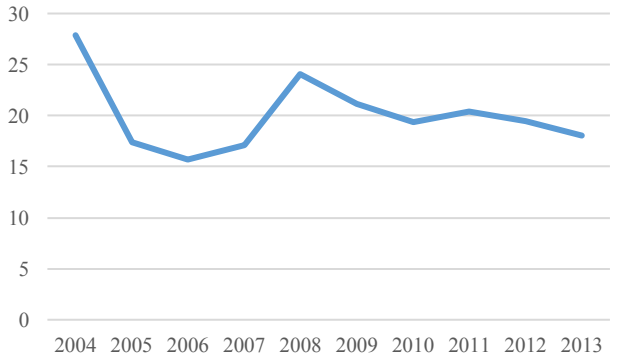


Figure 16. Average development of the monthly subscription fee in Euros from 2004 to 2015 (x-axis: years; y-axis: price values in Euros and price parity cleaned)

Combining the trends of the three considerations, we can conclude that the competition intensified, the penetration raised and customer prices decreased over the considered period. In addition to the achieved research hypotheses, the descriptive results give enough indications, which we will examine in the following section with the regression analyses.

V. REGRESSION ANALYSIS

A. Reliability and Validity Analysis

As introduced above, we proceed with a pooled ordinary least square regression. In the first step, the used concepts of regulations, competition and broadband penetration will be tested if the data can be trusted and if the data is reliable.

The results of the reliability analysis are visualized in Table III. For the further analysis, we use the following concepts: (a) regulations with a binary coding, (b) regulations with a year coding, (c) competition indexes, (d) broadband penetration rates, (e) prices, and (f) broadband connection speeds.

Following Cronbach, the Alpha values higher than 0.7 (0.6) stand for a good (acceptable) reliability [32][52][53]. Based on the results in Table III, 4 of the named concepts reach a good and 2 acceptable reliability.

After the testing of the reliability, the exploratory factor analysis includes the assessment of Kaiser-Meyer-Olkin criterion (KMO), the significance test from Bartlett, and the examination of the cumulative variances to evaluate the validity of the collected data [33][34][54]-[56]. To reach a good validity, the concepts should reach significant p values ($p<0.05$) in the Bartlett-Test and KMO values above 0.7 [33][34][54[56].

Table IV shows KMO-values above 0.6 (except the concept of competition indexes). Following Field [33] and Hair et al. [24], KMO-values above 0.6 describe an acceptable validity. For the concepts of the duration of the regulations, customer prices and broadband connection speeds, the KMO-values are even above 0.7, which indicate a good validity of the collected data/aspects. For the concept of the competition indexes, the validity of the achieved data could not be proved. The good validity scores (except competition indexes) are also supported by the significant results of the Bartlett-Test and the results of the cumulative variances higher than 50%, which indicate high explanation rates of the variances of the collected data [54]-[56]. Mostly, the reliability and validity of the collected data are proved.

B. Analysis of Variances

The consideration of the results of the analysis of variances (ANOVA) presents how the implementation of regulations affects in the mean, the developments of market concentrations, disparity and broadband penetration rates. Furthermore, we also implement a dummy variable if there is any influence if the specific country is a member of the European Union or the Association of Southeast Asian Nations (ASEAN). In the next steps, we present for the regulations displayed in Table V the model fit of the

ANOVA, the significance analysis if there would be a difference in means and how much would be the difference in average.

TABLE III. RELIABILITY ANALYSIS

Research Concepts	Cronbach's Alpha
Regulations (binary coding)	**0.662**
Regulations (duration coding)	**0.774**
Competition Indexes	**0.726**
Penetration Rates	**0.787**
Prices	**0.764**
Broadband Connection Speeds	**0.698**

TABLE IV. VALIDITY ANALYSIS

Research Concepts	KMO	Bartlett-Test	Cumulative Variance
Regulations (binary coding)	0.664	p< 0.000	60.24%
Regulations (duration coding)	0.780	p< 0.000	62.81%
Competition Indexes	0.526	p< 0.000	81.73%
Penetration Rates	0.651	p< 0.000	66.69%
Prices	0.713	p< 0.000	55.06%
Broadband Connection Speeds	0.702	p< 0.000	68.10%

Unbundling

Table V shows that the implementation of unbundling would lead to a significant difference in means of the development of broadband penetration. In comparison with the countries, which do not have implemented unbundling, the implementation of unbundling causes a five times higher broadband penetration per inhabitant in average. The F-Ratio is above the value of 3.87 [33] and therefore, a good model fit and a systematic variation is identified.

Furthermore, unbundling would lead to a significant difference in means in the consideration of the market concentration of the three biggest network operators. In average, countries, which have already implemented the unbundling regulation, have reduced market concentration values by HHI3 = 1,600 in mean. The systematic variation is better described than the unsystematic ones and due to a F-Ratio of 26.72 a good model fit is assumed.

Lastly, the examination of the difference in means in the consideration of the disparity of operators shows that in average, the implementation of the unbundling regulation is not significant. Consequently, there is no significant difference in means and a poor model fit. Due to the not existing difference in means, we possibly assume that unbundling would not affect the disparity of the network operators in a national broadband market.

Line Sharing

In Table V, we can also see that the implementation of line sharing would lead to a significant difference in means of the development of broadband penetration. Contrary to the countries, which do not have implemented line sharing, line sharing has averagely a 1.75 times higher broadband penetration per inhabitant. The F-Ratio is 44.58 and due to a value above 3.87 [33] a good model fit and a systematic variation can be concluded.

Additionally, line sharing is also a cause for the significant difference in means in the examination of the market concentration of the three biggest network operators. An implementation of line sharing would result in a 1,400 lower Herfindahl value. This indicates that the line sharing regulation could be an indicator for decreasing market concentrations. A systematic variation and a F-Ratio of 57.20 describe a good model fit.

The mean analysis of line sharing as influence factor for the development of disparities between the network operators illustrates no significant differences in means. The F-Ratio of 0.051 describes also a bad model fit. Consequently, we predict that line sharing would not possibly affect the changes in the disparity.

Bitstream

The implementation of bitstream lead also to significant differences in means for the development of broadband penetration. In comparison to the previously considered regulations of unbundling and line sharing, bitstream describes a slightly weaker effect, because countries, which have implemented bitstream, have only 1.4 times higher broadband penetration than countries, which did not implement bitstream as regulation for the broadband market. The F-Ratio is quite above the value of 3.87 [33] and therefore, a good model fit and a systematic variation can be concluded.

In comparison to the previous considered regulations of unbundling and line sharing, the results of the competition analyses are quite reversed. The implementation of bitstream do not imply a significance in means. The F-Ratio describes with a value of 0.92 a poor model fit [33]. Hence, we expect that bitstream does not change the relations of market concentrations in the considered broadband markets.

Oppositely, we identify a significant difference in means in the consideration of the implementation of bitstream. However, countries, which have implemented the bitstream regulation, possess in average a 1.2 times higher LI-value than countries, which did not implement this kind of regulation. Thus, bitstream would lead in average to a bigger disparity between the biggest network operator and the two following ones. The mean analysis describes a F-Ratio of 49.15, which covers a good model fit. The systematic variation is better than the unsystematic ones.

TABLE V. RESULTS OF THE ANALYSIS OF VARIANCES

ANOVA	Penetration per Inhabitant	HHI – Market Concentration	LI - Disparity
Unbundling	F-Ratio = 64.68 p < 0.05 good model fit difference in means	F-Ratio = 26.72 p < 0.05 good model fit difference in means	F-Ratio = 3.83 p > 0.05 bad model fit no difference in means
Line Sharing	F-Ratio = 44.58 p < 0.05 good model fit difference in means	F-Ratio = 57.20 p < 0.05 good model fit difference in means	F-Ratio =0.051 p > 0.05 bad model fit no difference in means
Bitstream	F-Ratio = 24.18 p < 0.05 good model fit difference in means	F-Ratio = 0.92 p > 0.05 bad model fit no difference in means	F-Ratio = 49.15 p < 0.05 good model fit difference in means
Resale	F-Ratio = 0.06 p > 0.05 bad model fit no difference in means	F-Ratio = 28.03 p < 0.05 good model fit difference in means	F-Ratio = 0.105 p > 0.05 bad model fit no difference in means
VULA	F-Ratio = 16.11 p < 0.05 good model fit difference in means	F-Ratio = 7.20 p < 0.05 good model fit difference in means	F-Ratio = 0.267 p > 0.05 bad model fit no difference in means
Fiber Regulation	F-Ratio = 17.36 p < 0.05 good model fit difference in means	F-Ratio = 28.03 p < 0.05 good model fit difference in means	F-Ratio = 0.211 p > 0.05 bad model fit no difference in means
Membership	F-Ratio = 0.20 p > 0.05 bad model fit no difference in means	F-Ratio = 4.33 p < 0.05 good model fit difference in means	F-Ratio = 1.034 p > 0.05 bad model fit no difference in means

Resale

Compared to the other regulations, the implementation of resale does not lead significant differences in means of broadband penetration rates. The F-Ratio of 0.06 describes nearly no change in means and the value illustrates a bad model fit [33].

However, the implementation of resale lead in average to a change of market concentration. The ANOVA is significant, which covers a significant difference in means. The value of the F-Ratio is 28.03, which implies a good model fit. National broadband markets, which have implemented resale as regulation, have averagely a 1,000 lower Herfindahl value than the countries, which have not implemented this kind of regulation. Consequently, we assume here that the resale regulation would lead to a reduced market concentration and a higher intensity of competition.

Finally, the implemented resale regulation does not lead to a significant difference in the mean analysis and the F-Ratio is quite poor, which indicates a bad model fit. We expect here for the further analysis that the resale regulation does not lead to significant changes in the development of the disparity between the operators.

VULA

Additionally, there is also a significant difference in means in the consideration of the average broadband penetration per inhabitant. Averagely, the implementation of VULA would lead to 1.38 times higher broadband penetration in comparison to the countries, which refuse the implementation of the VULA regulation. The F-Ratio is above the value of 3.87 [33] and therefore, a good model fit and a systematic variation is identified.

Considering the influence in the developments of the means in market concentration and penetration regarding the implementation of VULA, the means of the regarded market concentrations are significantly lower in the case, when the regulatory authority have decided to implement the VULA regulation. An implementation of VULA results in a 700 lower Herfindahl value. In average, the Herfindahl values are 700 less when VULA is implemented. In comparison to the previously considered significant model fits, the F-Ratio with a value of 7.20 indicates a quite weak significant model fit. Generally, the implementation of the VULA regulation lead to a higher intensity of competition and a lower market concentration.

In consideration of the impact of the implementation of VULA on the development of the disparity between the

network operators, there is no significant difference in means and a poor model fit. Due to the not existing difference in means, we possibly assume that the VULA regulation would not affect the disparity of the network operators in a national broadband market.

Fiber Regulation

Lastly, there will be now considered how would be the impact of possible implemented fiber regulation on the development of the broadband market. Unlike the other regulations, fiber regulation means different kind of access regulations in general and do not itemize a specific regulation. Based on the analysis of variances, we can see that fiber regulation could be also positive for the development of broadband penetration rates. On average, countries, which have implemented several fiber regulations, indicate a 1.28 times higher broadband penetration in comparison to countries, which do not have implemented them. The value of the F-Ratio is 17.356 and illustrates a good model fit.

Also, the implementation of fiber regulation reduces the market concentration in the considered broadband markets on average. There is a significant difference in means that countries, which have implemented fiber regulations, exhibits a 800 lower Herfindahl value than the countries, which refuse to implement fiber regulations. The systematic variation is better described than the unsystematic ones and due to a F-Ratio of 28.03 a good model fit is assumed.

The assessment of the difference in means in the consideration of the disparity of operators averagely shows that the implementation of fiber regulation is not significant. Concluding, there is no significant difference in means and the value of the F-Ratio is below 1 and describes a poor model fit. Due to the not existing difference in means, we possibly assume that fiber regulations do not lead to significant changes in the disparity of the network operators in a national broadband market.

Membership

In the last row, we have implemented the consideration of the influence of membership. We did this approach, because especially the European Union stipulates many regulatory approaches and the most countries follow these specifications. The consideration of the mean analysis regarding the impact of membership on the development of broadband penetration, no significance of the difference in means can be concluded. The value of the F-Ratio with 0.20 is also quite poor. Due to the broadband penetration development depends often on the national conditions, our expectation was here that the membership does not lead to a significant impact.

Contrary, membership leads to a significant difference in means when the market concentration values are considered. In average, countries, which are member in the EU or ASEAN, have a reduced Herfindahl value of 400. The F-Ratio of 4.33 is weakly above the critical bound of 3.87. However, the assumption can be hold and the model fits.

Lastly, membership does not lead to a significant difference in the mean analysis. The F-Ratio in Table V is below the above named bound and therefore, a poor model fit can be concluded. Despite the membership in a community can reduce the market concentration and would lead to more intense competition, the disparity between the operators exist further.

C. Further Approach

The first results of the regression analyses show that the calculated market concentrations correlate significantly (p-values below 0.05 [57]) with the development of the broadband connection speeds. The result supports the assumption that a stronger competition could lead to higher broadband connection speeds.

In addition, the same significant correlations between broadband penetration rates and market concentrations exist (p-values below 0.05 [57]). The correlations imply that higher competitive intensities and stronger competitive behaviors lead to rising broadband penetration rates.

Due to the focus on the impact of the regulatory behaviors on the market developments, we also find significant correlations (p-values below 0.05 [57]) between the single regulations unbundling, line sharing, bitstream, resale, fiber regulation, VULA (binary coding), and the development of market concentrations, broadband penetration rates and prices. Regarding the duration, how long the regulations are implemented in the specific broadband markets, the same significant correlations can be found.

However, the correlations between the regulations and the calculated market concentrations are negative (except for bitstream). Generally, regulations are able to reduce the concentration in a broadband market and increase the intensity of competition. Only the bitstream regulation correlates significantly positive with market concentrations. This indicates that the implementation and persistence of bitstream would strengthen the market power of the biggest network operators. If this relationship does really exist this kind of regulation does not fulfill the target of regulatory intervention. The governmental and regulatory intervention is performed to create better competitive market conditions and entrance possibilities for competitors. If a regulation strengthens the market power of one network operator (often the incumbent), then this kind of regulation should not be implemented.

Concerning the influence of the regulations on the development of broadband penetration rates, all different kinds of regulations support the growth of the broadband adoption and lead to a positive impact on broadband penetration. We will test this possible connection using also the pooled ordinary least square regression approach. The different regulations open various access possibilities to enter the broadband infrastructure, which means, new entrants come into the market with their own customers who may not yet have had access to these broadband structures. As a result, additional customers tend to be connected to the existing broadband infrastructure. In addition, the service providers are intensifying their customer acquisition

measures. Overall, the broadband penetration is thereby increased.

Finally, the implementation and the implemented duration of the regulations correlate significantly positive with the monthly subscription fees. This indicates that regulations would lead to increased prices.

Due to the status of a work in progress, the analyses are in an ongoing status and the results of the regression analyses are not finally completed.

VI. CONCLUSIONS AND FUTURE WORK

As aforementioned, the status of the paper is a work in progress and therefore, improvements in the results and in ongoing research will be necessary. Currently, we have collected the secondary data and have started to analyze the competitive intensities, broadband penetration rates and customer prices. Furthermore, we tested the reliability and validity of the collected data and we examined the data in the analysis of variances (ANOVA). Following this first overview, we will evaluate the above-mentioned hypotheses using the pooled ordinary least square regressions to test the established regression equations.

Despite the named conditions and the different developments in the national broadband markets, the general trend presents increasing competitive structures in fixed broadband markets. Combining the results of the HHI and LI analysis, the incumbents in each national broadband market have lost market shares and the disparity between the different providers is decreasing. As shown in the results, few countries (especially in Asia) still have very powerful incumbents and a general statement concerning all considered countries cannot be done at this status of work.

We have indicated that in all considered broadband markets, the fixed broadband penetration and adoption increase.

Finally, overall the customer prices decrease from 2004 to 2015. However, the examination over the recent years (2015 to 2017) shows that the monthly subscription fees are quite on a stable level.

At this time in evaluation work, the results are on an advanced but not final stage. For the concluding remarks in this topic, the ongoing research has to be deepened.

REFERENCES

[1] E. Massarczyk and P. Winzer, "The Impact of Regulatory Frameworks and Obligations on Telecommunication Market Developments – Analysis of the European and Asian Broadband Markets and Regulatory Frameworks", In K. Daimi & S. Semenov (Eds.), The Thirteenth Advanced International Conference on Telecommunications (AICT 2017, IARIA) [25. - 29. June 2017, Venice]. Conference Proceedings and Thinkmind Library, pp. 56-64 (ISSN: 2308-4030, ISBN: 978-1-61208-562-3)

[2] P. Koutroumpis, "The economic impact of broadband on growth: A simultaneous approach", Telecommunications Policy, Volume 33 (9), pp. 471-485, 2009.

[3] International Telecommunication Union, "The state of broadband 2014: broadband for all", Report from the broadband commission, pp. 16-23, 2014. (http://www.broadbandcommission.org/Documents/reports/bb-annualreport2014.pdf), [retrieved: 02.2018]

[4] W. Briglauer, "The impact of regulation and competition on the adoption of fiber-based broadband services: recent evidence from the European Union member states", Springer Verlag, pp. 450-468, 2014.

[5] Monopoly Commission, "Special Report 61 – Telcommunication 2011: Strengthen investments and secure the competition", in German: Monopolkommission "Sondergutachten 61 – Telekommunikation 2011: Investitionsanreize stärken, Wettbewerb sichern", pp. 24, 40-41, 55, 76-86, 2011. (http://www.monopolkommission.de/sg_61/s61_volltext.pdf), [retrieved: 02.2018]

[6] W. Briglauer and K. Gugler, "The deployment and penetration of high-speed fiber networks and services: Why are EU member states lagging behind?", Telecommunications Policy, Volume 37, pp. 819-835, 2013.

[7] W. Distaso, P. Lupi, and F. M. Maneti, "Platform competition and broadband uptake: Theory and Empirical evidence from the European Union", Information Economics and Policy, Volume 18 (1), pp. 87-106, 2006.

[8] H. Gruber and P. Koutroumpis, "Competition enhancing regulation and diffusion of innovation: the case of broadband networks", Springer Science + Business Media, New York, Volume 43 (2), pp. 168-195, 2013.

[9] R. L. Katz, "The present and future of the telecommunication in Costa Rica", in Spanish: "El presente y futuro de las telecommunicaciones de Costa Rica", 4ta Expo-Telecom Costa Rica – Telecom Advisory Services, LLC, pp. 14, 2011.

[10] R. L. Katz and T. A. Berry, "Driving demand for broadband networks and services, signals and communication technology", Springer Verlag, pp. 5-40, 135-200, 2014.

[11] U. Stopka, R. Pessier, and S. Flößel, "Broadband study 2030 – Prospective services, broadband adoption and demand", in German: "Breitbandstudie 2030 – Zukünftige Dienste, Adoptionsprozesse und Bandbreitenbedarf", pp. 42-50, 60, 166-164, 2013.

[12] T. Tjelta et al., "Research topics and initial results for the fifth generation (5G) mobile network", 1st International Conference on 5G Ubiquitous Connectivity (5GU), pp. 267-272, 2014.

[13] R. L. Katz and F. Callorda, "Mobile broadband at the bottom of the pyramid in Latin America", Telecom Advisory Services, LLC, pp. 23-25, 2013.

[14] H. Gruber, "European sector regulation and investment incentives: European options for NGA deployment", In I. Spiecker and J. Krämer (Eds.), Network neutrality and open access Baden-Baden: Nomos, pp. 191-202, 2011.

[15] C. D. Piros and J. E. Pinto, "Economics for investment decision making: Micro, Macro, International Economics". CFA Institute – Investment Series, John Wiley & Sons Inc., New Jersey, pp. 41-51.

[16] P. Winzer and E. Massarczyk, „Obstacles of fiber roll-out", in German: "Viele Faktoren – Was hemmt den Glasfaserausbau in Deutschland?" In: NET, 2015, 69. Jg., H. 3, p. 38-41

[17] P. Winzer and E. Massarczyk, "How Does Improving the Existing DSL Infrastructure Influence the Expansion of Fiber Technology?" Paper presented at the 17. International Conference on Broadband Communications, Networks, and Systems (ICBCNS) [28. - 29. June 2015, London], Conference Proceedings (eISSN: 1307-6892), p. 3934-3941

[18] Bundesnetzagentur, "Annual Report 2013 – Strong networks – consumer protection", in German: "Jahresbericht 2013 – Starke Netze im Fokus – Verbraucherschutz im Blick", pp. 70-81, 2013. (http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Bundesnetzagentur/Publikationen/Berichte/20

14/140506Jahresbericht2013Barrierefrei.pdf?__blob=publicationFile&v=4), [retrieved: 05/2017]

[19] Bundesnetzagentur, "Definition of market regulation", in German: "Definition von Marktregulierung", 2013. (http://www.bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Marktregulierung/marktregulierung-node.html) [retrieved: 05/2017]

[20] I. Cava-Ferreruela and A. Alabau-Munoz, "Evolution of the European broadband policy: Analysis and perspective", pp. 1-17, 2005.

[21] W. Kerber, "Competition Policy – Vahlens compendium for economic theory and economic policy", in German: "Wettbewerbspolitik. Vahlens Kompendium der Wirtschaftstheorie und Wirtschaftspolitik", Volume 2 (8), p. 302, 2003.

[22] W. Kiesewetter, L. Nett, and U. Stumpf, "Regulation and competition in European mobile telecommunication markets", in German "Regulierung und Wettbewerb auf europäischen Mobilfunkmärkten", WIK – Wissenschaftliches Institut für Kommunikationsdienste, 2002.

[23] L. Waverman and P. Koutroumpis, "Benchmarking telecoms regulation – The Telecommunications Regulatory Governance Index (TRGI)", Elsevier – Telecommunications Policy, Volume 35, pp. 450-468, 2011.

[24] J. Bouckaert, T. van Dijk, and F. Verboven, "Access regulation, competition, and broadband penetration: An international study", Elsevier – Telecommunications Policy, Volume 34, pp. 661-671, 2010.

[25] O. Falck, J. Haucap, and J. Kühling, "Economic growth oriented telecommunication policy: intentions and options", in German: "Wachstumsorientierte Telekommunikationspolitik Handlungsbedarf und -optionen", Studie im Auftrag des Bundesministeriums für Wirtschaft und Technologie, 2013. (http://www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/Studien/wachstumsorientierte-telekommunikationspolitik-handlungsbedarf-und-optionen,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf) (retrieved 02.2018)

[26] S. Wallsten, "Broadband and unbundling regulations in OECD countries", AEI-Brookings Joint Center Working Paper No. 06-16, pp. 1-28, 2006.

[27] T. Apolte et al., "Vahlens compendium for economic theory", in German: "Vahlens Kompendium der Wirtschaftstheorie und Wirtschaftspolitik", Verlag Franz Vahlen, Volume 9 (2), pp. 404-411, 2007.

[28] I. Schmidt, "Competition Policy and law", in German: "Wettbewerbspolitik und Kartellrecht", Volume 7, Stuttgart, pp. 49-55, 2001.

[29] M. Motta, "Competition policy – theory and practice", Cambridge University Press, Cambridge, United Kingdom, 2004.

[30] W. K. Viscusi, J. E. Harrington Jr., and J. M. Vernon, "Economics of regulation and antitrust", MIT Press, Volume 4, Cambridge, Massachusetts, pp. 155-162, 2005.

[31] S. Bicheno, "South Korea to add fourth mobile operator", Telecoms, 2015. (http://telecoms.com/423611/south-korea-to-add-fourth-mobile-operator/), [retrieved: 02.2018]

[32] L. J. Cronbach, "Coefficient Alpha and the internal structure of tests. Psychometrika, Volume 16, pp. 297-334, 1951.

[33] A. Field, "Discovering statistics using SPSS", Sage Publications Ltd., Volume 4, 2013.

[34] J. F. J. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, "Multivariate data analysis", Macmillan, New York, NY, Macmillan, Volume 3, 1995.

[35] W. Hugentobler, "Liberalization of the telecommunciation market", in German: „Liberalisierung in der Telekommunikation", 1999. (http://www.bundesnetzagentur.de/SharedDocs/Downloads/EN/BNetzA/PressSection/ReportsPublications/AeltereDaten/TKLiberalisationId2053pdf.pdf?__blob=publicationFile) 13/09/2016 (retrieved 02.2018)

[36] T. Ovington, R. Smith, J. Santamaría, and L. Stammati, "The impact of intra-platform competition on broadband penetration", Telecommunications Policy 41 (2017), pp. 185-196.

[37] D. Belson, "Akamai's State of the Internet", Akamai Technologies Q1 2012 Report, Volume 8 (1), pp. 5-32, 2012. .

[38] D. Belson, "Akamai's State of the Internet", Akamai Technologies Q1 2013 Report, Volume 8 (1), pp. 5-32, 2013.

[39] D. Belson, "Akamai's State of the Internet", Akamai Technologies Q1 2014 Report, Volume 8 (1), pp. 5-32, 2014.

[40] D. Belson, "Akamai's State of the Internet", Akamai Technologies Q3 2015 Report, Volume 8 (1), pp. 5-32, 2015.

[41] International Telecommunication Union, "Yearbook of Statistics 2014 – Telecommunication/ICT Indicators 2004-2013", 2014.

[42] World Bank, Data Bank World Development Indicators to Price Parities, Gross Domestic Product per Capita, Internet and telecommunication indicators. (http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators) (retrieved 02.2018)

[43] OECD, Broadband Portal – broadband penetration rates, broadband statistics. (http://www.oecd.org/sti/broadband/) (retrieved 02.2018)

[44] International Telecommunication Union, "The state of broadband 2015: broadband as a foundation for sustainable development", Report from the broadband commission, pp. 12-21, 39-44, 2015. (http://www.broadbandcommission.org/Documents/reports/bb-annualreport2015.pdf), [retrieved: 02.2018]

[45] International Telecommunication Union, "The state of broadband 2012: achieving digital inclusion for all", Report from the broadband commission, pp. 16-32, 2016. (http://www.broadbandcommission.org/Documents/reports/bb-annualreport2012.pdf), [retrieved: 02.2018]

[46] International Telecommunication Union, "The state of broadband 2016: broadband catalyzing sustainable development", Report from the broadband commission, pp. 16-32, 2016. (http://www.broadbandcommission.org/Documents/reports/bb-annualreport2016.pdf), [retrieved: 02.2018]

[47] International Telecommunication Union, World Telecommunication Database ICT Indicators 2015.

[48] E. M. Rogers, "Diffusion of Innovations", Vol. 3, New York / London 1983, p. 247.

[49] OECD, "OECD Communications Outlook 2007 – Information and Communications Technologies", Organisation for Economic Cooperation and Development.

[50] OECD, "OECD Communications Outlook 2011 – Information and Communications Technologies", Organisation for Economic Cooperation and Development.

[51] OECD, "OECD Communications Outlook 2013 – Information and Communications Technologies", Organisation for Economic Cooperation and Development.

[52] C. Fornell and D. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," Journal of Marketing Research, vol. 18, issue 1, 1981, pp. 39-50.

[53] R. Hossiep, "Cronbachs Alpha," [German] "Cronbachs Alpha," In Wirtz, M. A. (editor): Dorsch – Lexikon der Psychologie, vol. 17. Verlag Hans Huber, Bern, 2014.

[54] S. Fromm, "Data Analysis with SPSS Part 1," [German] "Datenanalyse mit SPSS für Fortgeschrittene," Arbeitsbuch,

vol. 2, VS Verlag für Sozialwissenschaften, GWV Fachverlage, Wiesbaden, 2008.

[55] S. Fromm, "Data Analysis with SPSS Part 2," [German] "Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten," Lehrbuch, vol. 1, VS Verlag für Sozialwissenschaften, Springer, Wiesbaden, 2010.

[56] N. M. Schöneck and W. Voß, "Research Project," [German] "Das Forschungsprojekt – Planung, Durchführung und Auswertung einer quantitativen Studie," vol. 2. Springer Wiesbaden, 2013.

[57] F. Brosius, "SPSS 8 Professional Statistics in Windows," [German] "SPSS 8 Professionelle Statistik unter Windows," Kapitel 21 Korrelation, International Thomson Publishing, vol. 1, 1998.

Consortium Blockchains: Overview, Applications and Challenges

Omar Dib, Kei-Leo Brousmiche, Antoine Durand, Eric Thea, Elyes Ben Hamida
IRT SystemX, Paris-Saclay, France
Email: {first.lastname}@irt-systemx.fr

***Abstract*—The Blockchain technology has recently attracted increasing interests worldwide because of its potential to disrupt existing businesses and to revolutionize the way applications will be built, operated, consumed and marketed in the near future. While this technology was initially designed as an immutable and distributed ledger for preventing the double spending of cryptocurrencies, it is now foreseen as the core backbone of enterprises by enabling the interoperability and collaboration between organizations. In this context, consortium blockchains emerged as an interesting architecture concept that benefits from the transactions' efficiency and privacy of private blockchains, while leveraging the decentralized governance of public blockchains. Although many studies have been made on the blockchain technology in general, the concept of consortium blockchains has been very little addressed in the literature. To bridge this gap, this article provides a detailed analysis of consortium blockchains, in terms of architectures, technological components and applications. In particular, the underlying consensus algorithms are analyzed in details, and a general taxonomy is discussed. Then, a practical case study that focuses on the consortium blockchain technology Ethermint is performed in order to highlight its main advantages and limitations. Finally, various research challenges and opportunities are discussed.**

Keywords–Consortium Blockchain; Distributed Ledger Technology; Smart Contract; Consensus Algorithm; Data Privacy and Security; Scalability; Benchmark.

I. INTRODUCTION

The blockchain technology has attracted increased interests worldwide in recent years, with emerging applications in key domains, including finance, energy, insurance, logistics and mobility. Indeed, this technology is expected to disrupt businesses and markets on a global scale.

A blockchain is essentially a trustless, peer-to-peer and continuously growing database (or ledger) of records, *aka.* blocks, that have been verified and shared among the participating entities [1]. Each block typically contains a timestamp, a cryptographic hash value of the previous block, and a set of transactions data. Once a new block is validated by consensus and written to the ledger, transactions cannot be altered retroactively without the collusion of the network majority. This technology was initially designed as a public transaction ledger to solve the double spending problem in the Bitcoin digital currency system [2], without the need for any third party or trust authority.

This technology per se is not novel, but is rather a combination of well-known building blocks, including peer-to-peer protocols, cryptographic primitives, distributed consensus algorithms and economic incentives mechanisms. A blockchain is more a paradigm shift in the way applications and solutions will be built, deployed, operated, consumed and marketed in the near future, than just a technology. Blockchain is secure by design and relies on well-known cryptographic tools and distributed consensus mechanisms to provide key characteristics, such as persistence, anonymity, fault-tolerance, auditability and resilience.

More recently, smart contracts [3] have emerged as a new usage for blockchains to digitize and automate the execution of business workflows (*i.e.,* self-executing contracts or agreements), and whose proper execution is enforced by the consensus mechanism. This makes the blockchain technology particularly suitable for the management of medical records [4], notary services [5], users' identities [6] and reputations [7], data traceability [8], vehicles' history [9], *etc.*

In this context, the blockchain technology is foreseen as the core backbone of future smart cities and enterprises by enabling the interoperability and collaboration between organizations, while enhancing their security, data and process management. Although many studies have been made on the blockchain technology in general [10], the application of this technology to business environments, beyond digital currencies, has been very little addressed in the literature. Indeed, several challenges will need to be addressed to unlock the tremendous potential of blockchains especially before this paradigm shift becomes technically, economically and legally viable in enterprises environments.

These challenges concern the technical aspects of blockchains, including its architecture, deployment, governance, scalability and data privacy. Private and consortium blockchains emerged from this perspective as appropriate architecture concepts for business environments, where restrictions are applied on who is allowed to participate to the network. While the former approach assumes that the network is operated by a single entity, a consortium blockchain operates under the leadership of a group of entities, thus enabling collaborative business transformation among organizations and innovative business models. Last but not least, the legal aspects of blockchains represent a challenge, especially in Europe, where this technology should be analyzed in the light of upcoming new regulations, such as the *General Data Protection Regulation* (GDPR) (Regulation (EU) 2016/679 [11]), and whose objective is to strengthen users' data privacy and protection within the European Union.

This article aims at bridging the gap between blockchain technologies and their potential benefits to business environments, by providing a detailed analysis of consortium blockchains, in terms of architectures, technologies and applications. In particular, the underlying consensus algorithms are analyzed in details, and a general taxonomy is discussed. Then, a practical and experimental use case is discussed in order to assess the performance of the consortium blockchain technology Ethermint. The performance indicators analyzed are the time required to validate a transaction, the number of transactions validated per second, as well as the average inter

blocks delay and the size of the blockchain data folder. Parameters that have been varied are the number of validators, the number of transactions submitted per second and the topology of the network. The results highlighted some limitations in terms of transactions validation time and storage requirements that may hinder the usage of Ethermint to deal with some real world use cases.

The remainder of this article is organized as follows. Section II discusses the technical aspects of blockchains in terms of taxonomy, system architecture, and building blocks. Section III provides a detailed analysis of consensus algorithms that are used today in private and consortium blockchains. Section IV discusses a general classification of blockchains applications and highlights typical use cases in the finance, energy, mobility and logistics sectors. Section V presents a practical case study based on the Ethermint consortium blockchain technology in order to highlight its main advantages and limitations. Section VI draws and discusses various research challenges and opportunities. Finally, Section VII concludes the article.

II. Blockchain Technology Overview

While public blockchains enable parties to make transactions in a secured manner, in trust-less environments, they show certain limitations when applied to industrial use cases. Indeed, we believe that aspects, such as controlled data reversibility (**i**), data privacy (**ii**), transactions volume scalability (**iii**), system responsiveness (**iv**) and ease of protocol updatability (**v**) that are crucial for the majority of corporate applications are not covered by public blockchain implementations. These shortcomings led industrials to develop alternative blockchain technologies tackling the aforementioned aspects and intended for restricted audience. These technologies can generally be classified into two categories: private and consortium blockchains [12]. The distinction between them comes down to the governance scheme along with the infrastructure (see Table I). In private blockchains, one entity rules the whole system whereas members of consortium blockchains share the authority among them. Accordingly, the infrastructure is centralized in case of private blockchains but, similarly to common distributed databases, data is replicated on multiple nodes that belong to the single owner. In contrast, consortium blockchains are deployed in a decentralized manner on multiple hardwares managed by different owners (or companies). Moreover, data is not necessary homogeneous among consortium nodes since some blockchains allow private transactions leading to knowledge fragmentation (*i.e.,* private transactions are shared by subsets of participants).

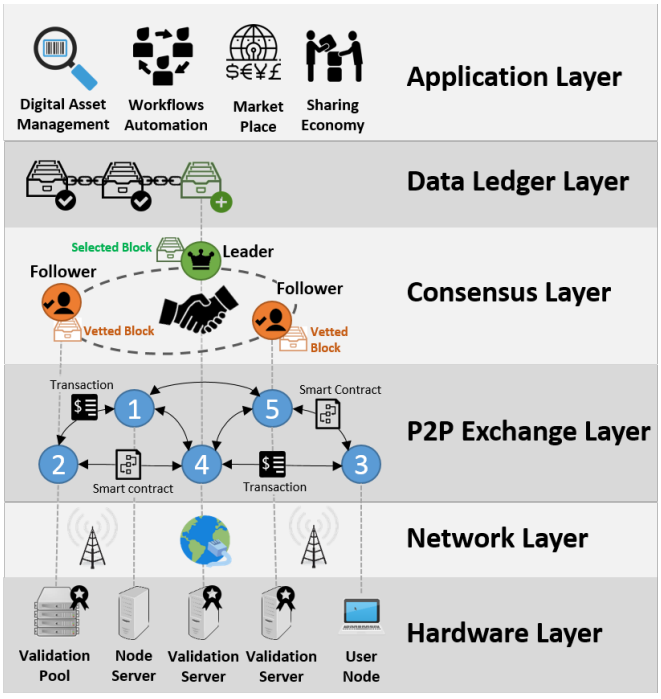


Figure 1. Blockchain High Level Architecture

Figure 1 shows the typical high level architecture of blockchain and its main layers. However, it should be noted that blockchain architecture is not yet standardized, and other representations exist in the literature, such as [13].

In the following, the term consortium blockchains will encompass both private and consortium blockchains. In the next sections, we overview the architecture of both public and consortium blockchain and highlight their inherent differences.

A. Data structure (Data Ledger Layer)

The data structure of a blockchain, whether public or consortium, corresponds to a linked list of blocks containing transactions also referred to as the "ledger". Each element of the list, has a pointer to the previous block and embodies its hash value as illustrated in Figure 2.

This hash value is the key element of the blockchain integrity, and is computed thanks to a one-way hash function that maps data of arbitrary size to a non-invertible hash value of fixed size. Indeed, if an adversary tries to modify the content of a block, anyone can detect it by computing its hash and comparing it to the one stored in the next block to see the inconsistency. In order to avoid this detection, the adversary could try to change all the hashes from the tampered block to the latest block. However, this is not feasible without the consent of a significant proportion of validator nodes, depending on the consensus algorithm (see Section III). Hence, the "immutability" characteristic of blockchain: data are tamper-proof.

On the other hand, consortium chains members can come to an agreement and alter previous blocks (**i**). In order to prove that data were not tampered and preserve the auditability of the ledger, it is common to periodically publish the hash of a block onto a public blockchain. By doing so, one can be assured that blocks in the interval of two published hashes have not been modified.

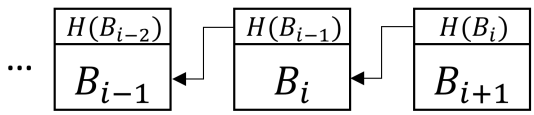


Figure 2. List of linked blocks with hash pointers

B. Network and privacy (Network/P2P Exchange Layer)

Along with its data structure, a blockchain is based on a peer-to-peer network that links its members. Members participate to the network through their blockchain client node. Each

TABLE I. Blockchain Classification

Property	Blockchain Governance		
	Public	Consortium	Private
Governance Type	Consensus is public	Consensus is managed by a set of participants	Consensus is managed by a single owner
Transactions Validation	Anynode (or miner)	A list of authorized nodes (or validators)	
Consensus Algorithm	Without permission (PoW, PoS, PoET, etc.)	With permission (PBFT, Tendermint, PoA, etc.)	
Transactions Reading	Any node	Any node (without permission) or A list of predefined nodes (with permission)	
Data Immutability	Yes, blockchain rollback is almost impossible	Yes, but blockchain rollback is possible	
Transactions Throughput	Low (a few dozen of transactions validated per second)	High (a few hundred/thousand transactions validated per second)	
Network scalability	High	Low to medium (a few dozen/hundred of nodes)	
Infrastructure	Highly-Decentralized	Decentralized	Distributed
Features	Censorship resistance Unregulated and cross-borders Support of native assets Anonymous identities Scalable network architecture	Applicable to highly regulated business (known identities, legal standards, etc.) Efficient transactions throughput Transactions without fees Infrastructure rules are easier to manage Better protection against external disturbances	
Examples of technologies	Bitcoin, Ethereum, Ripple, etc.	MultiChain, Quorum, HyperLedger, Ethermint, Tendermint, etc.	

node has a local copy of the whole linked list (or the most recent part of it in case of *light nodes* [14]). When retrieving the list for the first time, a node verifies the integrity of the blocks by computing all the hashes and keeps verifying each new block.

Depending on the implementation of the blockchain, the network can be either public (*i.e.,* anyone can access it) or permissioned (*i.e.,* only accounts that are allowed can participate). This restricted access to the network in consortium blockchains ensures data privacy (**ii**). Moreover, some blockchains allows to control data visibility at a more finer grain by enabling data encryption at transaction level (*e.g.,* [15]).

The identity of a participant is defined by his cryptographic asymmetrical key pair. The public key is derived to obtain his unique address, which serves as his public identity. The private key is used to sign transactions and guarantee their authenticity (*i.e.,* other participants can verify the signature using the associated public key).

In order to add data to the blockchain, a node sends a transaction request to the network. The prime data fields of a transaction in most technologies are the addresses of both sender and receiver, data values that are being communicated and the signature of the sender. These transactions' requests are then picked by some special nodes called *miners* also referred as to block generators or validators on consortium blockchains.

C. Security vs scalability (Consensus Layer)

On public blockchains, miners are nodes that are willing to share their computational power to add blocks to the blockchain in return of some reward. The way they are rewarded depends on the implementation of the blockchain protocol, however, it often involves a fee (*i.e.,* the node who was asking to add data pays the miner to do it) and/or creation of value (*e.g.,* in case of Bitcoin, the blockchain mints value and gives it to the miner who has added a block). Thus, miners are in competition: they all want to add the next block but only one or few of them will achieve it in a random way for each new block.

The process for selecting the actual node that will add the next block among all the miners is referred as a consensus protocol. For instance, Bitcoin and Ethereum use the Proof of Work (PoW) consensus [2], where miners have solve a computationally demanding cryptographic puzzle to prove their commitment. This protocol enables the random selection of the next miner and prevent adversary nodes from adding fraudulent blocks since their probability to be retained is too low compared to the procured energy and time to solve the puzzle. This leads to the 51% attack: if more than half of the power of the network is allied with the adversary, then his version of the ledger will always end up being the main one.

In a trust-less public configuration where miners are anonymous, this consensus is crucial for the integrity and security of the data. On the other hand, in consortium chains, validators are preliminary known according to the governance scheme and are trusted to some degree. Indeed, in majority of consortium protocols, validators are defined at the genesis of the blockchain. Some technologies enable the dynamic addition and retrieval of validators but these actions are always under the control of the current validators. Therefore, the security and required computational power can be lowered by using less demanding consensus algorithms. This reduction of complexity in the consensus protocol leads directly to an increased scalability in terms of transactions throughput (**iii**). Indeed, consensus such as PoW delay data transmission, *e.g.,* 10 minutes for each 1Mb block for Bitcoin, whereas protocols for consortium blockchains can make optimal use of the network and reach throughput on the order of thousands transactions per second. An overview of the major consensus algorithm for consortium blockchains is proposed in Section III.

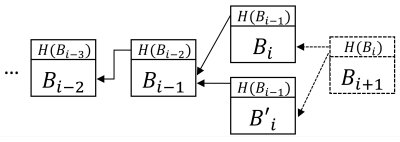


Figure 3. Fork

D. Forks and responsiveness

Once a miner's block has been published, it is added to the blockchain and the information is broadcast. Due to network effects, there are cases where multiple miners blocks are published at the same time. It is also possible that some miners propose other block versions when they do not agree

with the content of the most recent one. These cases are called a *fork:* the blockchain splits into multiple branches. However, the nodes goal is to converge towards acknowledging a unique and same version of the blockchain. In practice, the Proof of Work consensus achieves this result by requiring miners to work on the longest branch, with the intuition that eventually the longest branch will be the same for all nodes.

Thus, even if a transaction has been validated, we cannot be sure that it will remain on the main chain. In Bitcoin, users usually wait six blocks of confirmation before considering a transaction (*i.e.,* its block) as valid, that being $6 \times 10min = 1h$. In that sense, transactions are never definitely accepted, there is only the probability of reversal that decrease exponentially as the chain grows.

Hence, there is a correlation between the probability of fork occurrence and the *responsiveness* of the blockchain (*i.e.,* time to wait for a transaction to be validated and adopted by the rest of the network). On consortium blockchains, the use of adapted consensus algorithm allows for "block finality": once a block has been validated, it remains on the main chain and forks are not allowed. This increases the system responsiveness by shrinking waiting time for confirmations (**iv**).

E. Forks and updates

Miners software is sometimes updated to fix bugs or add functionalities. This also can create forks, as different nodes might handle transactions differently depending on their software versions. We usually distinguish:
- "soft forks" where the transactions considered valid by the new version are also valid for the old version.
- "hard forks" where the transactions considered invalid by the old version might be valid for the new version.

While it is complicated to synchronize the software of public blockchains due to the huge amount of anonymous participants and potential disagreements among them, it is easily feasible on consortium chains where members know each other and can quickly come to a mutual agreement (**v**).

F. Smart Contracts

The concept of smart contract has been introduced by Ethereum and consists in computer programs, which executions results are verified by miners. Their deployments and executions are triggered by users through transactions. Each node participating to the blockchain has a local virtual machine (*e.g.,* EVM in Ethereum [16]) in addition to the linked list and executes smart contracts, according to the transactions that have been validated and maintain a globally shared state. By inheriting the security level offered by transactions, blockchain's smart contracts protocol benefit from the properties of the blockchain: security, integrity, no intermediary, transparency and availability. Some consortium blockchains offer the possibility to restrict the visibility of these contracts to a subset of members in the same way as private transactions (*e.g.,* [17]).

This overview of blockchain architecture and components highlighted the main differences between public and permissioned implementations. In the next section, we present major consensus algorithms used in consortium blockchains that enables high transaction throughput compared to the regular algorithm such as the Proof of Work or the Proof of Stake.

III. Consensus Algorithms

As we saw in the previous section, the process that adds blocks to the linked list is a major factor to security and scalability. This is done with a *State Machine Replication* (SMR) algorithm, which makes the network agrees on a unique, constantly growing, ordered set of transactions. The *consensus* algorithm, is the part that makes the nodes agree on a single piece of data [26], *i.e.,* the block that is going to be added. However, since they are both very closely related, the terms are often used interchangeably.

To solve consensus, an algorithm has to provide two properties: Safety and Liveness. Informally, Safety means that once nodes confirmed a new entry to the transaction log, it will not be changed later. Liveness means that if a honest user has sufficient connectivity and tries to submit a transaction, then it will eventually get accepted.

The reason why blockchains are deemed to be *trustless* is because those algorithms are *Byzantine Fault Tolerant* (BFT), that is, other nodes may behave arbitrarily, including in malicious ways. Therefore, blockchains users do not need to trust others. However, it has been proved that no consensus algorithm can tolerate more than a third of the nodes being malicious [27]. Note that blockchains may tolerate more byzantine faults, but will not satisfy the traditional definition of consensus. This is most often the case with public blockchains that do not have transaction finality (or block finality, as mentioned earlier), *i.e.,* transactions are guaranteed to be irreversible only with a sufficiently high probability.

A. Algorithms

In this section, we describe prominent consensus algorithms that are considered today in consortium blockchains where miners are designated nodes (*i.e.,* validators). In Table II, we synthesize their characteristics: *Fault tolerance* indicates whether the algorithm tolerates arbitrary (*e.g.,* malicious) behavior, or only crash failure. The *Threshold* further quantifies the amount of faults tolerated. *Confirmation time*, also known as latency, is the time elapsed between the submission of a transaction, and its definitive acceptance. *Throughput* is the number of transactions per second that the system can sustain as a whole, and *Scalability* is the number of nodes that can participate in the consensus. We warn the reader that the figures for the last three columns are not based on rigorous experimental studies, but they were collected from online blogs and websites. Indeed, these performances are not expected to be reachable simultaneously for the three metrics, and depends largely on the algorithm configuration and network settings.

a) Practical Byzantine Fault Tolerance (PBFT): PBFT is the first high performance consensus algorithm with an optimal byzantine fault tolerance [28]. As such, it has been widely implemented and tested, and served as a basis for a wide range of newer algorithms.

This method works with a leader that does the ordering of transactions. Then consensus is reached in three phases: First, the leader broadcasts the new request, *e.g.,* transactions or block (pre-prepare). Then, each validator signs and broadcasts a *prepare* message for the request. If enough prepare messages have been received, then validators broadcast *commit* messages, and when enough commits have been received, the request is finally accepted.

TABLE II. BENCHMARKING OF CONSENSUS ALGORITHMS

Consensus algorithm	Fault Tolerance	Threshold	Confirmation time	Throughput	scalability
Tendermint Core	Byzantine	33%	5s [18]	10k tx/s [18]	100 nodes [18]
PBFT	Byzantine	33%	1s [19]	50k tx/s [19]	30 nodes [19]
Hashgraph	Byzantine	33%	n/a (claimed 1s) [20]	n/a (claimed 100k tx/s) [20]	n/a
SCP	Byzantine	partitioning	up to 15s [21]	1-10k tx/s [22]	-
PoET	Byzantine	TEE failure	n/a	n/a	very high [23]
DiversityMining	Crash	tunable, $\leq$ 50%	n/a	1k tx/s [24]	n/a
Raft	Crash	50%	1s [25]	up to 30k tx/s [25]	n/a

If at some point the leader does not behave as expected (*e.g.,* if there was no answer after some timeout), the next leader in the round robin order takes over through a complex procedure named *view change*. Informally, view change can be seen as a weaker variant of consensus, because nodes need to reach agreement on the fact that the leader changed.

This algorithm is able to tolerate a maximum of a third of the nodes being malicious, which is the maximum possible in this setting. It also makes the assumption that the network is partially synchronized since nodes need to be able to skip faulty leaders, *i.e.,* after a timeout. A PBFT implementation is currently in development for Hyperledger Fabric. Moreover, IstanbulBFT has been developed as a blockchain-friendly variant of this algorithm, and is implemented in Quorum [29].

b) Tendermint: Tendermint is a consensus algorithm that has been designed with applications for consortium blockchain in mind [30]. It is somewhat similar to PBFT, but additionally provide safety even when more that one third of the nodes get corrupted [31]. For each new block, a leader node is selected in a round-robin manner until one is able to commit the block. For a block to be committed, it has to gather more than two third of votes of validators within a given time period, through a procedure similar to the three-phases commit from PBFT.

Tendermint-core is provided as a stand-alone consensus engine, and has been implemented to function with the Ethereum Virtual Machine (EVM) in Hyperledger Burrow and Ethermint. Other blockchain frameworks based on Tendermint can be found on [32].

c) Proof of Elapsed Time: Proof of elapsed time is an initiative that aims at removing the computational cost induced by the usual proof of work approach, by leveraging Trusted Execution Environment (TEE) compatible hardware, such as Software Guard Extensions (Intel SGX).

For each block, miners wait for a given random time. The first miner for which the waiting time has elapsed is selected to validate a block before repeating this process. In other words, the miner with the shortest waiting time is elected as the leader. To prove that the node actually waited the required time, the waiting procedure is executed within the TEE, which produce a proof of its execution [33]. In essence, this is the same mechanism as in PoW-based consensus, except that the cryptographic puzzle is replaced by a hardware-enforced wait procedure. Note that the trusted hardware can also enforce that all nodes follow the protocol in its entirety (*i.e.,* are honest), thus making any crash-tolerant protocol withstand malicious faults as well. However, the precise construct of PoET allows to keep the advantages of PoW-based consensus, namely, the great scalability in the number of nodes. This does come with the same performance cost than public blockchains, since the waiting delay must be large enough to have a low probability of a (network-induced) fork.

The downside is that the security is guaranteed only if the TEE platform vendor is trusted. Moreover, if one is willing to assume that nodes may still be compromised (*e.g.,* due to implementation bugs), then only a few number of them would be sufficient to compromise the whole system [23]. PoET is primarily used within the Hyperledger Sawtooth platform [34].

d) Stellar Consensus Protocol (SCP): The Stellar Consensus Protocol [35] is the algorithm that was developed to power the Stellar network. It breaks the usual prerequisite of a unanimously accepted membership list by letting participants independently choose the nodes they trust. Each participant knows some nodes that are considered as trustful, *i.e.,* its *neighborhood*, and waits that the majority of them agree on a new transaction before considering it as valid. Consensus within a neighborhood is reached by first iteratively nominating candidate values, then by committing one. Those procedure are based around a primitive that allows nodes to ratify statements, which is done with two round of voting.

This approach allows each node to only be aware of a limited number of neighbors, while enabling an efficient network-wide consensus. On the other hand, security relies on each user good configuration, and requires that each neighborhood sufficiently intersects with each other. Additionally SCP may be less suited to consortium blockchains, due to its federated architecture, where one neighbor alone will not be trusted. Therefore, to actually benefit from SCP, a given blockchain platform should either anchors to the Stellar network, or try to spawn a new independent network.

e) Hashgraph: Hashgraph is an asynchronous protocol, which means that it makes no assumptions on the network delays [36]. This allows the nodes to be connected through an unreliable network such as the Internet. Its core idea is to piggyback the underlying broadcast protocol (*i.e.,* gossip) to reach consensus. By explicitly recording the network-layer execution of the broadcast into a graph similar to a blockchain, it is able to track events that has been sufficiently spread in the network to reach consensus. The Hashgraph authors show that if nodes are constantly synchronizing, then a given event will end up sufficiently deep in the graph so that one can be assured that a majority of nodes are able to tell that this event is valid.

Swirlds [37], the company that developed this algorithm, made available an implementation through a closed source Java SDK in alpha version. However, at the time of writing there was no independent benchmark of Hashgraph, which makes assessing its performances difficult.

TABLE III. BENCHMARKING OF ENTERPRISE BLOCKCHAIN PLATFORMS AND TECHNOLOGIES.

Company	Platform	Consensus Algorithm	Smart Contracts	Private Transactions	Popularity ♠	Activity (Github) ♠
Coin Sciences Ltd	MultiChain	DiversityMining	No	No	Low	Medium
Quorum	Quorum	pluggable: IstanbulBFT ①, Raft	EVM	Yes	High	Medium
IBM	Hyperledger Fabric	pluggable ②: Kafka	Chaincode	Yes ③	High	High
R3	Corda	pluggable: Raft, BFT-SMaRt	JVM	Yes	Medium	High
SWIRLDS	Hashgraph	Hashgraph	No ④	No	Low	N/A
Stellar	Stellar	SCP	No	No	High	Medium
ParityTech	Parity	pluggable: Ethereum, Aura ⑤, Tendermint	EVM	No	High	High
Intel	Hyperledger Sawtooth	PoET	EVM	No	Low	High
Monax, Intel	Hyperledger Burrow	Tendermint core	EVM	No	Medium	Medium
All In Bits Inc.	Ethermint	Tendermint core	EVM	No	Medium	Medium

♠ Rough estimation based on GitHub metrics and online presence (*e.g.,* download count, community hubs activity, *etc.*)
① IstanbulBFT is an adaption of PBFT for blockchains
② Other consensus to be added, or unofficial: PBFT, BFT-SMaRT, HoneyBadgerBFT
③ Stand alone private transactions are not possible, but participants can set up private channels
④ Hashgraph let the transactions semantics be implemented by the application, but they will not be checked during consensus
⑤ Aura is a simple consensus engine developed by ParityTech, but it is not well specified and there is no assessment of its soundness

f) Diversity Mining consensus: DiversityMining is a consensus algorithm developed by MultiChain [38]. It is based on leader election, but where PBFT uses timeouts to handle leader failure, DiversityMining allows a fraction (subject to a parameter named *diversity*) of them to claim leadership independently, directly by broadcasting a block. Thus, the amount of tolerable faults is directly set by this parameter. Forks resulting from this process are handled the same way as in Bitcoin, with the longest-chain rule. As a result this algorithm does not provide transaction finality and does not fit the usual definition of consensus in the permissioned model.

Moreover, this algorithm only tolerates crash faults (*i.e.,* not malicious nodes) [39] and thus should be compared with other crash tolerant systems, like Raft [40] or Apache Kafka. At the time of writing, there was no available assessment of DiversityMining performances with a high number of nodes. However, since it exhibits a communication pattern similar to public blockchains, it can be expected to have viable performance with large sized network, at least under optimal conditions.

g) Raft: Raft [40] is a consensus algorithm that has been designed to be understandable and modular. It tolerates up to 50% of crashed nodes, which is the maximal threshold for this kind of faults. It is based on a leader that does the ordering of transactions. Leader election is triggered by a timeout, but contrary to PBFT, the timeout is randomized for each server. As a result, the elected leader (if successful) will be the first to timeout. Once there is an available leader, it can simply broadcast transactions to impose its version of the transaction log. Raft has been widely adopted and has a great number of implementations, so its robustness and practical performances are well known [41].

B. Benchmarking Existing Technologies

Blockchain is currently under extensive research and development, leading to a high market fragmentation, with more than 20 different technologies and frameworks, which have been released by companies, open-source communities and universities. Table III compares the key characteristics of some popular blockchain technologies, especially for the context of enterprise and consortium based case studies. The column consensus algorithm lists the consensus engine that have a compatible implementation. In column Smart Contract we give the mechanism that execute the smart contract (*e.g.,* virtual machine, specification), if this feature is available. In Private Transactions, we specify whether the platform allows client to send transaction whose contents are only available to the recipients, possibly including their identities. Then we give a rough estimation of the platform popularity and activity based on github metrics and online presence (*e.g.,* download count, community hubs activity, *etc.*)

Two points should be noted regarding private transactions: First, if the blockchain allows arbitrary data in transactions, it is always possible to add encrypted data using the recipients public key, but then the validators cannot verify the semantics of the transaction (*e.g.,* double spending), although some solution exists to this issue [42][43]. Second, in the particular consortium setting, it is always possible to instantiate new blockchains for a specific subset of users, which is conceptually what Hyperledger Fabric does with its *channels*. However, this is limited in the sense that anything that happens within a channel cannot have an influence on something external to it, *e.g.,* a monetary transaction cannot be redeemed outside the channel where it has been made, only the members can vouch for it. In the corresponding column, we state "Yes" when the platform has a working implementation of a feature that allows for some level of privacy, including the mentioned techniques, "No" otherwise.

As it can be seen from Table II, consensus algorithms in the classical setting, *i.e.,* that tolerate 33% byzantine nodes, have similar characteristics: they show good performance but do not scale well when the number of nodes increase. There are proposals to address this issue such as SCP or PoET, but they imply alternative models that are subject to caveats, especially regarding trust assumptions.

Therefore, an application that requires cooperation from a small set of distrusting entities will be able to use a more classical consensus algorithm such as Tendermint or PBFT, and consequently will have a wide range of choices for a blockchain framework. Then, the determining factor is related to the additional features that are specific to each blockchain platform, such as the ones listed in Table III.

On the other hand, applications that require a larger network, *e.g.,* in a scenario where each registered user of the decentralized application needs to participate to the consensus, the choice will have to involve some performance trade-off. For instance, PoET may meet the scalability requirements, but not provide a low enough confirmation time. As a result, it may be difficult to find a fitting blockchain platform that also have a consensus algorithm meeting those requirements. Finally, applications that does not actually need a trustless platform can rely on crash tolerant systems where scalability is less an issue. Therefore, the choice for a blockchain framework will be driven by the specific characteristics of each platform, similarly to the first case.

IV. Applications and Case Studies

The advent of the blockchain technology has enabled a wide range of new applications. In this section, we introduce a general classification of these applications, followed by examples of case studies in key domains.

A. Classification

In a centralized world, an ecosystem is organized around one predominant actor. This architecture has advantages: it is easy to manage, stable and the responsibilities are clear. But it also has drawbacks: a monopolistic approach creates barriers to entry and hinders innovation, it can also result in an unbalanced value distribution among participants.

With the blockchain, we move towards a distributed approach with 4 categories of benefits: data traceability for a new trust paradigm, systems interoperability for process automation, flexibility for services enhancement, token-based ecosystems for a shared governance.

1) Data traceability for a new trust paradigm: The blockchain can be seen as an immutable distributed ledger where transactions are timestamped by block. The derived properties (transparency, integrity, immutability) create the trust conditions that enable a new framework for asset transactions. Indeed, from a vision perspective, the blockchain is to value exchange what the Internet is to information exchange: an interconnected network developed, maintained and updated by the participants for their benefits, a common ground that is safe, neutral, disintermediated and universal. Based on this framework, it is now possible to better trace, manage and monetize data. This offers new perspectives from machine to machine transactions, to identity attributes sharing, to user-centric data marketplaces.

2) Systems interoperability for process automation: The blockchain can also help bringing down domains silos. As an example, the transport infrastructure will more easily interoperate with the energy infrastructure. How so? As a shared ledger, the blockchain creates the conditions for process standardization, via shared on-chain data models, smart contracts and rules, between any actors that would like to work together. These can be actors within a domain, say a supplier and a customer, or actors from different domains, as in the electro-mobility example. Once standardized, processes can then easily be automated by using smart contracts (also known as chaincodes in the Hyperledger vocabulary).

3) Flexibility for services enhancement: Based on these interoperable systems, it is now possible to develop new features or services. Having access to more shared data and more standardized processes will definitely help. But the new operational organization will help too. Because there is no going through a third party, everyone on the blockchain is free to implement new features or services directly, at its own pace (and its own risk). More freedom also means more responsibilities. But this flexibility is good for services evolution and improvement. For example, it becomes easy to create ad-hoc temporary offers. It is also becomes easy to provide personalized services, by leveraging the shared data mentioned above.

4) Tokens-based ecosystems for a shared governance: Beyond the traceability, interoperability and flexibility benefits, the blockchain can help fundamentally transform ecosystems by leveraging digital assets, also known as tokens. Indeed, tokens can be used to track and monetize transactions, but they can also be considered as a great tool to materialize the governance rules and maintain an equilibrium among actors. The minting protocol and the trading rules put in place will reflect the consortium view on how to steer behaviors and share the value created. As an example, a green token could be created and used only in environmental-friendly scenarios. Going a step further, a decentralized autonomous organization (DAO) is another way to create new ecosystems. A DAO is an organization that relies on the blockchain to manage interactions between participants. Rules are implemented in smart contracts executed on the blockchain, so the governance is executed automatically with transparent rules and immutable actions.

B. Case Studies

While we classified above the blockchain use cases in several categories, we can now look at concrete examples in different application domains.

1) Finance and Insurance: The first blockchain application was the cryptocurrency Bitcoin. But many use cases have followed since. As an example, it can be used at the ”data level” to issue and trade assets, such as bonds, in a decentralized market place (see the proof-of-concept from Caisse Des Depots in France). At the ”infrastructure level” Chaincore implements the distributed ledger technology for clearing and settlement, as a way to lower costs and improve efficiency. The blockchain can also help with processes such as KYC (Know Your Customer), by sharing the proof of identity and not the data itself between banks (see KYC-chain as an implementation example). At the ”service level”, we can imagine new offers such as personalized short term insurances [44], created on the spot and taking into account diverse pieces of information about the beneficiary. Finally, crowd-sharing an insurance deductible can be a good DAO application in the insurance sector.

2) Energy: With the rise of solar panels and other green sources of energy, the energy production is becoming more decentralized and offers a promising field for blockchain applications. As an example, the distributed ledger technology can be used to certify the data, the source of energy production, therefore, guaranteeing that it is environment-friendly. It can also be used to trade energy at the local grid level, between individual producers and consumers (see the proof-of-concept

from LO3 Energy in Brooklyn or [45]). We can imagine further benefits in the home where devices can schedule their energy charging to optimize costs and exchange data autonomously between them. This can be a totally new ecosystem model, where tokens are directly exchanged between parties to incentivize appropriate (green) behaviors.

3) Mobility: In this sector, the distributed ledger technology can be used to safely store the car data (for example, its mileage [9] or certificates [46]). We can also look at electro-mobility use cases, where the mobility infrastructure (say electric vehicles) interoperates more easily with the energy infrastructure (say the charging points). Another application example is chasyr, which is a blockchain-based ride-sharing platform that matches passengers and drivers. So, this is basically an uber-like service, in a decentralized architecture. One last example would be a decentralized transportation ecosystem, where people can use a same token to ride on a bus, rent a bike or carpool, without any central authority to organize its operation.

4) Logistics: In this sector, the distributed ledger technology can be used to track an asset. For example, Everledger tracks diamonds to ensure their authenticity, Provenance can track food origin to guarantee its sanitary safety. Another example would be using a blockchain to create a collaborative IT system, which matches transporters and customers timetable for efficient delivery.

V. A Practical Case Study - Analyzing the EtherMint Technology

In this section, we provide a detailed study related to the performance evaluation of the Ethermint consortium blockchain technology. Being based on Ethereum, this technology inherits all the capabilities including the EVM and smart contracts. Moreover, the consensus relies on the Tendermint protocol. By combining those two characteristics, a versatile protocol and a lightweight consensus that remains secure, we believe that Ethermint is a great candidate for various use cases.

Although extensive studies have been conducted to assess the performance of the Tendermint consensus protocol such as [31] or other blockchain technologies such as [22], to the best of our knowledge, the literature has not provided yet any detailed study related to the technical performance of Ethermint. That is the first reason behind using this technology for benchmarking. The second reason lies in the fact that many industrials, such as in [9], are looking for the usage of this technology for their own use cases. However, they do not have yet any detailed assessment of the performance of this technology. Therefore, for the sake of helping them testing this technology, this study is conducted.

The remainder of this section is organized as follows. First we give a detailed description of the Tendermint consensus protocol before presenting Ethermint itself (Sections V-A and V-B). Then, we specify the experimental setup and the results regarding the transaction validation speed depending on multiple factors such as the transaction load and the network topology (Sections V-C and V-D).

A. Tendermint: Overview and Architecture

Tendermint [30] can be seen as a software for replicating an application on many machines in a secure and consistent manner. Tendermint protocol guarantees that machines compute the same state and it can tolerate up to 1/3 of malicious or failing machines.

From a functional architecture point of view, Tendermint consists of two main components: a blockchain consensus engine called Tendermint Core, which is used to ensure that the same transactions are recorded on every machine in the same order, and a generic application interface called the Application blockchain Interface (ABCI) that is used to enable the transactions to be processed in any programming language.

In contrast to most current blockchain solutions (such as Bitcoin) that come pre-packaged with built in state machines, Tendermint can be used for replicating state machines related to applications written in any programming language or development environment. Thanks to these features, Tendermint has been widely used in a variety of distributed applications including blockchain platforms such as Ethermint [47] and Hyperledger Burrow [48].

From a consensus point of view, Tendermint belongs to the family of BFT (Byzantine Fault Tolerance) consensus protocols [49]. More precisely, participants in this protocol are called validators; their main role is to propose blocks of transactions and vote on them. Blocks come in the form of a chain and only one block is committed at each height.

As in most consensus protocols, a block may fail to be committed due to many reasons such as network connectivity or malicious behaviors. In such a case, the Tendermint protocol moves to the next round, and a new validator is designed to propose a block at this height level. The selection of a proposer is proportional to its validation power. Considering the voting phase, two stages are required to successfully commit a block; a pre-vote and a pre-commit. For any block to be committed, more than 2/3 of validators must pre-commit for it in the same round. We show in Figure 4 the optimal validation scheme of transactions using Tendermint.

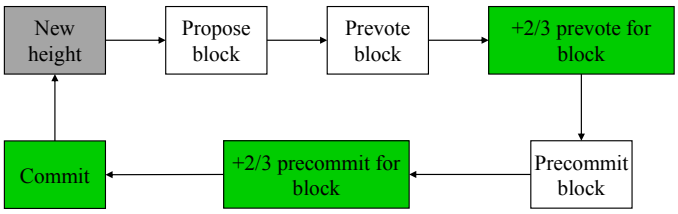


Figure 4. Tendermint consensus: optimal workflow

Finally, it should be noticed that Tendermint is overall qualified as a weakly synchronous protocol since validators continue to do their work only after hearing from more than 2/3 of the validators set. Furthermore, validators have also to wait for a small amount of time in order to receive a complete proposal block from the proposer before voting to move to the next round.

B. Ethermint: Ethereum + Tendermint

As previously mentioned, Tendermint has a generic application interface that enables the transactions to be processed in any programming language. Indeed, to use Ethereum with a Tendermint consensus protocol, Ethermint has been developed. Using Ethermint was first introduced in May 2017, as part of

Tendermint's goal to launch the COSMOS hub [50], the first blockchain in the Cosmos network, which is a decentralized network of independent parallel blockchains.

The key idea of Ethermint is to enable Ethereum to run on top of Tendermint. This allows developers to have all the nice features of Ethereum, while at the same time benefit from Tendermints consensus protocol implementation. Tendermint combined with Ethereum is supposed to result in fast block times, transaction finality while also getting the goodies of smart contracts.

In the next sections, the performance of the blockchain technology Ethermint is considered. This will encompass explaining the experimental setup, describing the assessment workflow, as well as, analyzing several indicators related to the performance of this technology.

C. Experimental Setup

To assess the performance of Ethermint, several parameters are studied and many performance indicators are considered. The evaluation process consists of dynamically deploying a blockchain network on an Openstack virtual machine [51] having the following properties (20 GB of RAM and 6 Virtual Central Processing Unit). The used Tendermint version is $0.12.0$ and $0.5.3$ for Ethermint.

Parameters used for building the blockchain are: the number of nodes n, the number of validators v and the network topology t. By this latter, we mean how nodes are connected to each other. In this work, the network may be ***complete*** (each node i is directly connected to all other nodes), in the form of ***line*** (each node i is connected to node $i+1$ except for the last node n), ***cycle chain*** (line and the last node is connected to the first node), ***enhanced chain*** (a node i is connected to nodes $i+1$ and $i+2$; the $i+2$ node of the node $n-1$ is the first node; the $i+1$ node of the last node is the first node and the $i+2$ node of the last node is the second node).

The above explained topologies are represented in Figure 5. The main reason why the network topology is considered is due to the fact that the Tendermint consensus protocol lies in a very important gossiping phase between all nodes in the network in order to accomplish the validation steps (look at Figure 4 for more information). Therefore, any topology will certainly affect the speed at which the information circulates in the network.

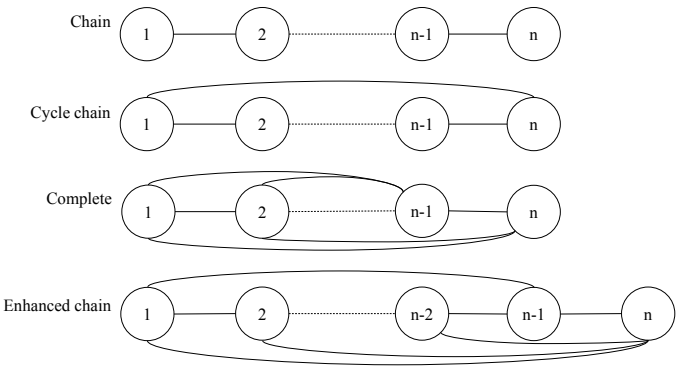


Figure 5. Network topology

To perform the deployment process, a NodeJS web service has been developed. The aforementioned parameters are used as an input for this service. A ready to use blockchain respecting those parameters is generated as an output. More precisely, the service works as follows: A NodeJS server always listens on a specific port; once a correct request is received, n containers are built (n corresponds to the number of nodes in the blockchain). Each container will hold a Tendermint node associated with an Ethermint node.

To deal with the validators for the Tendermint consensus, the first v Tendermint nodes will be selected. Besides, each node in the validators set will ask for the public key of all other validator nodes in order to build a common genesis file. This latter contains the list of validators and the power associated with each of them. We assume in this work that validators have the same validating power.

Once the genesis file is ready and successfully shared, Tendermint nodes will start and so will do Ethermint nodes. Moreover, the web service will check that all containers are successfully started and the blockchain is ready to use by asking for the first block information; a success flag is returned if and only if the blockchain already starts validating blocks (in this case empty blocks are generated). Moreover, another service has been developed in order to destroy the blockchain. By destroying, we mean removing all containers and data related to a blockchain.

Besides, a separate Java program that is run on a 8 GB of RAM is used for interacting with the blockchain, as well as, sending transactions and computing performance indicators. As for all Ethereum based blockchains, instances of web3j client have been used. To assess the blockchain performance, this program will dynamically create and remove a blockchain by calling the corresponding services.

Moreover, this program will be in charge of sending transactions in asynchronous mode (*i.e.,* we do not wait for the transaction to be validated). The number of transactions to be sent in second as well as, the sent interval duration and the dynamic blockchain parameters are considered as input for this program. The sent transaction consists of adding an element to a map in the smart contract that is written in Solidity. The map assigns a random value to the account calling the smart contract.

At the end of each scenario (*i.e.,* after the assessment duration), a validator blockchain node is contacted in order to fetch all validated transactions and map them with the sent transactions. By doing so, several metrics can be computed such as the duration between the sending time of a transaction and the time at which the transaction has been written on the blockchain (the validation time of a transaction). Furthermore, the number of transactions that each block contains is computed, as well as, the time between two blocks. We show in Algorithm 1 the assessment workflow from a high level point of view.

D. Experimental Results

We start this section by analyzing the impact of both the number of transactions sent per second, and the number of validators on the blockchain performance. By this latter, we mean the number of transactions validated per second. The network topology in this scenario is fixed to be complete. The results are compared with the ideal case line where all sent transactions are validated in one second or less.

Algorithm 1: Assessment workflow

input : Number of nodes n,
Number of validators v,
Topology of the network t,
Assessment duration d
Number of transactions per second (frequency f)

output: Performance indicators (average transaction validation time, average block size, number of transactions validated per second, etc)

```
1  begin
       // BlockChain Deployment
2      buildBlockChain(n,v,t)
3      waitForBlockChainToBeReady()
       // Send Transactions
4      while duration < d do
5          for i = 0; i < f; i++ do
6              tx ← PrepareTransaction()
7              send(tx) // Send Asynchronously
8
9          end
10         sleep(1s)
11     end
       // Extract performance indicators
12     ComputeAverageValidationTime()
13     ComputeAverageBlockSize()
14     ComputeAverageTransactionsPerSecond()
15 end
```

As can be seen from Figure 6, the results show that more the number of validators increases, more the number of transactions validated per second decreases. This can be explained by the fact that more communications are required to pre-vote and pre-commit a block when the number of validators becomes important. Besides, it has been noticed that the number of transactions sent per second has an impact on the output of the blockchain.

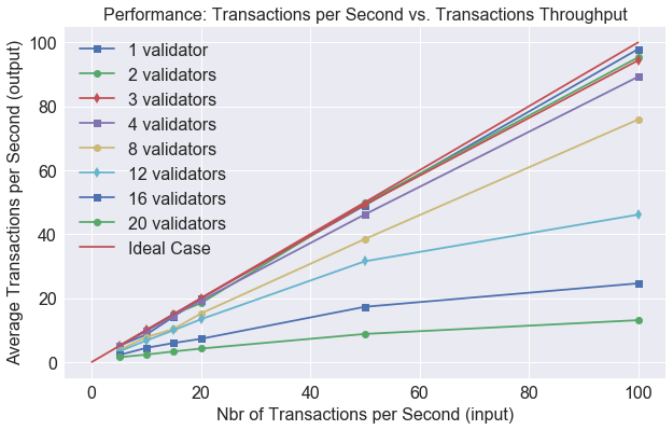


Figure 6. Average transactions per second

More precisely, more transactions sent per second, more the network accumulates some delays to validate transactions. For instance, in the worst case when 100 transactions are sent per second, a network of 1, 2, 3 or 4 validators can almost validate them in one second or less. However, for a network containing 8, 12, 16 or 20 validators, the blockchain requires several seconds to validate all the input.

When it comes to assessing the impact of the size of validators set on the average transaction's validation time, the results in Figure 7 show that more the number of validators is important, more the time required to valid a transaction increases.

For instance, in a network containing 1 validator, the average validation time is very low (less than 1 second), however, that increases to approximately two minutes when 20 validators are considered. The confidence interval is also given in this figure.

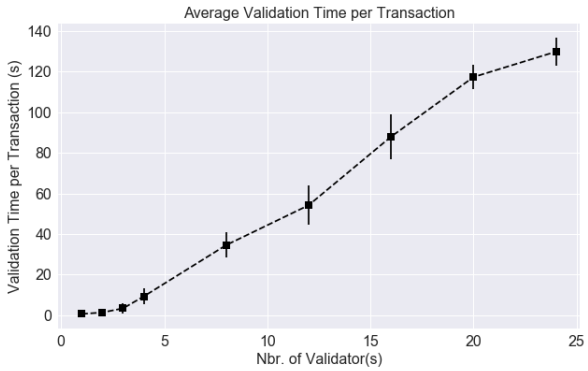


Figure 7. Validation time

From both Figures 6 and 7 it can be said that a compromise has to be made between the number of validators considered, the number of transactions sent per second and the desired blockchain performance. More precisely, an additional effort has to be made in order to select the appropriate number of validators in the network, and to fix the adequate number of transactions that the blockchain can deal with.

This usually depends on the use case where the blockchain is used. For instance, within a blockchain based energy market place, that requires validating 100 of energy exchange transactions every 20 seconds, it is obvious that it wont be possible to achieve that using more than 4 validators.

In the following, the impact of the network topology is studied. As previously mentioned, four topologies are considered: complete, chain, cycle chain and enhanced chain. To do the assessment, the number of validators has been fixed at 20 and the input transactions number is varying.

The results in Figure 8 indicate that the topologies are in the following order (from best to worse) enhanced chain, cycle, chain, complete regarding their performance in terms of the number of transactions that have been validated. This can be explained by the fact that in a complete network, the validator nodes spend more time in order to synchronize their state with the rest of the network. As a result, such nodes have less time in order to accomplish the validation work.

In a chain topology, the time required to spread an information (*i.e.,* send or receive votes) is quite high in comparison with other topologies. In contrast to those latter, the enhanced chain topology, represents a good compromise between the spread information time and the synchronization effort. More precisely, each node is to be synchronized with only four nodes

(see Figure 5) and the time required to spread information is more optimized in comparison with other topologies.

In the following the time between two blocks is studied. The results in Figure 9 show that more the number of validators increases, more the inter-blocks delay increases. This is coherent with what have been previously obtained. Indeed, in a network containing an important number of validators, a validator node requires additional time in order to inform/get informed about the state of other nodes. That will certainly delay the validation/creation of new blocks in the network.

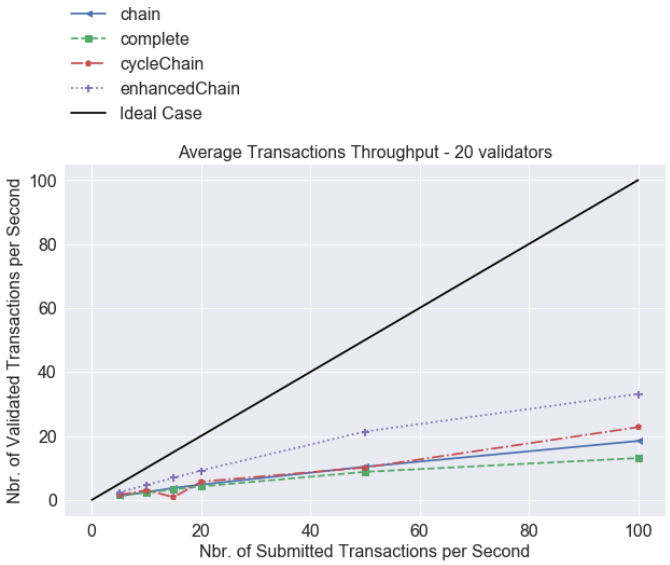


Figure 8. Network topology effect

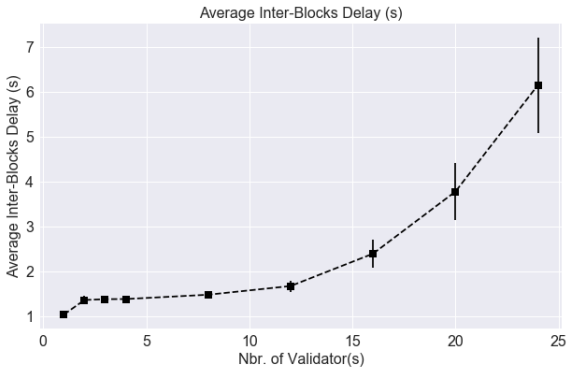


Figure 9. Average inter blocks delays

The last parameter studied in this paper is the size of the blockchain. By this we mean the actual size of the folder containing the blockchain information. For this purpose, four scenarios have been considered. The first one consists of sending 20 transactions per second during 2 hours to a network containing 1 validator.

The results in Figure 10 show that the blockchain size increases with time. After two hours, the final size increases to approximately 90 Megabytes. When 20 validators are considered, the size is not so important as in the previous scenario since few transactions will be validated. As a result, it can obviously be said that the size of a blockchain is proportional to the transactions that the blockchain writes. That is, more validated transactions will obviously result in higher blockchain size.

A final note is related to the sudden decrease in the size. This is actually due to a compression mechanism used in Ethermint. More precisely, at each increase of approximately 35 megabytes, a compress mechanism is applied by which approximately 15 megabytes are gained.

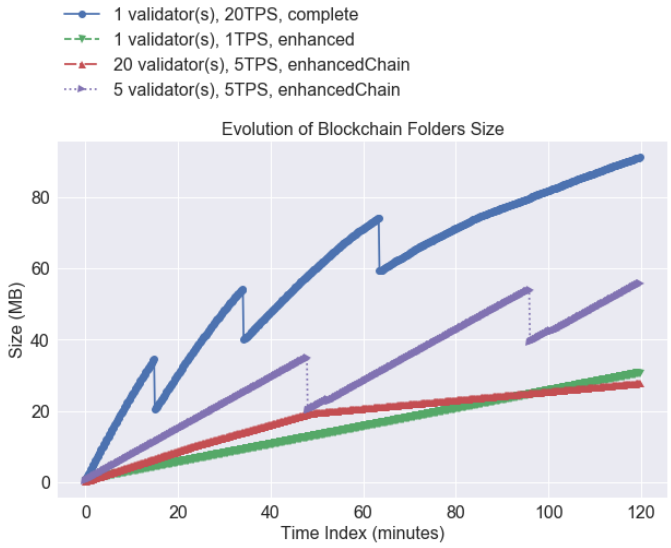


Figure 10. Blockchain folder data evolution

From the above results, it can be concluded that the performance of the Ethermint technology will certainly limit its usage in practice to specific use cases where the time required to valid a transaction is very high, and the number of validators joining the network is not very important. Besides, the storage space can also constitute a bottleneck for this technology to be used for real world applications. For instance, running the blockchain for several weeks will certainly be a problematic for certain use cases.

VI. Research Directions and Opportunities

Blockchain is currently under extensive research and development from both the academia and the industry, however, there are still major challenges to be overcome before mass market penetration and adoption. In this section, we highlight major research directions and opportunities that we believe are important to investigate.

A. Data Analysis and Visualization

A blockchain being no more than a ledger of transactions between accounts, data from a blockchain can be seen as no more than nodes connected by occasionally existing multi-property edges. Under which structural form should they be tackled depends on the aimed analysis. From a blockchain *network supervision* point of view, crucial in a private companies consortium, the relevant data aggregation level is the

block, with a time-series scheme. From the point of view of auditing the *quality of the user activity*, transactions should be considered the atomic level to investigate, under a graph scheme, and more specifically under a time-varying graph (TVG) scheme [52].

The aim of efficiently auditing a blockchain brings several challenges:

1) Real-time analysis: Because of the possibility of forks, there is no such thing as absolute reliability of the data retrieved from the blockchain. It is decreasingly high toward the most recent blocks data, as one only get the version of the ledger stored on a node at a given time, so that a blockchain-specific time-dependent reliability weight has to be determined. This procedure must be highly dependent on the chosen consortium governance scheme.

2) Exploitable visual representation of TVG: From a graph point of view, each edge (transaction) represents a unique and directed communication bridge between nodes, having an infinitesimally narrow time-width. To be able to graphically analyze a blockchain networks, or to compute common graph indicators such as centrality or community borders, systematic smart ways to define edges weight based on non-Dirac delta function in time have to be conceived.

3) Smart contract internal transactions unraveling: Unless explicitly coded as so, the transactions from and to smart contracts, or from smart contracts to users, are no written down in the ledger, and this can be used for transaction obfuscation allowing token laundering [53], Ponzi scheme [54] or other uses where the blockchain only serves itself. In order to determine whether or not blockchain transactions are related to real-world event, or more generally what it is used for, studies on specific key quality indicators related to smart contract have to be conducted.

B. Blockchain Audit

Data immutability is generally put forward when referring to blockchain technologies. However, as already discussed in Section II.A, the written data could still be tampered and the blockchain rebuilt as long as the majority of the participants (or miners) have reached a consensus. This is especially true in consortium and private blockchains where the number of miners is generally limited in comparison with public blockchains.

In this context, it becomes extremely difficult for a regulation authority to audit consortium based blockchains and to check whether the data and transactions have been tampered with or not. A commonly adopted solution consists in piggybacking data hashes from the consortium blockchain into the Bitcoin network, by embedding those hashes inside the OP_RETURN field of Bitcoin transactions. However, this contribute in polluting and increasing the size of the Bitcoin network with nonsense and non-financial data.

More recently, alternatives solutions have been proposed to reduce the impact of piggybacking on public blockchains, including the concepts of side-chains and notary chains whose main objective is to make it extremely hard for malicious users and/or the network participants to alter the blockchain data.

C. Governance

The governance in a private blockchain assigns authority and responsibility among the consortium members. It determines nodes that will be able to create blocks (*i.e.,* miners), to read/write data, to contribute in the consensus mechanism (*e.g.,* voting for a miner) and/or to participate in decisions for the system evolution (*e.g.,* operations management [55], software updates, allow new nodes to join the system etc.). This power distribution has an impact not only within the system but also on the business model of the use case.

Costs linked with the system activity such as the system set-up, its execution or maintenance are shared within the consortium according to the governance scheme. It also affects future incomes or losses at a business level since the governing nodes decide the rules of the system. For example, the majority of governing nodes can agree to allow the membership of a new entity within the consortium, and which is competitor of another member who has no power over this decision. Hence, this could jeopardize the viability of the system.

The viability of the system can also be affected by the governance definition. In many cases, to be durable, the consortium has to be able to grow by allowing new members to integrate the system. It is the case for example of new services over blockchain like dematerialized car service books. More companies join the consortium such as car manufacturers, car repair shops or insurance companies, the more durable and available is the system. On the other hand, the power is dissolved with the growth of consortium.

One should also take into account the impacts on the business model when building the governance scheme as it will be discussed in the next section.

D. Incentives and Business Models

Blockchain solves the issues of trust between actors in situations of exchange where the temptation of cheating is high by *removing* this need of trust. Any business model based on a solution that would not claim to solve a trust issue would inevitably fail, as its solution could be replaced by a less constraining and probably already existing centralized system.

In a blockchain whose users are exclusively individuals, the pecuniary incentives must ensure that, because members either receive additional incomes or just lessen their expenses, they find a financial interest in participating to the process.

However, in a consortium of commercial entities, it should be pointed out that the simple fact *not to* be part of the consortium might represent a handicap that could lead to loss of turnover or customers attrition, because of the latter attraction to blockchain promises and interest in financial incentives.

E. Data Privacy

Data privacy is an imperative for enterprise blockchains. But lets first distinguish anonymity and privacy. A transaction is considered anonymous if we cannot identify its owner, whereas a transaction is called private if the object and the amount of transaction are unknown.

We have seen many schemes on public blockchains to improve privacy: Stealth Addresses, Pedersen Commitments, Ring signature, Homomorphic encryption, Zero-knowledge-proof. No scheme can hide the sender, the receiver and the

amount at the same time, so we see actual implementations mixing these techniques in order to achieve the desired level of privacy. In addition, there are some known drawbacks such as computational time, so further research is needed. But we can expect that these initiatives on public blockchains will drive improvement on enterprise blockchains privacy as well. Interested readers may refer to [56] for more detailed information on privacy issues in blockchain.

F. Security

Guaranteeing End-to-End security means identifying vulnerabilities and mitigating risks at each element level and at the system level. This goes beyond looking at the blockchain building blocks (consensus, distributed network, cryptographic tools) and includes evaluating the virtual machine, the Smart Contracts, the Oracle, the user client, the hardware component, the keys management and PKI, etc. Some areas of research are the following: Formal verification of smart contracts, Usage of trusted platform modules for key storage, Identification of the different types of attack vectors and their counter strategies (sybil attacks, double spending attacks, distributed denial of service attacks, botnet attacks, storage specific attacks, censorship, *etc.*), Audit (detect issues a priori or a posteriori), Supervision (detect issues during run time). Interested readers may refer to [57] for more detailed information on security challenges in blockchain.

G. Scalability

As usual, there is always a trade-off between costs, security and performance. Because participants are known in enterprise blockchains, the scalability issue is therefore easier to solve, as compared to public blockchains. Yet, in order to achieve scalability, we first need to keep in mind the usage context and the performance metrics we want to optimize: transactions throughput, validation latency, number of participant nodes, number of validating nodes, energy costs, computation costs, storage costs or other criteria? As always, remember the trade-off principle: A round robin consensus algorithm will scale well, but the participants need to be honest. A PBFT algorithm can recover from malicious behaviors (up to 1/3) but the validating nodes should not be too many (tens of nodes at most) if the system is to work [58]. All in all, scalability is an active area of research and we can mention some initiatives such as: fragmenting the global ledger into smaller sub-ledgers run by sub-groups of nodes, removing old transactions in order to optimize the storage, using a hierarchy of blockchains (transactions are done at a higher level and settled optionally afterwards in the blockchain), and so on.

VII. Conclusions

The blockchain technology represents a major paradigm shift in the way business applications will be designed, operated, consumed and marketed in the near future. In this paper, we discussed in details the concept of consortium blockchains, in terms of architecture, technological components and applications. In particular, we analyzed and compared various consensus algorithms. An experimental study was then proposed in order to assess the performance of the Ethermint technology. The performance indicators analyzed are the time required to validate a transaction, the number of transactions validated per second, as well as the average inter blocks delay and the size of the blockchain data folder. Parameters that have been varied are the number of validators, the number of transactions submitted per second and the topology of the network. The results highlighted some limitations in terms of transactions validation time and storage requirements that may hinder the usage of Ethermint to deal with some real world use cases. Finally, we highlighted some major research challenges that need to be addressed before achieving mass market penetration, including the issues related to governance, audit, scalability, incentives, data privacy and security.

Acknowledgment

This research work has been carried out under the leadership of the Institute for Technological Research SystemX, and therefore granted with public funds within the scope of the French Program Investissements d'Avenir.

References

[1] E. Ben Hamida, K. L. Brousmiche, H. Levard, and E. Thea, "Blockchain for Enterprise: Overview, Opportunities and Challenges," in The Thirteenth International Conference on Wireless and Mobile Communications (ICWMC 2017), Nice, France, Jul. 2017.

[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008, accessed: 2017-05-25. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[3] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," IEEE Access, vol. 4, 2016, pp. 2292–2303.

[4] M. Mettler, "Blockchain technology in healthcare: The revolution starts here," in 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Sept. 2016, pp. 1–3.

[5] "stampd.io: A document blockchain stamping notary app," accessed: 2018-05-25. [Online]. Available: https://stampd.io/

[6] A. Yasin and L. Liu, "An online identity and smart contract management system," in 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), vol. 2, June. 2016, pp. 192–198.

[7] R. Dennis and G. Owen, "Rep on the block: A next generation reputation system based on the blockchain," in 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), Dec. 2015, pp. 131–138.

[8] F. Tian, "An agri-food supply chain traceability system for china based on rfid blockchain technology," in 2016 13th International Conference on Service Systems and Service Management (ICSSSM), June. 2016, pp. 1–6.

[9] K.-L. Brousmiche, T. Heno, C. Poulain, A. Dalmieres, and E. Ben Hamida, "Digitizing, securing and sharing vehicles life-cycle over a consortium blockchain: Lessons learned," in Proceedings of 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 1st International Workshop on Blockchains and Smart Contracts (BSC), Feb. 2018.

[10] Z. Zheng, S. Xie, H.-N. Dai, and H. Wang, "Blockchain challenges and opportunities: A survey," International Journal of Web and Grid Services, 2017, pp. 1–23.

[11] "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016," accessed: 2018-05-25. [Online]. Available: http://eur-lex.europa.eu/eli/reg/2016/679/oj

[12] V. Buterin, "On Public and Private Blockchains," 2015, accessed: 2018-05-25. [Online]. Available: https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/

[13] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in 2017 IEEE International Congress on Big Data (BigData Congress), June. 2017, pp. 557–564.

[14] "wiki: The Ethereum Wiki -," Feb. 2018, accessed: 2018-05-25. [Online]. Available: https://github.com/ethereum/wiki

[15] "Quorum | J.P. Morgan," accessed: 2017-07-01. [Online]. Available: https://www.jpmorgan.com/country/US/EN/Quorum

[16] "Ethereum," Feb. 2017, accessed: 2018-05-25. [Online]. Available: https://fr.wikipedia.org/w/index.php?title=Ethereum&oldid=134391073

[17] "Hyperledger Project," accessed: 2018-05-25. [Online]. Available: https://github.com/hyperledger/

[18] E. Buchman. Tendermint: Byzantine fault tolerance in the age of blockchains. Accessed: 2018-05-25. [Online]. Available: https://allquantor.at/blockchainbib/pdf/buchman2016tendermint.pdf (2018)

[19] N. Knezevic. A high-throughput byzantine fault-tolerant protocol. Accessed: 2018-05-25. [Online]. Available: https://infoscience.epfl.ch/record/169619/files/EPFL_TH5242.pdf (2018)

[20] Hedera hashgraph platform. Accessed: 2018-05-25. [Online]. Available: https://www.hederahashgraph.com/platform#speed (2018)

[21] O. Band. Stellar load testing results for the kin ecosystem. Accessed: 2018-05-25. [Online]. Available: https://medium.com/kin-contributors/stellar-load-testing-results-for-the-kin-ecosystem-64c4d8676e69 (2018)

[22] Get started with the basics of the stellar network. Accessed: 2018-05-25. [Online]. Available: https://www.stellar.org/how-it-works/stellar-basics/ (2018)

[23] L. Chen, L. Xu, N. Shah, Z. Gao, Y. Lu, and W. Shi, "On security analysis of proof-of-elapsed-time (poet)," in Stabilization, Safety, and Security of Distributed Systems, P. Spirakis and P. Tsigas, Eds. Cham: Springer International Publishing, 2017, pp. 282–297.

[24] Multichain developers q&a: About throughput performance. Accessed: 2018-05-25. [Online]. Available: https://www.multichain.com/qa/5556/about-throughput-performance (2018)

[25] D. Ongaro. Consensus: Bridging theory and practice. Accessed: 2018-05-25. [Online]. Available: https://ramcloud.stanford.edu/~ongaro/thesis.pdf (2014)

[26] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," ACM Transactions on Programming Languages and Systems (TOPLAS), vol. 4, no. 3, 1982, pp. 382–401.

[27] M. Pease, R. Shostak, and L. Lamport, "Reaching agreement in the presence of faults," Journal of the ACM (JACM), vol. 27, no. 2, 1980, pp. 228–234.

[28] M. Castro and B. a. Liskov, "Practical Byzantine fault tolerance," in Third Symposium on Operating Systems Design and Implementation (OSDI), vol. 99, 1999, pp. 173–186.

[29] J. P. Morgan. Quorum description. Accessed: 2018-05-25. [Online]. Available: https://github.com/jpmorganchase/quorum/wiki (2018)

[30] J. Kwon, "Tendermint: Consensus without mining," 2014, accessed: 2018-05-25. [Online]. Available: https://tendermint.com/static/docs/tendermint.pdf

[31] E. Buchman, "Tendermint: Byzantine fault tolerance in the age of blockchains," Ph.D. dissertation, 2016, accessed: 2018-05-25. [Online]. Available: https://allquantor.at/blockchainbib/pdf/buchman2016tendermint.pdf

[32] Tendermint's software ecosystem. Accessed: 2018-05-25. [Online]. Available: https://tendermint.com/ecosystem (2018)

[33] V. Costan and S. Devadas, "Intel sgx explained." IACR Cryptology ePrint Archive, 2016, p. 86, accessed: 2018-05-25. [Online]. Available: https://eprint.iacr.org/2016/086.pdf

[34] Intel. Hyperledger Sawtooth description. Accessed: 2018-05-25. [Online]. Available: https://sawtooth.hyperledger.org/docs/core/releases/latest/introduction.html (2018)

[35] D. Mazieres, "The stellar consensus protocol: A federated model for internet-level consensus," 2015, accessed: 2018-05-25. [Online]. Available: https://www.stellar.org/papers/stellar-consensus-protocol.pdf

[36] L. Baird, "Hashgraph consensus: fair, fast, byzantine fault tolerance," Swirlds Tech Report, Tech. Rep., 2016, accessed: 2018-05-25. [Online]. Available: http://www.swirlds.com/wp-content/uploads/2016/06/2016-05-31-Swirlds-Consensus-Algorithm-TR-2016-01.pdf

[37] I. Swirlds. Swirlds website. Accessed: 2018-05-25. [Online]. Available: https://www.swirlds.com/ (2018)

[38] "MultiChain | Open source private blockchain platform," accessed: 2018-05-25. [Online]. Available: http://www.multichain.com/

[39] C. Cachin and M. Vukolić, "Blockchains consensus protocols in the wild," arXiv preprint arXiv:1707.01873, 2017, accessed: 2018-05-25. [Online]. Available: https://arxiv.org/pdf/1707.01873.pdf

[40] D. Ongaro and J. K. Ousterhout, "In search of an understandable consensus algorithm." in USENIX Annual Technical Conference, 2014, pp. 305–319.

[41] "The raft consensus algorithm," accessed: 2018-05-25. [Online]. Available: https://raft.github.io/

[42] B. Parno, C. Gentry, J. Howell, and M. Raykova, "Pinocchio: Nearly practical verifiable computation," Cryptology ePrint Archive, Report 2013/279, 2013, accessed: 2018-05-25. [Online]. Available: https://eprint.iacr.org/2013/279

[43] R. Mercer, "Privacy on the blockchain: Unique ring signatures," arXiv preprint arXiv:1612.01188, 2016, accessed: 2018-05-25. [Online]. Available: https://arxiv.org/pdf/1612.01188.pdf

[44] M. Raikwar, S. Mazumdar, S. Ruj, S. S. Gupta, A. Chattopadhyay, and K.-Y. Lam, "A blockchain framework for insurance processes," in Proceedings of 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 1st International Workshop on Blockchains and Smart Contracts (BSC), Feb. 2018.

[45] J. Horta, D. Kofman, D. Menga, and A. Silva, "Novel market approach for locally balancing renewable energy production and flexible demand," Dresden, Germany, Oct. 2017.

[46] N. Lasla, M. Younis, W. Znaidi, and D. Ben Arbia, "Efficient distributed admission and revocation using blockchain for cooperative its," in Proceedings of 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 1st International Workshop on Blockchains and Smart Contracts (BSC), Feb. 2018.

[47] Ethermint. Ethermint project description. Accessed: 2018-05-25. [Online]. Available: http://ethermint.readthedocs.io/en/develop/ (2017)

[48] Hyperledger. Hyperledger project description. Accessed: 2018-05-25. [Online]. Available: https://www.hyperledger.org/projects/hyperledger-burrow (2017)

[49] K. Driscoll, B. Hall, H. Sivencrona, and P. Zumsteg, "Byzantine fault tolerance, from theory to reality," in International Conference on Computer Safety, Reliability, and Security. Springer, 2003, pp. 235–248.

[50] Cosmos. Cosmos white paper. Accessed: 2018-05-25. [Online]. Available: https://cosmos.network/about/whitepaper (2017)

[51] OpenStack. OpenStack description. Accessed: 2018-05-25. [Online]. Available: https://www.openstack.org/ (2018)

[52] A. Casteigts, P. Flocchini, N. Santoro, and W. Quattrociocchi, "Time-varying graphs ans dynamic networks," International Journal of Parallel, Emergent and Distributed Systems, vol. 27, no. 5, 2012, pp. 387–408.

[53] M. Moser, R. Bohme, and D. Breuker, "An inquiry into money laundering tools in the bitcoin ecosystem," in eCrime Researchers Summit (eCRS), 2013. IEEE, 2013, pp. 1–14.

[54] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia, "Dissecting ponzi schemes on ethereum: identification, analysis, and impact," CoRR, vol. abs/1703.03779, 2017.

[55] T. Sato and Y. Himura, "Smart-contract based system operations for permissioned blockchain," in Proceedings of 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 1st International Workshop on Blockchains and Smart Contracts (BSC), Feb. 2018.

[56] M. Conti, S. K. E, C. Lal, and S. Ruj, "A survey on security and privacy issues of bitcoin," CoRR, vol. abs/1706.00916, 2017.

[57] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," CoRR, vol. abs/1802.06993, 2018.

[58] M. Vukolic, "The Quest for Scalable Blockchain Fabric: Proof-of-Work vs. BFT Replication," Open Problems in Network Security (iNetSec), 2015, pp. 112–125.

Context Aware Control Schemes for the Performance Improvement of V2X Network Slices

Alexandros Kaloxylos
Department of Informatics and Telecommunications
University of Peloponnese
Tripoli, Greece
email: kaloxyl@uop.gr

Prajwal Keshavamurthy, Panagiotis Spapis, Chan Zhou
Huawei German Research Center
Munich, Germany
email: prajwal.keshavamurthy@huawei.com
panagiotis.spapis@huawei.com
chan.zhou@huawei.com

***Abstract*— Network slicing for 5th Generation (5G) networks enables the support of multiple logical networks, called slices, which aim to be tailor-cut network solutions for specific services for the vertical industries (e.g., transportation, smart factories, health industry etc.). Although, considerable effort has been taken to define the generic framework for network slices, it still remains open how the network performance can be further optimized by taking into consideration the specificities of each use case. At the same time, the work for the specification of network functions to support autonomous driving is picking up speed. However, up to this moment, it is still not addressed how contextual information can serve the optimization of a vehicle-to-everything (V2X) slice. This paper provides in detail the latest status of the 3GPP standardization process related to slicing. It also introduces two new mechanisms called Context Enhanced MOBility management (CEMOB) and Context-Aware Resource Pre-allocation (CARP). The former improves the existing mobility management process while the latter serves the minimization of the communication delay among vehicles. The point we make with these two mechanisms is that by taking advantage of contextual information the performance of network control functions can be significantly improved. Towards this end, we quantify the merits of our mechanisms and we present how these are integrated into a V2X slice.**

Keywords-network slicing; mobility management; pre-allocation of resources; V2X communications.

I. INTRODUCTION

Using contextual information for future mobile networks has become lately a hot topic [1]. 5G networks target, apart from the support of the telecommunications sector, also the communication needs of "vertical industries" like autonomous driving in transportation, smart factories, new health services, etc. An extensive list of 5G use cases can be found in [2] and [3]. A thorough examination of the verticals has identified that these sectors have diverse requirements. These requirements are mapped to different network Key Performance Indicators (KPIs). The KPIs indicatively include throughput, transmission reliability, latency, energy consumption, blocking probability, etc. Services and applications for the vertical industries have different requirements and thus, different values for the aforementioned KPIs. It is widely accepted that no single network can support efficiently all these different use cases.

Thus, it appears that the deployment of parallel logical networks over the same network infrastructure is a necessity. These logical networks may have network functions (NFs) configured differently or even introduce new network functions both in the Radio Access Network (RAN) [4] as well as the Core Network (CN) [5].

The 3rd Generation Partnership Project (3GPP) has defined a network slice to be " A logical network that provides specific network capabilities and network characteristics" [6]. A "Network Slice" is implemented by a "slice instance" that in its turn is created by a "network slice template". The latter is a template that defines a complete logical network including the NFs, their interfaces and their corresponding resources.

Network slicing has been intensively investigated during the past years both by industry and academia. There are several research proposals that target full flexibility in terms of selecting, organizing and deploying NFs [7]. At the same time, 3GPP has already delivered the first phase specifications for 5G networks that include also the support for slicing. The standardization activities have followed a sensible path and have re-used existing NFs or share NFs across different slices as much as possible, focusing essentially on the enhanced Mobile BroadBand (eMBB) slice. The use cases to be supported, as well as their requirements, have been thoroughly studied [8], but current specifications do not provide fully tailor-cut solutions for them. In order to do this, it is needed to work really closely with the representatives of the so called "vertical industries" (e.g., transportation, health, factories, energy). This is needed to understand not only the requirements and the operational environment, but also the contextual information produced and how this can be used to optimize network functions.

For example, the newly founded 5G Automotive Association (5GAA) [9] is working towards such a direction. Still, the activities for proposing mechanisms driven by such organizations, that are expected to affect the standardization process, are in primitive steps.

In the current paper, extending our previous work presented in [1], we present the latest status of the standardization activities related to network slicing. We also

provide two novel mechanisms for vehicular communications, which can be easily deployed using slicing solutions. The first one is a new mobility management mechanism for autonomously driven vehicles. It takes advantage of contextual information that is possible to be used by the standardized 5G NFs. The second is a resource pre-allocation scheme that uses available contextual information to meet the stringent requirements of certain V2X use cases, by minimizing the communication delay among vehicles. These are exemplary schemes to highlight that different use cases need very different solutions. Thus, we believe that it is important that solution providers take into consideration the specificities of each use case. We also present how the new mechanisms can be supported by 5G networks.

The rest of the paper is organized as follows. In Section II, we provide the latest status of 3GPP in relation to slicing. Section III discusses how mobility management is planned to be supported in the technical specifications and why we consider this not to be efficient for moving vehicles. In Section IV, we provide the details of CEMOB on how to extend the 5G network functions to improve mobility management for moving vehicles. In Section V, we present quantitative results that illustrate the benefits of our scheme. In Section VI, we discuss how the allocation of resources takes place in 5G networks and what is the expected delay, while in Section VII we present the CARP mechanism, that minimizes the communication delay among vehicles and we analyse its performance improvements. In Section VIII, we summarize the key findings of the paper. Finally, Section IX concludes the paper and describes future directions.

II. SLICE SUPPORT IN 3GPP

3GPP has decided to treat 5G specifications in two phases. The first one is just recently completed (Release 15). This phase addresses a more urgent subset of the commercial needs. Phase 2 is to be completed by March 2020 (Release 16) for the IMT 2020 submission, having addressed all identified use cases & requirements. In relation to slicing, several working groups are currently progressing on the key elements and procedures that have to be specified.

In [6] and [10], the 5G network architecture is presented. There, a list of technical key issues, as well as potential solutions for slicing are presented. For example, in these documents the issues of slice selection, slice isolation, sharing of NFs among slices, multi-slice connectivity, management of slices, etc. are being addressed.

The first set of specifications has addressed a number of key principles. The first principle is that NFs, previously incorporated into monolithic network components, are now decomposed to smaller modules. The target is to allow a synthesis and configuration of the NFs on a per slice type basis. A second principle is the further splitting of user and control plane functions to facilitate a more flexible evolution of NFs. A third key principle is the exposure of NFs to external services through appropriate APIs. This is expected to allow a better collaboration among network operators and service providers.

Figure 1 presents a summary of the supported NFs. The control plane functions in the CN are considered to be the following:

- **Unified Data Management (UDM):** supports the Authentication Credential Repository and Processing Function (ARPF).
- **Authentication Server Function (AUSF):** supports the authentication of end users
- **Policy Control function (PCF):** supports unified policy framework to govern network behaviour and provides policy rules to control plane functions
- **Core Access and Mobility Management Function (AMF):** supports mobility management, access authentication and authorization, security anchor functions and context management
- **Session Management Function (SMF).** supports session management, selection and control of UP functions, downlink data notification and roaming
- **User Plane Function (UPF):** is the anchor point for inter/intra RAT mobility, supports packet routing and forwarding, QoS handling for user plane, packet inspection and policy rule enforcement
- **Network Exposure Function (NEF):** provides a means to securely exchange information between services and 3GPP NFs.
- **NF Repository Function (NRF):** maintains the deployed NF Instance information when deploying/updating/removing NF instances
- **Network Slice Selection Function (NSSF):** supports the functionality to bind a UE with a specific slice.

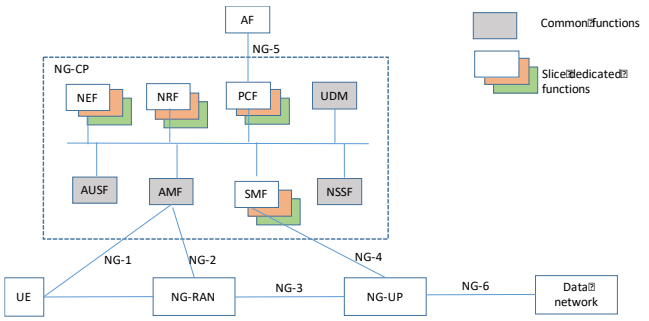


Figure 1: 5G service based architecture (adapted from [6])

Note that some of these functions are common for all slices, while others can be dedicated for different slices. A User Equipment (UE) may access multiple slices concurrently via a single RAN. For such cases, it is assumed that the involved slices should share some control plane functions, like the AMF. The abovementioned logical network allows the support of Application Functions (AF) and provides connectivity to typical external data networks.

Moreover, it has been agreed that RAN will be slice-aware so as to treat slice traffic according to the customer needs. Moreover, RAN shall support resource isolation among slices so as to avoid shortage of shared resources in

one slice to break the service level agreement on another [11].

Detailed alternative solutions have been proposed on how RAN is involved in slice selection by passing an appropriate identifier to the core network elements. Currently, slicing for RAN essentially focuses on different scheduling schemes for various slices and also by providing different L1/L2 configurations. Moreover, it is considered that even if a UE is connected to multiple slices, a single Radio Resource and Control (RRC) entity will be used. Other radio access protocols (i.e., Packet Data Convergence Protocol – PDCP and Radio Link Control - RLC) can used on a per slice basis.

Every slice is identified by a Single Network Slice Selection Assistance Information (S-NSSAI) identifier. This identifier consists of a Slice/Service Type (SST) and a Slice Differentiator (SD). The former defines essentially the features and network services to be offered by a slice, while the latter is used to select among different slices of the same type. Currently, only 3 SST values have been agreed to be supported. These are a) eMBB, b) Massive Internet of Things (MIoT), and c) Ultra Reliable Low Latency Communications (URLLC) [6]. This information is exchanged as part of non-access stratum signalling through the RAN.

In [12], the life cycle of a network slice is described by the following phases: a) Preparation phase, b) Instantiation, Configuration and Activation phase, c) Run-time phase and d) Decommissioning phase.

Overall, 3GPP has defined the framework for slice deployment, operation and selection. However, there are no detailed solutions about how each slice type will be different from another. This is crucial gap that has to be addressed. In order to achieve the desired performance for each slice, new mechanisms are needed. These mechanisms must take advantage of the characteristics and the environment where the slices for the vertical industries will be used. As we will present in the following chapters, taking advantage of contextual information of autonomously driven vehicles (e.g., the street geography, the path to be followed by a vehicle), one can improve considerably control functions like mobility management and decrease the communication delay for critical services like collaborative collision avoidance.

III. CURRENT STATUS FOR MOBILITY MANAGEMENT IN 5G NETWORKS

Mobility management for legacy systems is performed as follows. The network is divided into non-overlapping regions called Tracking Areas (TAs). Idle UEs have to inform the network each time they cross the border of such areas or when a timer, typically set at 54 minutes, expires. However, this design was initially static and the cost for re-arranging the coverage areas of TAs was quite high. Moreover, a problem appeared from excessive Tracking Area Update (TAU) messages due to the movement of the users near the TA borders. That is why the notion of Tracking Area Lists (TAL) was introduced. TALs were assigned on per UE basis and allowed the overlapping of TAs. The algorithm to define the TAL is proprietary and the operator decides according to his strategy whether to allocate large or short TALs for each UE. Whenever a UE has to be discovered (e.g., delivering data to it, incoming call etc.), paging is executed in a subset or all the cells inside a TAL according to the operator's strategy [13]. If a subset of the cells of the TAL is paged there is a risk of increased delay due to page misses. On the other hand, if all the cells are paged, there is an increased signalling cost. The size of the TAL relates to a signalling tradeoff. Small TALs have reduced paging signalling cost but require frequent TAU. If large TALs are used, the signalling cost is high but fewer TAU notifications are needed.

Even with these improvements, it has been noticed that whenever idle UEs switch into connected mode, signalling has again to be exchanged up to the core network and more specifically the Mobility Management Entity (MME) in 4G networks or the AMF in 5G networks. Inside the MME or AMF, contextual information for each UE (such as security credentials) is kept. Considering that smartphones have a number of applications (e.g., Facebook, Skype, Viber, etc.) that wake up asynchronously and exchange small amount of information, this creates a significant signalling load for the aforementioned network components.

This is why, for the 5G systems mobility for idle terminals had to be redesigned [11]. In the latest specifications, the RAN-based Notification Area (RNA) has been defined. This can be considered as a smaller subset of a TAL and consists of a number of base stations (called gNBs in 5G terminology). While inside an RNA, an idle UE can move from one gNB to another, without informing the network about its exact location. Also, a new state called RRC_INACTIVE is introduced. Whenever a UE is in this state, then its context information is kept locally at its last serving gNB. Thus, a UE avoids contacting the CN entities (i.e., AMF) whenever the UE switches again to the connected mode. This addresses the needed minimization of signalling load caused by the frequent waking-up of end devices (e.g., smartphones) towards the CN.

If the UE wakes up and becomes connected under a new gNB inside the same RNA, then it uses the *RRCConnectionResume* message to force the new gNB retrieve its context from the last serving gNB. The new gNB may also trigger a path switching by communicating with the AMF. Paging a UE takes place from the last serving gNB to all gNBs that are members of the RNA. These procedures are illustrated in Figure 2. On top of these messages, note that whenever a UE crosses the RNAs borders it needs to receive the gNBs identifiers that are members of the new RNA.

This mechanism treats indeed several of the inefficiencies present in existing cellular systems like 4G mobile networks. However, as explained in [14], the RAN based mobility management scheme suffers from excessive load for high moving UEs. This is why, Hailu and Säily [14] suggest a hybrid scheme where a typical *CN mobility management* takes places for high moving UEs, while a *RAN based mobility* management is executed for slow moving UEs. To achieve this, the UEs have to report their mobility status to the CN at some intervals or during specific events

(e.g., during location update). Moreover, the authors also indicate potential delay issues that may arise if there is no direct interface between the last serving gNB and the new one. In such a case, signalling between gNBs has to travel through the CN. The lack of a direct link between base stations is not uncommon in commercially deployed mobile networks. Note that in the current standard specification both the typical CN as well as the RAN mobility management scheme are supported.

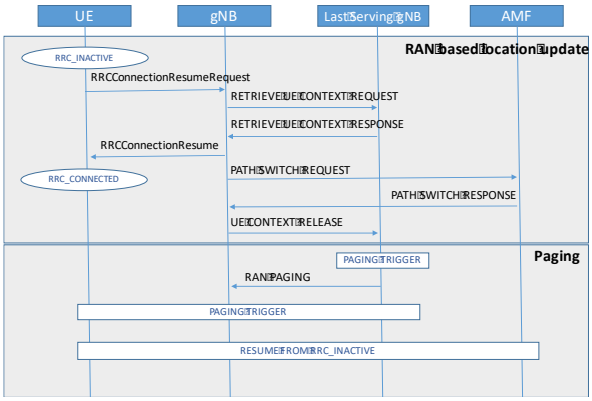


Figure 2: RAN based mobility management (adapted from [11])

Based on the above discussion, it is clear that the adoption of the RAN based mobility management scheme will be beneficial for some of the 5G use cases but inefficient for others. An example of a non-applicable use case is the one of the autonomously driving vehicles. This is because vehicles are expected to change their velocities quite often when moving for example inside an urban environment. Having the vehicles reporting their mobility status frequently, it will cause an excess signalling overhead to the network. On the other hand, the CN mobility management scheme will also suffer, as we have explained, from frequently awaking vehicles that will want to exchange information with their neighbours for a short period (e.g., to perform a manoeuvre). To optimize a control procedure like mobility management for moving vehicles, one has to take advantage of contextual information that can be easily available to the operator as we will discuss in the next section.

IV. CEMOB: CONTEXT ENHANCED MOBILITY MANAGEMENT

A. Algorithm Description

Autonomous driving is one of the key targets of the industry for the next decade. 3GPP has already specified an architecture and the related mechanisms to support inter-vehicle communication as well as their access to service specific servers (i.e., V2X application server - [15]). The support of such services introduces additional contextual information that if used, it can greatly improve the network control operations for a mobile network. More specifically, it is expected that in order to form a route, a vehicle will communicate with a server to receive the path to be followed. These servers can also estimate the time a vehicle will need to be at a certain position in the path. Such functionality exists even today with well-established applications like Google maps or any other GPS navigators. Obviously, these applications are unaware about the deployed base stations of a mobile operator. However, for 5G networks passing a route information to an operator, it is going to be an easy task to perform.

As we discussed in Section II, the NEF allows for Service providers (e.g., Google maps) to communicate this path the mobile network in a secure way. A translation function is then able to transform path coordinates to a list of gNBs that will serve the UEs when they reach specific areas at specific times. Furthermore, the specific geography of the roads can significantly assist in determining the exact cells a vehicle is going to pass through. Such information can be used to considerably optimize the mobility management procedure by optimizing the TALs allocation and at the same time improving the paging strategy. Additionally, the modularization of 5G network functions facilitates their optimum placement in the RAN or CN network components. For the V2X case it makes sense to keep part of the mobility management functionality close the moving vehicles (i.e., at the gNBs), since unnecessary frequent communication with CN entities like the AUSF can be avoided.

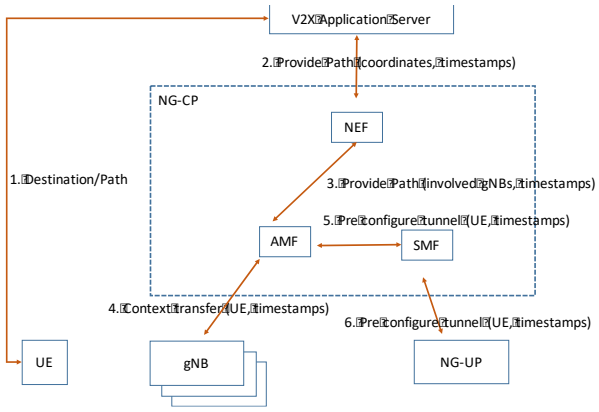


Figure 3: Mobility management for vehicles in 5G networks

In Figure 3 we present how a new mobility management scheme called CEMOB (Context Enhanced MOBility management). It is designed especially for vehicles operating inside 5G networks.

Whenever a UE/vehicle wants to reach a specific destination, it will communicate with a V2X application server and it will receive the path so as the computer inside the car will start the autonomous driving functions (step 1). Upon calculation of such a path by the V2X application server, the information in terms of coordinates and timestamps (time when the vehicle will be at a specific point) can be communicated to the mobile operator. This will take place through the NEF entity (step 2). The NEF can also translate the coordinates into specific gNBs and forward further this information to the involved AMFs (step 3). These entities on their turn can transfer the UE context to the involved gNBs (step 4). Moreover, they will communicate

with the corresponding SMFs so as to pre-configure the data path for the vehicles (steps 5 and 6). Note that this pre-configuration does not imply that resources will be allocated for large period of times but rather only for a short time for which a vehicle is expected to be in a certain area. Obviously, it is possible that a vehicle (or the respective V2X application server) may be required, due to traffic conditions, to modify and re-calculate a path. Such information will again be communicated through the NEF entity and the new information will be passed to all involved gNBs.

Note that the communication of a UE with a V2X application server located inside the domain of the mobile operator, can take place in terms of a few tenths of millisecond [16]. Thus, any updating of network components by the server will take place very rapidly. During such short time, a vehicle will not have changed its position no more than a few meters. So, any mobility management action, like paging is not going to be seriously affected.

Although Figure 3 illustrates the placement of the AMF inside the CN, part of its functionality can be placed in the gNBs. Consider for example the case where a UE/vehicle wants to communicate with a neighbouring one inside its own RNA. If part of the AMF functionality is placed at the gNB level, the communication request will stop at the serving gNB of the calling UE. The serving gNB's mobility management function will perform the paging to the called UE/vehicle. To do this, it will send a paging message to its cell as well the neighbouring ones, since it is already aware of the vehicles that are under the RNA vicinity during a specific time.

Since the actual location of the communicating UEs is well known with a pretty good accuracy, there is no need to communicate with the CN NFs to acquire a larger searching area (i.e., TAL). Also, there is no need for transferring the UEs' context information in the RAN in a reactive manner. This information is pre-fetched in the gNBs during the execution of step 4, as presented in Figure 3.

The benefits of CEMOB are manifold. Firstly, the mechanism is fully optimized for moving UEs independently of their speed. Firstly, it is not necessary to communicate with the network the UE's mobility status. Also, tt is not necessary to revert to the typical CN mobility management scheme if a vehicle's speed is high and the network signalling reaches a high paging load. Similarly, there is no need to switch to the RAN based mobility management scheme at a low moving speed.

Secondly, there is no need to exchange control messages for UE location updates (i.e., TAU) over the wireless interface which is the bottleneck for any wireless system. Note that the execution of a typical TAU message exchange requires the communication of a considerable number of signalling messages as described in [10].

In the case of CEMOB, the delay for transferring the context information of a UE from a serving to a new gNB is zero, since this information is in place beforehand. This delay in the RAN based mobility management scheme can be significant, as we have already explained for the cases where the gNBs have no direct interface and their communication takes place through the CN.

The paging cost for CEMOB is significantly lower than the CN and the RAN based mobility management schemes. The already known geography of the streets can minimize the number of cells that need to be paged only to the few ones that are serving street segments. All the aforementioned benefits are possible because CEMOB takes advantage of service related contextual information that can be available to the NFs of the mobile operator in a standardized way.

Finally, note that the modularized architecture and the slice support allow different NFs to be used for different logical networks (i.e., slices) and even place them at different network components. This means that CEMOB may be used only for the V2X communications network slice. Other slices, like the eMBB may use the existing solutions for mobility. This is possible since network slices can be configured differently for each use case.

B. Performance Analysis

To evaluate the performance of CEMOB we compare it with the CN and RAN based mobility management schemes. In order to calculate the signalling cost during paging, we follow the analysis presented in [14]. Let M be the number of cells and N the number of gNBs. As an exemplary analysis, we consider 3 cells to be supported by a single gNB. The RAN based scheme requires M messages to be transmitted over the radio link, plus N-1 messages to be transmitted from the last serving gNB to the neighbouring gNBs located inside the same RNA. As for the CN based mobility management scheme, M messages need to be transmitted over the radio interface. Additionally, N messages will be sent from the CN to the gNBs as well as 6 additional messages are exchanged on a per UE basis to inform the CN NFs that a UE is currently in the RRC_INACTIVE state [10].

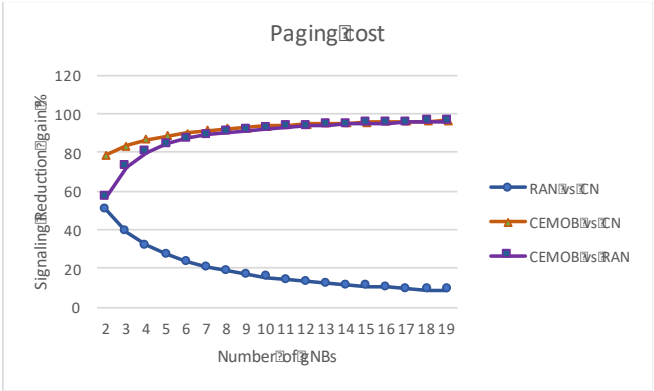


Figure 4: Paging cost CEMOB vs. RAN based vs CN based

Concerning CEMOB, the knowledge of the position of a UE with a high accuracy, even under some time coarse time period, requires paging only the gNB where the vehicle is camped under it. Also, knowing the topology of the streets and the direction of the vehicle, it is easy to make sure that there will be no page miss, by also paging the previous and the following gNBs from the estimated camped gNB.

Considering an inter site distance among gNBs of even 500m, the vehicle is paged in an area of 1.5 km that makes the probability of success rather high.

As shown in Figure 4, as long as the number of gNBs increases, the signalling reduction gain of the RAN-based mobility management scheme, compared to CN based one, is rather low. On the other hand, CEMOB outperforms these two schemes considerably since we take advantage of the accurate information about the location of the UE/vehicle CEMOB's relative gain is improved when the number of gNBs in an area increases since the baseline mobility management schemes need to page a larger number of gNBs.

To estimate the number of messages to be exchanged during a location update we perform the following analysis. As shown in Figure 2, for the RAN based scheme, 7 messages need to be exchanged every time a UE crosses the border of an RNA or when it resumes an RRC connection in a gNB that is different from the last serving gNB. A similar number of messages is needed for the CN based scheme, but this time the communication takes places between a gNB and AMF, instead of the last serving gNB. For the CEMOB case, the UE context needs to be transferred to all gNBs of an area before the UE enters into it. Also, in case a UE selects with a probability p, a different path for any reason, then it will communicate again with the V2X application server and the context will have to be updated again to all the gNBs of an area.

To perform an evaluation of CEMOB for the signalling load we consider an area of 15 gNBs containing 3 RNAs. We also consider that a street has two lanes. According to [17], the vehicle traffic flow with *measurement at a point* is "the number of vehicles that passes a point on a highway or a given lane or direction of a highway during a specific time interval". Traffic flow q is expressed in vehicles/hour and is given by:

$$q = \frac{n_t}{t} \quad (1)$$

where n_t is the number of vehicles passing a particular point in a defined period t. Related to the flow of vehicles the space headway parameter can also be used to derive q [17]. The average space headway $\overline{hs}$ is defined as the distance measured between the front ends of two successive vehicles (as the sum of the vehicles' in-between space and a vehicle's length). Based on this parameter the traffic flow can be calculated as:

$$q = \frac{\bar{v}}{\overline{hs}} \quad (2)$$

where the flow q is calculated as the average speed $\bar{v}$ of the vehicles divided by their average space headway. Based on this, we are able to calculate the traffic flow of vehicles passing through the 3 RNAs border areas per hour. Our assumption is that for the baseline schemes (i.e., CN and RAN based), a UE will resume its connection once every 5 cells. Having also a fixed road topology and assuming a uniform distribution of vehicles with fixed space headway distance among them, it is easy to calculate the number of vehicles in this area. Using this number, we can select a probability that some of the vehicles will change their path, so CEMOB will have to update all the gNBs of an RNA.

Figure 5 presents the results for different vehicle speeds (from 20 to 60km/h) and different space headways (from 4.5 to 22.5 meters). For this experiment, we consider that every 30 sec the 20% of the vehicles will request a path update.

As seen from the figure, CEMOB significantly outperforms the baseline schemes. The reason is that the on-demand context transfer requires a lot of signalling even if this is requested from one gNB to another inside an RNA. In such cases, the CN has to be notified so that path switching is performed.

On the other hand, CEMOB has to notify the gNBs only once and pre-configure the RAN-CN communication path at the same time. For a small number of cells, even for the exemplary topology under consideration, this means a considerable signalling reduction. Although CEMOB needs to update the gNBs every time a UE changes its path, this cost can be minimized by selecting a subset of gNBs to be updated at a time (e.g., only the time relevant part of the end-to-end path). In the case of the baseline schemes, the signalling cost is heavily affected by a complex process that may take place every time a UE is paged or whenever it switches from and idle to a connected state.

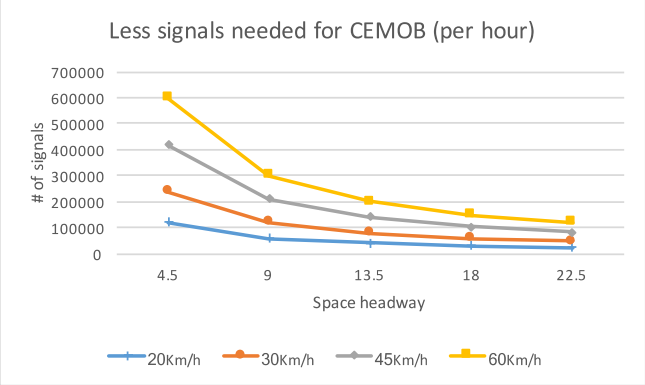


Figure 5: Signaling comparison between CEMOB and baseline scheme

Obviously, the penalty for CEMOB is the transfer of contextual information to many more gNBs (all the gNBs inside an RNA) compared to the baseline schemes where this information is transferred only from one gNB to another. To evaluate this penalty, we present the following analysis.

According to [18], the security information that needs to be transferred to the gNBs and is part of the contextual information is approximately 624 bits and consists of a) K-ASME key (256 bits), b) K-eNB key (256 bits) and c) NONCE (32 bits). Also the Globally Unique Temporary UE Identity GUTI (80 bits). This information needs to be transferred the corresponding gNBs that are involved either in the RAN based mobility management scheme or the CEMOB mechanism.

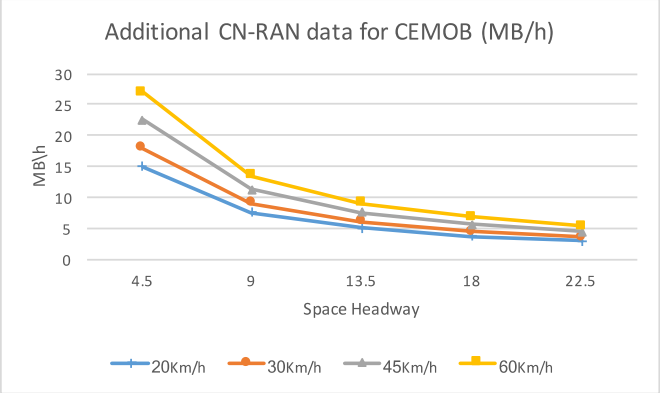


Figure 6: Additional data transfer needed for CEMOB

In Figure 6 we present the additional information needed to be transferred for CEMOB when compared to the baseline scheme in terms of MB/h for a fixed topology and different space headway among vehicles. The settings of this experiment were the same with the previous one (e.g., number of gNBs, size of an RNA, probability of changing path, etc.). As expected CEMOB always underperforms compared to the baseline scheme. The additional overhead of CEMOB for transferring contextual information is minimized when the space headway value increases since less vehicles are moving on the street and participate in mobility management functions. In all cases the additional amount of information that needs to be transferred over the wireline CN-RAN link for the case of CEMOB seems to be rather manageable for existing mobile networks. As shown in Figure 6 the worst case for CEMOB is for an average space headway of 4.5 meters for vehicles travelling at 60 Km/h. For this case only an additional 27 Mbps needs to be exchanged between the AMF and the gNBs over the wired part of the network.

Overall, CEMOB minimizes the signalling load and the interactions among NFs for mobility management procedures by taking advantage of contextual information that is related to the specificities of the V2X use cases. As we will demonstrate in the next Section the same principle may be applied to other network control functions like the management of network resources in a way that minimizes the communication delay among vehicles. This minimization is of paramount importance for autonomous driving applications.

V. Resource Allocation in Existing Systems

In cooperative automated driving (CAD) communications, vehicles need to communicate under strict delay and reliability constraints. In the existing schemes, a UE must obtain the resources from the scheduler in order to communicate. This procedure consists of the following steps (illustrated in Figure 7):

a) scheduling request

b) scheduling grant

c) UE processing

The average duration of the overall procedure involved in obtaining the first schedule grant is about 10msec [19] thus, failing to meet the delay requirements of various V2X use cases such as cooperative collision avoidance, cooperative lane change, emergency trajectory alignment that may require an end-to-end delay in terms 3-10msec [20].

On the other hand, the delay for a UE to be granted resources for transmission is linked with the transition from RRC IDLE state to RRC CONNECTED. Table I presents the control-plane delay budget for moving from IDLE to CONNECTED which in total is approximately 50msec. In addition, 3msec is required for resources to be granted by the scheduler.

The solution using the RNA concept [14], described in Section IV, introduces a "light" connected state (i.e., RRC_INACTIVE), where the UE is able to resume a connection with the *RRCConnectionResume* message. This procedure requires about 30msec for entering CONNECTED state from the light connected state (since interactions with the core network are omitted) [21]. Still, this solution cannot address the abovementioned strict delays.

If all vehicles are always CONNECTED even though the new transmissions will require only 3msec to transmit their data, resources are wasted for having a signalling channel ready for the usage, even when the transmissions are not planned. Considering the number of vehicles in the street this can be a significant waste of resources. In case of downlink communication, the scheduling delay is insignificant, but the delay associated with the paging needs to be taken into account if the vehicle is in IDLE mode. Furthermore, in case where a vehicle crosses the boundaries of two cells, a handover has to be executed. The execution of a typical handover process requires about 30-50msec. Again, such a delay is not acceptable based on the latest specifications of 3GPP.

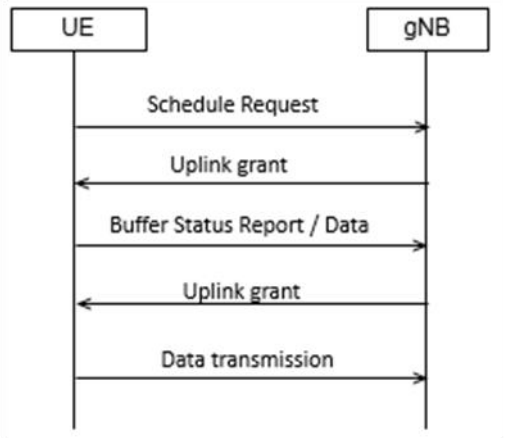


Figure 7: UE obtaining resources from the gNB

Summarizing, the abovementioned baseline procedures have two key drawbacks. The first being the increased delay in obtaining the schedule grant and the second being the lack of assurance in getting the transmission opportunity.

Table I: CP ESTABLSHMENT DELAY [22]

Step	Description	Duration
0	*Approaching area of interest*	
1	Average delay due to RACH scheduling period	5msec
2	RACH Preamble	1msec
3	Preamble detection and transmission of RA response (Time between the end RACH transmission and UE's reception of scheduling grant and timing adjustment)	5msec
4	UE Processing Delay (decoding of scheduling grant, timing alignment and C-RNTI assignment + L1 encoding of RRC Connection Request)	2.5msec
5	TTI for transmission of RRC Connection Request	1msec
6	HARQ Retransmission (@ 30%)	0.3 *5ms
7	Processing delay in eNB (Uu –> S1-C)	4ms
8	S1-C Transfer delay	Ts1c (2 – 15msec)
9	MME Processing Delay (including UE context retrieval of 10ms)	15msec
10	S1-C Transfer delay	Ts1c (2 – 15msec)
11	Processing delay in eNB (S1-C –> Uu)	4msec
12	TTI for transmission of RRC Connection Setup (+Average alignment)	1.5msec
13	HARQ Retransmission (@ 30%)	0.3 *5msec
14	Processing delay in UE	3msec
15	TTI for transmission of L3 RRC Connection Complete	1msec
16	**HARQ Retransmission (@ 30%)**	0.3 *5msec
	Total LTE IDLE to ACTIVE delay (C-plane establishment)	**47.5msec + 2 * Ts1c**

VI. CARP: CONTEXT-AWARE RESOURCE PRE-ALLOCATION

One of the key characteristics of the vehicular mobility is that they have restricted spatial distribution since the vehicles have specific dimensions and their mobility is confined to the dedicated road infrastructure which is of certain capacity. Consequently, the maximum number of vehicles on a road segment is known beforehand. Additionally, by considering safety aspects, inter-vehicle distance can be taken into account to know the maximum density of the vehicle-UEs.

The above observation leads to the outcome that, in certain cases, allocating resources in advance on a per-geographical area basis, rather than per-UE basis, will not be extremely costly for having collision free communication. Figure 8 illustrates such geographical area division in an intersection marked by the grid lines where each block in the grid can be allocated with resources beforehand. Also, the cost incurred by the pre-allocation can be further reduced if the resource pre-allocation is combined with spatial reuse by limiting the transmission power. Such an approach (i.e., pre-allocating resources in specific areas and for specific use) eliminates the delay in obtaining the resources. The gain of such an approach is that it limits the communication delay to values that are required by the most demanding V2X use cases (e.g., collaborative collision avoidance).

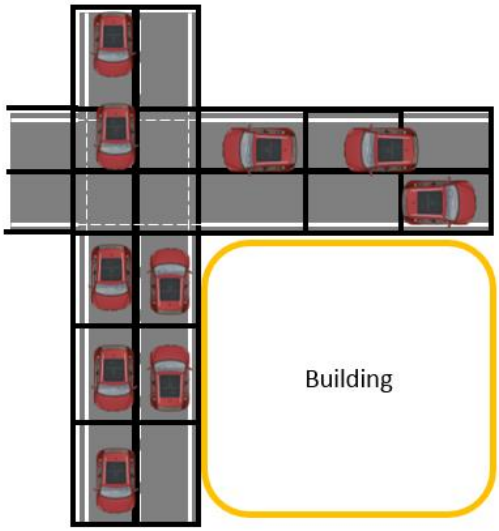


Figure 8: Pre-allocation Layout

The knowledge about the existence of these resources can be communicated to the vehicles in advance (during initial attachment or tracking area updates). Hence, it has the potential to meet the delay-bound requirements without the need for scheduling. By availing context information, the pre-allocation strategy can support collision free low delay-bound communication with a guaranteed delay. Obviously, the penalty of this scheme is that the pre-allocated resources are wasted if there is no need to be used. On the other hand, there are not really any other alternatives to support delays in the magnitude of a few milliseconds unless all vehicles are always in a connected state. But this requires even more resources from the network.

Context information about the streets and specific geography (e.g., crossings, junctions, highways, etc.) combined with information about the traffic limitations (e.g., speed limitations, etc.) facilitate proper splitting of the geographical area and allocation of the resources where these are needed (e.g., in crossroads). Thus, by using the vehicle context and road topology, tailor-cut to vehicular communications, resources can be pre-allocated to specific segments on the streets. Then, vehicles can use them as long as they are on the predefined place for predefined services that require low latency and high reliability.

We call this framework of pre-allocating resources in specific road segments and communicating this information to the vehicles beforehand CARP (Context Aware Resource Pre-allocation). To achieve its purpose, it is required from 5G networks to be aware of the street geography and also about the current vehicle traffic load in the streets. This is possible to achieve in 5G networks since the introduction of NEF allows application functions such as the V2X application server to exchange information both with control functions but also to the Network Management System.

By analyzing the context information for the vehicles and the streets, a centralized entity may further update the

pre-allocated resources accordingly. Figure 9 presents how this process takes place and the decision is distributed to the vehicles. Initially, the V2X application provides to the Network Management System (NMS) the vehicles and street context thought the NEF. The NMS, considering the street statistics and the vehicles information concludes about the proper gNBs configuration in the form of a Radio Resource Map (RR map). The NMS, by identifying segments where certain emergencies are highly probable, can proceed in certain optimizations. Then, the NMS communicates this information to the gNBs and the AMF, so as to configure the first ones and facilitate the vehicles information through the AMF. The informing of the vehicles can take place through the tracking area update procedure or every time a path diversion takes place.

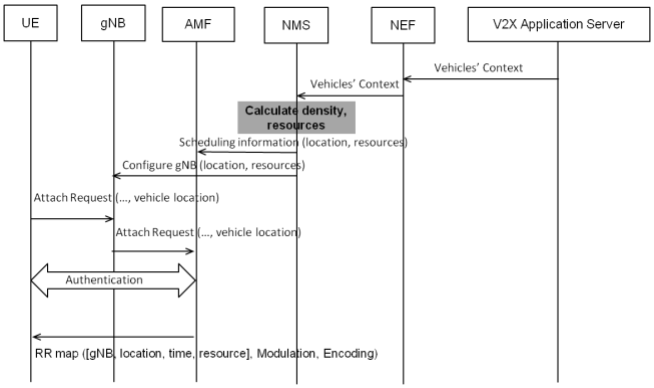


Figure 9: Provision oa radio resource map to a UE

VII. CARP PERFORMANCE EVALUATION

To analyze the capacity requirements of the pre-allocation scheme, an urban environment described by the urban information society use case in METIS I project has been used [23]. This urban topology is based on the Madrid-grid as shown in Figure 10. The dimension of the grid is considered to be 387m in width and 552m in height with lanes of 3m width. The length of a vehicle is assumed to be 4m and the number of microcells is assumed to be 24. Considering 8 horizontal lanes and 10 vertical lanes, the total length of lanes without overlap is (387-30)*8 + 552*10 = 8376m. Then, when the road is congested at its maximum capacity, the maximum number of vehicles on the road can be 8376/4 = 2094. The assumptions for the evaluation are presented in Table II.

Table II: Assumptions of the evaluations

Parameter	Value
Vehicle size	4m
Inter-vehicle distance	2.5sec
Packet size	100 bytes
Transmission interval	5msec
Number of gNB	24

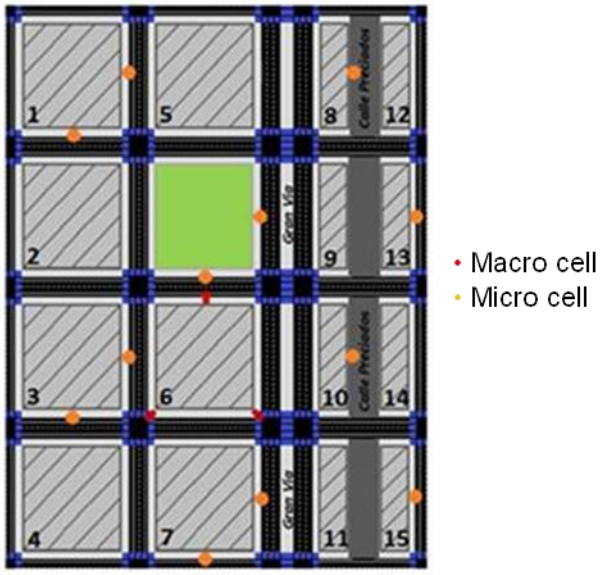


Figure 10: Madrid-grid [23]

Figure 11 presents the cost of the pre-allocated resources in terms of required resource blocks for guaranteeing transmission opportunities for various vehicular densities. Here, each resource block is considered to have 12 subcarriers with inter subcarrier spacing of 15 kHz as per the existing LTE system. The considered messages are of 100 bytes size and are required to be transmitted within 5ms. It is observed from the evaluations that the cost of pre-allocation can be as low as only 6 resource blocks when vehicles are moving with the speed of 15m/s and using a Modulation and Coding Scheme (MCS) 15. Even under the higher density scenarios and MCS 15, the cost associated is only about 17 resource blocks to achieve the required delay.

In this analysis we observe that in cases of low inter-vehicular distance a larger amount of resources for the pre-allocation (i.e., ~ 9-10 MHz) is needed. Whereas when the inter-vehicular distance is rather high the cost of pre-allocation of resources is quite low (~ 1-1.5 MHz).

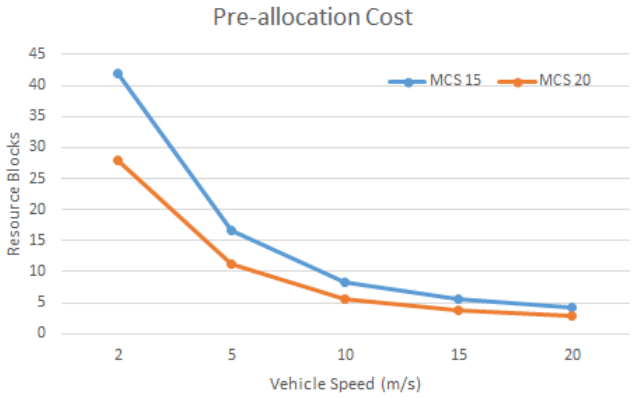


Figure 11: Pre-allocation Cost

However, the abovementioned case is rather extreme since we need to pre-allocate resources in the overall area

under consideration. A more realistic use case relates to the pre-allocation of resources only in the areas of interest (e.g., crossings) and only for certain distance from these points of interest. To analyze the cost of pre-allocation in such points of interest, intersection areas of the map are considered. In particular, one small intersection (cross of 2 vertical and 2 horizontal lanes) and one large intersection (cross of 6 vertical and 2 horizontal lanes) including area of 50m radius from the center of these intersection are analyzed. The cost analyses of these intersections are given in Figure 12 considering each case with MCS 15 and 20. It is observed that with about 4 to 6 resource blocks, pre-allocation can support low latency communications even under higher density intersections. This translates to the bandwidth requirement of only ~0.7 MHz – 1 MHz along with additional 10% of the required bandwidth for guard bands. This cost is rather acceptable assuming the 10MHz bandwidth is typically allocated for V2V communications [24].

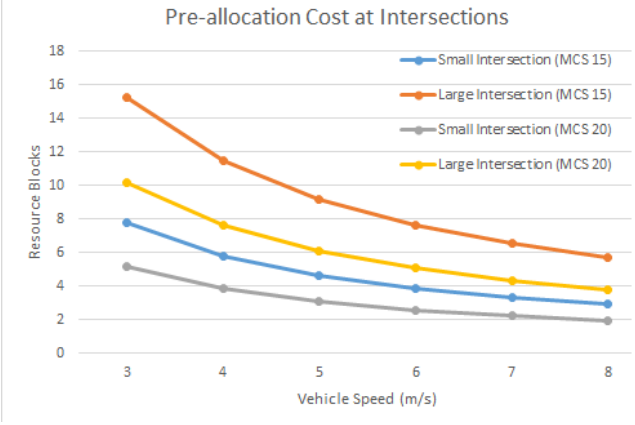


Figure 12: Pre-allocation Cost at intersection

VIII. Key Findings

As we have illustrated in the paper, the 5G architecture is flexible enough and gives a new opportunity to support very demanding use cases for the vertical industries. Having the appropriate functions in place (e.g., NEF) and by allowing the communication of the application functions (e.g., V2X application server) with the network components it is possible to collect the necessary static (e.g., streets geography) or dynamic (the moving path of a vehicle) contextual information and re-design the operation of control functions.

In our view, what is missing from the current version of the specifications is exactly this consideration of contextual information and how it can improve the control functions of the networks. Information about the path and the speed of a UE, the maximum number of UEs the geography of the streets (e.g., crossroads etc.) can assist considerably in minimizing the communication delay, increase the reliability and alleviate the signaling cost in a network.

This strategy of re-examining the network control operations based on the available contextual information can be adopted not only for the V2X case but also for all types of verticals (e.g., mIoT communications). We expect that in the future, for the next releases of 5G specifications, similar approaches will be followed for the further elaboration of the overall 5G architecture as a framework, as well as, the fine tuning of NFs that will support the operation of the different network slices on a per use case basis.

IX. Conclusions

This paper makes the case that although the specification of 5G networks is well underway and slicing is gradually reaching a mature status, several inefficiencies still exist. Standardization activities have sensibly focused on introducing new principles like NF modularization and the support of different numerologies in RAN and ported existing functionalities into the new principles.

What is still missing though are further optimizations, that can be realized if use case specific context information is taken into account. In this paper, we have presented a new mobility management scheme that outperforms the baseline for the case of high moving UEs, like the autonomously driven vehicles. By taking advantage of the knowledge of the path that a vehicle will follow and by tailoring cut the involved network functions (e.g., AMF, NEF) appropriately, then significant benefits can be achieved in terms of signalling reduction with a manageable penalty of additional information being moved inside the network.

Moreover, we have introduced a novel approach to pre-allocate resources in road segments, where this is required (e.g., crossroads) and use this information to minimize the communication delay among vehicles. This is achieved by allocating not a significant amount of resources. As a next step of the current work, we will evaluate the proposed schemes using event driven simulations.

References

[1] P. Spapis, C. Zhou, A. Kaloxylos, “On V2X Network Slicing: Using Context Information to Improve Mobility Management”, IARIA INNOV 2017, Athens, Greece, October 2017.

[2] NGMN Alliance, 5G white paper, v 1.0, 2016 available from: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf, access date 12-09-2017.

[3] S.E. Elayoubi, M. Fallgren, P. Spapis et al., “5G service requirements and operational use cases”, European Conference on Networks and Communications - EuCNC 2016, DOI: 10.1109/EuCNC.2016.7561024.

[4] P. Marsch et al., “5G Radio Access Network Architecture: Design Guidelines and Key Considerations”, IEEE Communications Magazine, vol. 54, issue 11, pp 24-32, November 2016.

[5] X. An et al., “Architecture Modularisation for Next Generation Mobile Networks”, European Conference on Networks and Communications - EuCNC 2017, DOI: 10.1109/EuCNC.2017.7980664.

[6] 3GPP, TS 23.501 “System Architecture for the 5G System; Stage 2 (Release 15)”, Version 15.0.0, December 2017.

[7] 5G-PPP Architecture Working Group, “View on 5G Architecture (Version 2.0), July 2017, available at: https://5g-ppp.eu/5g-ppp-revised-architecture-paper-for-public-consultation/, access date 12-09-2017.

[8] 3GPP, TS 22.261, “Service Requirements for the 5G System”, V16.2.0, January 2018.

[9] 5G Automotive Association - 5GAA, “The case for Cellular V2X for Safety and Cooperative Driving”, available at: http://5gaa.org/pdfs/5GAA-whitepaper-23-Nov-2016.pdf, access date 12-09-2017.

[10] 3GPP, TS 23.502, “Procedures for the 5G System”, Stage 2 (Release 15), Version 15.0.0, December 2017.

[11] 3GPP, TS 38.300, “NR and NG-RAN Overall Description”, Stage 2 (Release 15), Version 15.0.0, January 2018.

[12] 3GPP TR 28.801, “Study on management and orchestration of network slicing for next generation networks”, Release 15, Version 15.1.0, January 2018.

[13] K. Chatzikokolakis, A. Kaloxylos, P. Spapis, N. Alonistioti, and C. Zhou, “A survey of location management mechanisms and an evaluation of their applicability for 5G cellular networks”, Recent advances in Communications and Networking Technologies, vol. 3, no. 2, 2014.

[14] S. Hailu and M. Säily, “Hybrid paging and location tracking scheme for inactive 5G UEs”, European Conference on Networks and Communications - EuCNC 2017, DOI: 10.1109/EuCNC.2017.7980730.

[15] 3GPP TS 23.285, “Architecture enhancements for V2X services”, Release 14, Version 14.5.0, December 2017.

[16] R. Trivisonno, R. Guerzoni, I. Vaishnavi, and D. Soldani, “Towards zero latency software defined 5G networks,” in IEEE International Conference on Communication Workshop (ICCW), June 2015, pp. 2566–2571.

[17] T. V. Mathew and K. V. Krishna Rao, “Introduction to Tranportation Engineering”, Chapter 30, Fundemental parameters of traffic flow, May 2007, available at: http://nptel.ac.in/courses/105101087/downloads/Lec-30.pdf, access date 15-05-2018.

[18] 3GPP, TS 33.401, “Security Architecture”, Release 15, Version 15.2.0 January 2018.

[19] 3GPP, TR 36.912, "Feasibility study for Further Advancements for E-UTRA (LTE-Advanced)", Version 14.0.0, March 2017.

[20] N.G.M.N. Alliance, "Perspectives on vertical industries and implications for 5G", 2016, available at: https://www.ngmn.org/publications/all-downloads/?tx_news_pi1%5Bnews%5D=516&cHash=191d94c830ec204060fdc44deb5aef32, access date 16-05-2018.

[21] I. L. Da Silva, G. Mildh, M. Säily, and S. Hailu, "A novel state model for 5G Radio Access Networks," 2016 IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur, 2016, pp. 632-637.

[22] 3GPP, TR 25.912, "Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN)”, Version 14.0.0, March 2017.

[23] METIS, "Simulation guidelines", Deliverable D6.1, October 2013.

[24] 3GPP, TR 36.885, "Study on LTE-based V2X Services", Version 14.0.0, June 2016.

Energy-efficient Live Migration of I/O-intensive Virtual Network Services Across Distributed Cloud Infrastructures Leveraging Renewable Energies

Ngoc Khanh Truong*, Christian Pape*, Sven Reißmann†, Thomas Glotzbach‡ Sebastian Rieger*
*Department of Applied Computer Science
Fulda University of Applied Sciences, Fulda, Germany
{ngoc.k.truong, christian.pape, sebastian.rieger}@cs.hs-fulda.de
†Datacenter
Fulda University of Applied Sciences, Fulda, Germany
sven.reissmann@rz.hs-fulda.de
‡Department of Electrical Engineering and Information Technology
Hochschule Darmstadt University of Applied Sciences, Darmstadt, Germany
thomas.glotzbach@h-da.de

***Abstract*—Virtual infrastructures and cloud services became more and more important over the past years. The abstraction from physical hardware offered by virtualization supports an increased energy efficiency, for example, due to higher utilization of underlying hardware through consolidation. Also, the abstraction enables the ability to geographically move cloud services, e.g., to be able to benefit from lowest available energy prices and renewable energy. This article gives an overview on such migration techniques in distributed private cloud environments. The presented OpenStack-based testbed is used to measure migration costs along with the service quality of virtualized network services. Correspondingly, the article illustrates the impact of high memory and input/output (I/O) load on live migrations of network services and evaluates possible optimization techniques. The results gained from the experiments presented in this article, can be used to evaluate whether network services and virtual resources can be migrated to distant sites to reduce energy costs. A potential benefit of such migrations can be to leverage from fluctuating renewable energies across multiple data center sites. Possible improvements as well as side effects of this use case are presented in the evaluation regarding the live migration of virtual network services. Regarding virtual network services, potential drawbacks can result from additional latency when maintaining and using the virtual services across distant locations. To mitigate these effects, the article describes a way to identify dependencies and affinities between virtual and physical resources based on network flow data. The evaluation used a data set of characteristic networks flows from around one hundred virtual machines of the production environment at Fulda University of Applied Sciences. While respecting these requirements and dependencies, the optimization described in this article used weather data of multiple years of three different distant locations in Germany. Possible improvements of the utilization of renewable energies due adaptive placement and migration of virtual resources were evaluated using this data. Together with the detailed evaluation of the costs of these migrations, which especially rise for I/O-intensive migrations, e.g., for virtual network services, the results of this article can be used to increase the overall energy efficiency of data centers in distributed cloud infrastructures.**

Keywords–Cloud Computing; Network Services; Live Migration; Energy Efficiency; Renewable Energy.

I. INTRODUCTION

A solution for energy-efficient live migrations of I/O-intensive virtual network services across distributed cloud infrastructures was presented in [1]. In this article, these findings will be elaborated and their benefit for the use of renewable energy (RE) sources between distant data centers while limiting possible drawbacks of distributed virtual resources, e.g., due to resource dependencies and associated affinity groups, will be explained. Energy costs are an important factor for data centers and IT infrastructures as a whole. Drivers for the increasing costs over the last decade have been electricity prices, but also the growing energy demand of data centers and IT infrastructures. Regarding the electricity price, the changes in national energy policies to move from low-priced conventional, e.g., nuclear, power to renewable energies (e.g., in the European Union and especially in Germany), augur that energy costs will increase even further. While the percentage of the costs for network equipment and services have been negligible for data centers in the past, this is likely to change due to increased bandwidth and the steadily increasing number of network devices, amplified by the evolving "Internet of Things" and cloud-based services. Recent papers even state that the network power consumption could grow beyond 25% [2][3] of the total data center energy demand. This is especially likely for large data centers (i.e., Google, Amazon, Facebook), whose inner data center traffic is quickly increasing [4]. Since virtualization is used for compute, storage and network resources in modern data centers, these infrastructures support automatic provisioning and management of virtual resources, which can be used to optimize the energy efficiency. For example, virtual resources can be consolidated to reduce the required hardware based on the current load. During off-peak hours, resources and links can be powered down or use power management, while being quickly and automatically reactivated on demand. This also allows for elastic scalability [5], as well as adaptive scheduling, placement and migration of virtual resources. The scheduler can consider electricity prices and the availability of RE resources across multiple data centers [6]. Hence, an energy- and cost-efficient adaptive placement of virtual resources can be attained.

Regarding the network, cloud environments typically employ network virtualization to implement networking functions for their delivered services. These virtualized network services offer transparent use and flexibility regarding the underlying resources. Such services, e.g., in the form of virtual network functions (VNF), are not only getting more and more momentum in service provider networks (as defined, e.g., in the

network functions virtualization (NFV) reference model of the ETSI [7]), but also in virtualized network infrastructures as a whole. While the "on-demand self-service" and "rapid elasticity" paradigms [5] of cloud and software-defined infrastructures imply that virtual resources, including virtualized network services, can quickly be spawned and destroyed, not all use cases of virtual network services fulfill such a "pets versus cattle" approach [8] for cloud resources. Spawning and stopping new VMs or containers behind a load balancer for example, is easy to implement, while the migration of entire clusters of services and load balancers across cloud infrastructures, without losing network connectivity (e.g., sessions, traffic flows) requires special techniques. Virtualization and cloud environments typically allow for a transparent migration of virtual resources across different underlying hardware components (e.g., implementing a "live migration" technique). Hence, network services impose special requirements for live migrations. The network load, e.g., on VNFs, is typically higher than on back end servers, due to their function as a front end for multiple services or servers. This leads to a high I/O rate of the virtual machines (VMs) and containers offering such virtual network services (e.g., VNF). Sometimes, these I/O-intensive memory and network operations are therefore enhanced by using special acceleration functions of the underlying hardware, i.e., TCP offloading or single-root I/O virtualization (SR-IOV), which also hold specific constraints for live migrations, due to the fact that they are depending on local physical hardware (i.e., network interface cards).

In this article, an analysis of the impact of these implications for the migration of virtual network services across distributed cloud environments is presented. Our experimental approach uses an OpenStack-based testbed migrating virtual network services under load and evaluating the results. Additionally, techniques to improve the energy efficiency of the migration are discussed. By using a live migration, the services can be transferred seamlessly during operation instead of interrupting existing connections leading to additional energy being required to reestablish lost connections. However, the energy consumption of the migration itself needs to be optimized (e.g., limiting resources and time needed for the migration).

The migrated virtual network services can include typical NFV or SDN components (e.g., virtual switches, controllers). Regarding the energy efficiency, again the stated "pets versus cattle" paradigm cannot be applied to every migration of virtual network services. Besides the negative effect of the downtime while respawning, e.g., VNFs, also more energy is consumed, if components like switches, firewalls or load balancers are simply destroyed and lost connections or flows need to be reestablished, increasing the load on the underlying cloud infrastructure. Furthermore, more energy is consumed if constraints like placing the network functions close to back end servers etc. are not satisfied (e.g., due to resource dependencies or "affinity groups" of virtual resources). Hence, the energy consumption of the migration itself needs to be optimized (e.g., limiting the resources and time needed for the migration). This article also includes an outlook on further improvement by using containers and microservices, reducing the effort for live migrations regarding the state and data that needs to be transferred. The migration techniques for virtual network services can also be combined with upcoming network and server power management, to increase energy efficiency and power savings even further.

A potential beneficial use case for energy-efficient live migrations is the combination with RE sources for data centers, as already discussed, e.g., in [9]. This includes placing newly started virtual resources in sites, which are temporarily offering a high amount of RE resources, as well as migrating running virtual resources to such sites. This might require long distance migration and placement across geographically distributed data centers, e.g., in cloud infrastructures, to leverage from fluctuating RE sources (e.g., solar or wind power). However, especially in the case of highly distributed data centers, the already stated relevance of dependencies for the virtual resources (e.g., underlying compute, storage, network resources), needs to be considered. For example, this means that affinity groups for virtual resources (e.g., virtual or physical resources that need to be combined to form a service being offered to distant users) must be respected during the optimization of energy-efficient placement and migration of the virtualized resources. This way, the location independent placement of virtualized resources, due to the abstraction of the virtualization infrastructure from physical hardware, can be used to enhance the use of RE sources within data centers accounting for a large portion of national energy consumption [9].

The rest of this article is laid out as follows. Section II presents related work and defines the research questions of this article. In Section III, the state of the art in energy-efficient private clouds, as well as the usage of virtual network services and live migration of virtual resources in such infrastructures are described. The model for our approach is introduced in Section IV, describing the requirements for scheduling and migrations of virtual network services in private clouds, to support an energy-efficient placement while respecting corresponding requirements like dependencies and affinities of virtual and physical resources. Section V characterizes the testbed created to measure the impact of virtual network service migrations on the energy efficiency of private cloud infrastructures and presents the results of the evaluation. Additionally, Section VI discusses the potential to use energy-efficient placement and migration of virtual resources to leverage fluctuating renewable energies while respecting the earlier defined requirements regarding dependencies and affinities between virtual and physical resources. Finally, Section VII draws a conclusion, discusses the findings of the evaluation compared to the related work, and gives an outlook on further research in this area.

II. Related Work

Migration of virtual resources and its impact on application performance is subject of current research. The energy-efficient placement of VMs in an OpenStack-based environment is discussed in [10][11]. Indeed, these approaches target on the algorithms used for placing VMs based on temperature and cooling demands, but also focus on network requirements for the VMs. A vector-based algorithm for VM placement considering the availability of renewable energies is discussed in [12]. Furthermore, more general evaluations are given in [13] and [14]. These publications examine the relevant parameters for an energy-efficient placement of VMs in a data center. A basic analysis of VM migration costs and the impact of migration on application performances is discussed in [15]. In [16], an estimation of the energy consumption of physical servers running VMs and an algorithm for energy-efficient VM placement are described. The ElasticTree project [17] focuses

on energy-efficient computer networks by throttling network components using OpenFlow. Other projects like ECODANE [18] extend these ideas to also provide traffic engineering techniques. Constraints and requirements for energy-efficient placement of VMs related to their network connectivity were introduced in [19][20][21]. An evaluation of the power consumption during VM migration tasks is presented in [6]. This publication also includes a breakdown on different data center components like storage, network and compute resources. Furthermore, [22] discusses an energy-aware virtual data center architecture using software-defined networking (SDN). Finally, [23] introduces benchmark test metrics for performance and reliability monitoring and discusses related issues. A study comparing different hypervisors concerning migration time and efficiency is presented in [24]. The interference effects of simultaneously running migrations and the efficiency of different permutations of migrations are reviewed in [25].

Several publications also address the identification of affinity groups, e.g., between virtual resources and services. Migrations of complete groups including their underlying SDN network structures are introduced in [26]. The base method called LIME is formalized and its correctness is proofed. A deduplication strategy for transmitting identical memory pages only once during the migration of VM groups is outlined in [27]. In [28], the grouping of VMs with the focal point on performance of highly parallelized applications is described. Another algorithm for optimizing the migration of related VMs by reserving bandwidth along traffic paths is outlined in [29]. Statistical techniques are used in [30] to create representative groups of similar VMs in order to simplify monitoring. Management tasks identified for this subset of machines can be applied on all relevant virtual resources.

III. State Of The Art

The evolution of cloud services in IT infrastructures enables companies to speed up business processes and scale their services on demand. Physical servers, storage and network devices are consuming energy, but today these components are typically just the foundation for virtualized workload running on top of them. Moreover, in such highly virtualized environments, the virtual resources providing the services are the decisive consumers of power and bandwidth. Orchestration and automation techniques like SDN can help to optimize the power consumption in cloud infrastructures. To ensure the service quality and scalability along with the energy efficiency, it is necessary to investigate the behavior of these virtual resources, e.g., regarding available migration techniques.

A. Energy-efficient Private Clouds

Today, energy efficiency and power management is a foundation pillar in modern data centers. This is mainly driven by increasingly high energy costs and energy consumption in large-scale IT infrastructures. The sensitization for the sustainable use of resources like renewable energies, e.g., as a result of the Fukushima nuclear disaster and the consequential renunciation of nuclear energy for example in Germany and Europe, additionally supports this process of rethinking data center designs. Data centers are using a large amount of power not only for running the IT components and equipment, but also for cooling them. The ratio between energy consumed by IT equipment and the overall power consumption including cooling and energy loss in power supplies is known as the power usage effectiveness (PUE). This value describes the operational overhead of data centers and is an eligible candidate for optimization approaches.

The concept of cloud computing enables companies to better utilize their physical IT resources and empowers them to dynamically scale their services in a location-independent manner. To take advantage of these benefits, a consequent resource management must be deployed. Ideally, this means that currently not required compute resources, as well as their dependencies like upstream or downstream storage or network devices are partially or fully suspended or shut down. The consumption of energy in a common cloud environment depends on its directly associated physical infrastructure components like compute resources (i.e., central processing unit - CPU, random access memory - RAM), storage devices (i.e., storage area networks - SAN, network attached storage - NAS, local or direct-attached storage) and network components (i.e., routers, switches, firewalls). Thus, the power consumption of a service depends on the physical IT resources that are needed to provide it. However, VMs providing cloud services are not picky concerning their location of execution, as long as required dependencies are met at either site.

By migrating virtual resources across distant data centers in different regions, it is possible to optimize energy efficiency and cost. Such "follow-the-sun" data center services move their workload to different geographic regions to more efficiently balance computing demand while taking into account the latency for the end-users to access the service. Usually, the output of RE sources is fluctuating, which means that the energy is not always available when needed and also not necessarily produced near the point where it is consumed. Further, energy storage at industrial scale is not available yet. Related to that, this also leads to seasonal and regional energy price fluctuations. The cloud paradigm enables companies to move their workload nearby the currently available RE sources and to take advantage of the economic benefits by consuming energy at lower prices.

B. Migration of Virtual Resources in Private Clouds

Today's cloud software is providing a layer for scalable and elastic cloud applications that allows to deploy virtual network services (e.g., VNFs) like routers, load balancers or firewalls. Also, private cloud platforms like OpenStack already added a lot of these functions to their service portfolio. As a result, many industry-leading service providers are starting to use OpenStack as a platform to deliver reliable and scalable services and applications. This includes VMs running customer-facing applications, as well as virtualized storage and networking components needed for the service delivery. Of course, containers as a very thrifty and scalable building block for cloud services can also be provisioned and deployed in these infrastructures. However, to offer reliable, elastic and energy-efficient services, these resources have to be movable across the infrastructure components. This movability of virtual resources is mostly provided by VM migration from one node to another. The migration can be implemented live or online by transferring block storage of the VM or using a shared storage back end, and finally transmitting the main memory and CPU state. Furthermore, a VM can also be migrated offline by suspending, transmitting its state and resuming the machine consecutively. These approaches are described in detail in Section IV-B. When a VM does

not contain any essential data and the configuration can be realized by an automated provisioning mechanism, it is also possible to just destroy a VM or container on the source node and recreate or respawn it on the destination node. It is obvious that this technique minimizes network transfer costs and requirements for shared storage hardware but also implies that the cloud application or service is well-designed related to elasticity. Moreover, live migration techniques for containers are currently developed and discussed. While the small size of containers compared to VMs reduces the network traffic for the migration, saving the state of containers holds much more dependencies and hence is more difficult to implement [31].

IV. Energy-efficient Placement of Virtual Network Services in Private Clouds

The migration of virtual network services to regions where RE sources are currently available or where energy prices are lower, can substantially improve the overall energy efficiency of distributed data centers. However, if the costs for the migration are too high, e.g., due to a reduced performance of the migrated resources, the migration will be inefficient. For these reasons, when designing services, it is important to understand how the migration process is performed in the underlying infrastructure to restrict possible negative consequences of migration costs.

A. Scheduling

A common OpenStack-based cloud environment is based on multiple services. First of all, Nova, the compute fabric controller, encapsulates the hypervisor and is responsible for the execution of VMs. Block-level storage is provided by the Cinder service. It manages the complete life cycle of block devices for the virtual servers. The image service Glance stores disk and server images and their metadata and assures that they are available to the compute nodes. The networking component Neutron manages multi-tenant virtual networks supporting different network architectures. For example, traffic can be managed using SDN technologies like OpenFlow. Also, OpenStack Neutron already offers some virtual network services (i.e., VNF) like firewalls and load balancers as a service. While OpenStack contains additional components, this article is based on the OpenStack core services described above. Scheduling and placement of virtual resources in OpenStack environments is carried out by schedulers of the services given above. For example, the nova-scheduler checks which compute nodes can provide the requested resources. The decision is based on filters (i.e., based on capacity, consolidation ratio, affinity groups) that can be modified by an administrator.

B. Migration Techniques

One of the crucial points when performing the migration is to ensure that services should not be disrupted during the migration process, otherwise possible service-level agreements (SLAs) will be violated. OpenStack, which typically uses libvirt and the kernel-based virtual machine (KVM) hypervisor, provides three different migration types to move VMs from the source host to a destination host with almost no downtime: shared storage-based live migration, block live migration and volume-backed live migration [32]. Shared storage-based live migration, as the name states, requires a shared storage that is accessible from source and destination hypervisors. During the migration only the memory content and system state (e.g., CPU state, registers) of the VM are transferred to the destination host. This migration type in OpenStack can be performed using a pre-copy [33] or post-copy [34] approach. In the former, VM memory pages are iteratively copied to the target without stopping the services running on the migrated VM. Every change in memory state (i.e., dirtied memory) during the copy phase will trigger another transfer of modified memory pages. If predefined thresholds have been reached, e.g., the number of iterative copy rounds or the total amount of transmitted memory, or the amount of modified memory pages in the preceding copy round is small enough [35], the copy process is terminated, whereby the source VM is suspended, the source hypervisor copies the remaining modified memory pages and system state and resumes the VM on the destination. Depending on the dirtied page rate this switching can cause a downtime. A big issue of pre-copy migration arises at the iterative copy rounds. If the rate of memory changes exceeds the transfer rate over the network, then the copy process will run infinitely. This limit can be eliminated by post-copy migration, in which at the beginning of the migration the migrating VM is stopped on the original node, then the non-memory VM state is copied to the destination, after which the VM will be resumed on the target. In parallel, a prepaging will be performed. At this stage, the memory pages are proactively pushed by the source to the destination VM. Any access to the memory pages on the target VM that have not yet been copied, result in the generation of page faults, requiring to transfer the accessed memory pages over the network. This process is known as demand paging. Obviously, this behavior can solve the indefinitely migration problem, but can cause a huge degradation of VM performance because of the large amount of page faults transferred over the high-latency medium in comparison to pre-copy migration. Moreover, post-copy cannot recover the memory state of the migrated VM in the case of network failure during the transfer of the page faults.

As the requirement of a shared storage increases the financial burden, block live migration is considered more cost effective. No shared storage is required when the migration takes place. Hence, this migration type is especially useful when moving the VMs between two sites over long distances without having to expose their storage to one another. This type is very similar to Microsoft Hyper-V Shared-Nothing Live Migration feature [36]. Initially, not only a VM on the remote host is created, but also the virtual hard disk on the remote storage. During the migration, at first the virtual hard disk contents of the running VM must be copied to the target host. Changes of disk contents as a result of write operations will be synchronized to the destination hard disk over the network. After the migration of the VMs storage is complete, the copy rounds of memory pages are executed, which perform the same processes used for shared storage-based live migration. Once this stage is successfully finished, the target hypervisor will resume the VM, while the source hypervisor deletes the VM and its associated storage. Volume-backed live migration behaves like shared storage-based live migration since VMs are booted from volumes provisioned by Cinder instead of ephemeral disk, i.e., VM disks on shared storage. To achieve energy-efficient placement of VMs, the migration costs must be taken into account. These costs play an important role for the scheduling process to decide when and how often services should be migrated to remote hosts.

Two categories of parameters to calculate migration costs will be analyzed in this article: total migration time, which denotes how long the migration lasts from the start of copy rounds until the VM is resumed on the remote host, and performance loss, which focuses on the degradation of the services' performance during the migration process. Apparently, these costs are strongly impacted by the iterative copy rounds due to any modification on memory pages or disk contents. They should be thoroughly calculated to allow the scheduler to efficiently place services not only in terms of energy, but also their quality of service.

C. Communication Flows

In most cases a VM cannot be considered on its own. Also, several VMs are part of the provisioning of a service or business process. End-users access services from their devices via front end systems. These facades handle the communication with the end-user devices, the service itself is provided in cooperation with several other systems. This includes systems responsible for directory services, domain name resolution and identity management but also components of the business process itself like database management systems and application servers. For each exchange of data, a connection needs to be established between the involved systems, e.g., via the internet and transport layer.

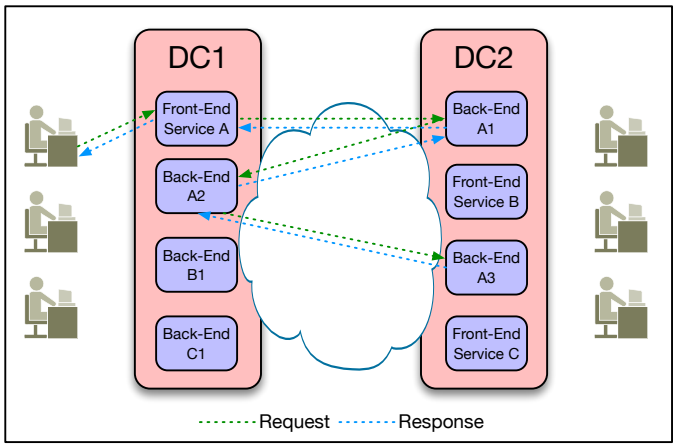


Figure 1. Suboptimal distribution of virtual resources across two data centers.

This clearly illustrates that latencies in communication flows between the involved systems have to be accumulated. The resulting overall latency is usually small when the systems reside at the same site or ideally inside the same hypervisor. But these latencies rise when the involved systems are distributed geographically. In most cases, the geographical positioning of the end-users can be neglected due to their uniform distribution, but a poor distribution strategy of the back end services can lead to negative effects on end-users latencies and an overall degradation of service quality. Figure 1 depicts a suboptimal distribution of virtual resources in two data centers. An end-user's request to the front end server results in additional requests to back end servers, which due to their poor placement across different data centers lead to a high end-user latency.

A favorable distribution is shown in Figure 2. Although the distance between end-user and the front end system is greater, a smaller overall processing time is achieved by grouping the involved server systems in one data center. These sets of

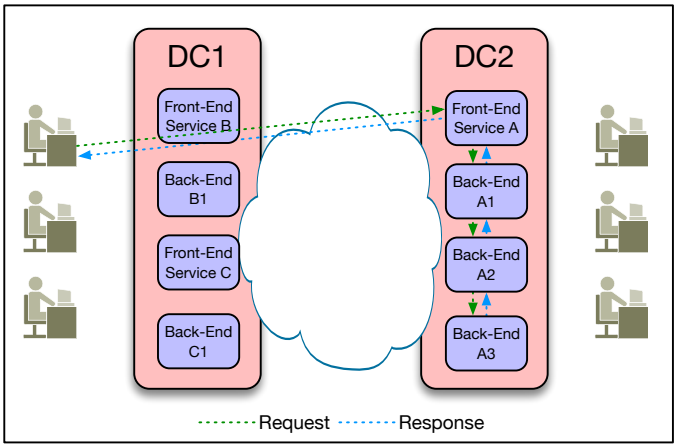


Figure 2. Favorable distribution of virtual resources across two data centers.

associated VMs are known as affinity groups. Typically they comprise the services and their dependencies, e.g., in the form of virtual resources.

D. Identification of Affinity Groups using Network Flows

There are different motivations for the collection of network traffic data, e.g., capacity planning, accounting or security monitoring (see for example [37]). The foundation builds the analysis of aggregated network flows. As stated in RFC 3917 [38], a so-called network flow can be seen as a set of internet layer packets that pass an observation point during a given time interval and share a common set of properties. The observable properties and the usable sampling mechanisms differ for the various technologies like NetFlow v5, NetFlow v9, IPFix and sFlow. For the sake of identifying affinity groups, all of these protocols provide source and destination addresses and basic counters. However, NetFlow v5 is only usable for monitoring IPv4 traffic. Today, the collection and aggregation of network flow data is not limited to physical switching or routing hardware. Especially, networks with redundant links and devices allow traffic to follow different paths from source to destination. A viable solution is to collect the network flow data as near as possible to the source or destination and to assure that all packets pass this observation point. In today's data centers and their high degree of virtualization, these data can also be collected using virtual network devices.

Listing 1. Configuration of NetFlow, sFlow and IPFIX using an Open vSwitch (see [39])

```
# ovs-vsctl -- set bridge vswitch \
  netflow=@netflow -- --id=@netflow create \
  netflow target=\"10.1.1.42:2055\" \
  active-timeout=30

# ovs-vsctl -- set bridge vswitch \
  sflow=@sflow -- --id=@sflow create \
  sflow agent=eth0 target=\"10.1.1.42:6343\" \
  header=128 sampling=64 polling=10

# ovs-vsctl -- set bridge vswitch \
  ipfix=@ipfix -- --id=@ipfix create \
  ipfix targets=\"10.1.1.42:4739\" \
  obs_domain_id=123 obs_point_id=456 \
  cache_active_timeout=60 cache_max_flows=12
```

For instance, VMware's Distributed Switch allows the collection of network flows and their export via NetFlow or

IPFix to an external collector. Furthermore, an Open vSwitch – used esp. in KVM- and Xen-based virtualization environments – supports the export of aggregated data using IPFix, NetFlow and sFlow. These virtual switches also form the basis for networking in OpenStack-based environments, as being used in this article. Listing 1 shows the commands for enabling flow collection and export using NetFlow, sFlow and IPFix.

Network flows and the associated communication endpoints can be used to build up affinity groups based on an aggregated metric and a given threshold. An affinity group can be seen as a partitioning for the set of VMs M. Thus, an equivalence relation R_P can be defined as follows:

a) Reflexivity: Each virtual resource $m \in M$ is part of the relation, so $\forall m \in M : (m, m) \in R_P$.

b) Symmetry: The related VMs whose communication volume exceed a given threshold s should also be included in the relation. So, let $\lambda(m_i \in M, m_j \in M)$ be a function that returns an aggregated metric (e.g., packet count or octet count) for two given VMs and it holds that $\lambda(m_i, m_j) = \lambda(m_j, m_i)$. Based on this constraint the symmetry is also given for the relation for machines that exceed a given threshold s, so $\forall m_i, m_j \in M, \lambda(m_i, m_j) \geq s : (m_i, m_j) \in R_P$.

c) Transitivity: The missing tuples for fulfilling the transitivity should also be added by $\forall (m_i, m_j) \in R_P, (m_j, m_k) \in R_P : (m_i, m_k) \in R_P$.

Based on this definition, a partitioning for the set M, i.e., the affinity groups, can be defined as $P_M = \{[m]_{R_P} | m \in M\}$. An example of such a breakdown can be seen in Figure 3.

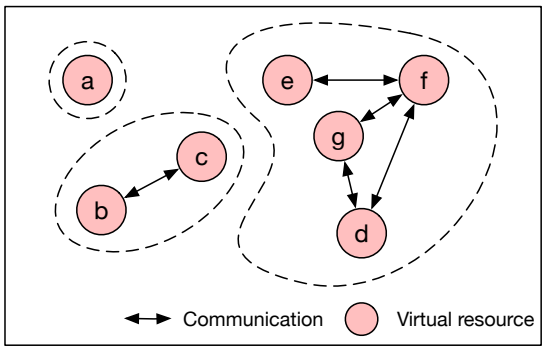


Figure 3. Partitioning of virtual resources in affinity groups based on their communication volume.

Beside the evaluation of the performance impact of VM migrations, the results also show that it is profitable to migrate VMs in order to group VMs geographically. In [40], an algorithm to optimize the usage level of renewable energies of distributed data centers was introduced. In addition to this goal, this algorithm uses network flow data to build affinity groups of VMs to avoid negative side effects. For example, these side effects could be high latency or bitrate capacity exhaustion. As a result, the consideration of the affinity groups allows for the minimization of the average distance across virtual resources that packets are traveling between. Thereby, a decrease of the overall service latency is enabled.

V. Evaluation

This experimental study concentrates on the impact of migration on memory- and I/O-intensive services. For this purpose, an experiment was set up in an OpenStack environment that is presented in the following sections.

A. Testbed Environment and Methodology

Our testbed environment consists of two physical servers that act as compute nodes and two NetApp E2700 providing block storage over 16 Gbit/s FibreChannel. Each of the compute nodes running Ubuntu 14.04 is equipped with two 8-core Intel(R) Xeon(R) E5-2650v2 2.60 GHz CPUs and 256GB of main memory. The nodes are connected using two 1 Gbit/s Ethernet interfaces over a Cisco C3750 switch. All migrated VMs run Ubuntu 14.04 with 1 vCPU, 2 GB of memory and 10 GB of disk space. In our study, migration costs of a web proxy as a virtual network service is analyzed. 10 VMs (Set 1) representing web proxy servers are initially launched on Nova-Compute 1 with a defined memory workload using the tool *stress*, which keeps dirtying a predefined amount of memory. Swapping was also activated to simulate additional I/O load on the service. If all memory for user space (1702 MB, 83% of memory size) is already allocated, inactive memory pages will be swapped out to disk.

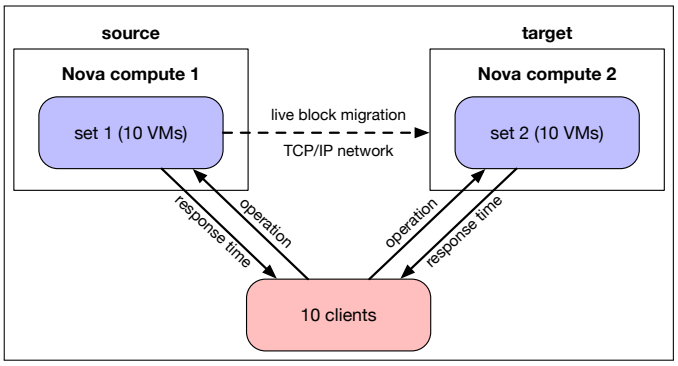


Figure 4. Overview of the methodology of the experiment.

The performance of each VM will be measured by 10 clients, each sending HTTP requests to the VMs in a fixed time interval. Additionally, extra load was produced on those VMs by sending other requests for various operations from the clients, such as searching a directory, writing a 20 MB file (disk I/O load) and generating 4096 bit RSA keys (CPU load). The response times for those requests are then used as a performance metric. After 15 minutes of measurement the same process is performed on 10 VMs of Set 2 on Nova-Compute 2. All source VMs are then concurrently migrated from Nova-Compute 1 to Nova-Compute 2 using block live migration. Block live migration were chosen due to its advantage in the case of moving the VMs located on two sites with large distance. While also 10 Gbit/s Ethernet is available in our servers and switches, the 1 Gbit/s NICs were used to better reinforce small effects of different migration parameters and changes. Furthermore, the number of concurrent migrations were varied to better understand the impact of the bandwidth on the migration. The performance of VMs on Nova-Compute 2 was also investigated to observe the influence of the migration on instances running on the target host. Figure 4 shows an overview of the methodology.

Besides several configurations that were necessary to implement a true live migration in OpenStack [32], the *max_requests* and *max_client_requests* parameters in libvirt had to be increased to 40, to support the large number of 10 concurrent migrations in the experiment. This parameter was changed in the libvirt configuration file and followed by

a restart of the libvirt service. The experiment was performed using a script and was repeated 10 times to ensure significant and reproducible results. All runs led to reproducible results. After changing a parameter in the experiment (e.g., the memory workload shown in Figure 5) it was run 10 times again.

B. Research Results and Discussion

Figure 5 demonstrates the experimental results for different memory workloads. The results show that the total migration downtime increases proportionally with stressed memory size caused by the iterative transfer of dirtied memory pages generated by the command-line tool *stress*. Another reason for this effect is the more intensive swapping of memory pages leading to a repeated modification of disk contents and thus more additional transfers over the network. In addition, the block live migration process in OpenStack will last longer, if the number of VMs migrated concurrently was reduced. The source of this impact is the overhead of nova-scheduler handling the migration requests.

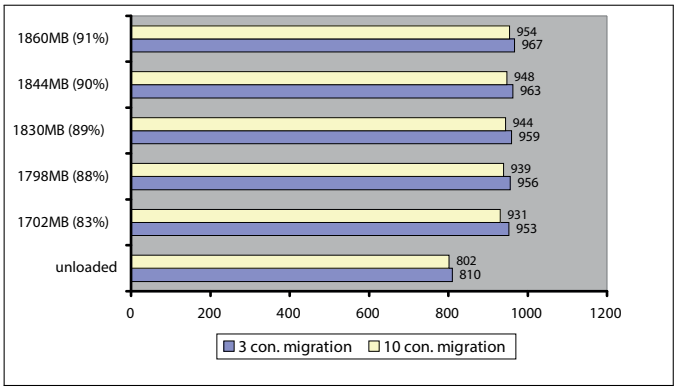


Figure 5. Total migration time (in seconds).

During the migration process, the performance degradation for search operations within the VMs significantly starting from 1830 MB loaded-memory (89% of total memory size) were observed. This degradation is shown in Figure 6, which demonstrates the response time for search operations on both sets before, during and after concurrently migrating 10 VMs of Set 1 to Nova-Compute 2. Response times were capped to a maximum of 60 seconds as seen in the figure for the second set before its creation. The average response time on Set 1 during the copy rounds rises from 2.299s to 5.606s, approximately 144%. Moreover, the migration of Set 1 to Nova-Compute 2 influences the VMs performance for search operations on this node. Particularly, the average search response time of Set 2 increases around 110% from 2.45s to 5.164s. After the VMs are moved to Nova-Compute 2, the performance of both sets is also decreased, by approximately 72% on Set 1 and 61% on Set 2, since Set 1 produces more I/O workload on the disk of the target host. The peak in Figure 6 during the migration denotes the switch process that was explained in Section IV-B.

Another conspicuous point is that the performance loss during the migration strongly depends on the amount of stressed memory as shown in Table I. The performance loss increases linear with the size of the memory workload. This could be due to the fact that the available amount of memory for buffer/cache used for I/O operations is too low so that more intensive I/O flush processes occur. Consequently, more disk synchronization must be performed over the network during

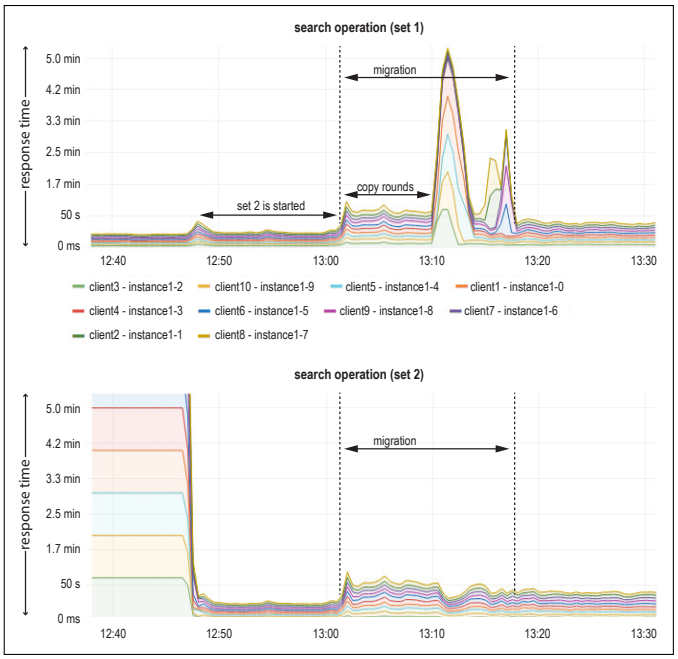


Figure 6. Performance of search operations with a memory workload of 1830 MB on Set 1 and Set 2 before, during and after the migration.

the migration, causing a slowdown in the response times. In Figure 6, one can recognize that the performance of Set 1 for the search operation slightly degrades when the VMs of Set 2 on Nova-Compute 2 are started, although they do not use a shared storage. For instance, the average response time of Set 1 increases from 2.067s to 2.532s (22.5%) in the case of 1830 MB loaded-memory, from 2.229s to 3.083s (39.7%) in the case of 1844 MB loaded-memory and from 2.856s to 6.858s (140%) in the case of 1860 MB loaded-memory. This result shows that many simultaneous intensive I/O operations on an extremely memory-intensive VM have an immense impact on the I/O performance of the underlying system in OpenStack and on the performance of I/O operations in hosted VMs, respectively. Nevertheless, this effect does not emerge if the stressed memory falls below 1830 MB, as well as for other non-I/O-related operations.

TABLE I. PERFORMANCE LOSS OF SEARCH OPERATION WITH DIFFERENT MEMORY WORKLOADS.

VM set	*Increased response time during migration (s)*		
	1830 MB	1844 MB	1860 MB
Set 1	3.307	4.389	6.241
Set 2	2.712	3.678	3.422

VM set	*Increased response time after migration (s)*		
	1830 MB	1844 MB	1860 MB
Set 1	1.655	2.527	3.431
Set 2	1.498	2.044	1.265

Last but not least, the performance of the main operation of the web proxy, serving HTTP requests, as well as the performance of the CPU-related operation, generating a 4096 bit RSA key, are only significantly impacted as the amount of stressed memory rises above 1860 MB. The average response time for HTTP requests to the migrating set grows from 0.166s to 0.785s during the migration process, whereas the one for the operation of generating an RSA key rises from 3.759s to

6.3s. This degradation effect arises only if those operations are carried out while other I/O-intensive operations such as a search for a file are running. When block live migration with separate operations were performed, the performance deviation did not occur. Therefore, it could be stated that not only I/O operations are strongly impacted by the migration process, but also have direct influence on the other operation types.

VI. Using Renewable Energies for Virtual Network Services in Private Clouds

As stated in the introduction in Section I, efficient placement and migration of virtual resources can be used to leverage RE sources. However, the energy output of RE sources is fluctuating. Hence, the energy is not always available when needed or vice versa. In addition, the energy is not always produced near its point of use (e.g., offshore wind energy). First of all, this leads to the necessity to store the energy in between or shift the consumption in time. Until now, the storage of energy is just conditionally feasible, as it is expensive and not available in industrial scale. In contrast, the possibility to shift the energy consumption of data centers with the help of an intelligent energy management is viable, as stated in the introduction of this article and in further detail, e.g., presented in [9]. Migrating VMs can help counteract the issue of fluctuating energy sources. However, as described in Section IV-B, each shift is associated with migration costs. VMs should only be moved if it is certain that a shift is worthwhile. Furthermore, it must be ensured that the VMs do not oscillate between different locations in short time. This could happen, for example, if the RE fluctuates sharply in short periods at different locations. Various optimization options have already been presented in this paper. To further optimize the number of shifts, this article proposes an additional idea besides the concept of migrating VMs based on available RE. To avoid additional shifts, VMs should be started directly where high RE power is available, or where high performance is expected over time. The adaptive placement of VMs can use the same underlying virtualization technology (i.e., virtual resource scheduling), as already described in Section IV-A. To reduce the possibility of required migrations, a weather forecast of the following day should be included in the placement decision. In order to ensure the energy supply, energy providers have been using weather forecasts for a long time for the prognosis of energy from wind turbines. Recently, such systems have also been used for photovoltaics (PV). The weather forecast can estimate how much energy will be available in certain locations. With this information, the VMs can now be started exactly where energy from RE is expected. It also should be accepted that energy from RE is not available immediately, or not consistently. For example, if there is a lot of energy from PV at location A according to the weather forecast for the following day, VMs should be started there. In this case, it may be necessary to accept that the VMs are started, e.g., at 8 a.m. in the morning, but the power from PV is not available until 10 a.m. in an acceptable size. A software fed with relevant information can create a schedule for such scenarios to circumvent these deficiencies and possible additional costs. Virtual resources (like VMs) that recur regularly and are present for a longer period of time per day, can be started at locations with high RE. Weather forecasts can be requested from various weather services. Artificial neural networks (ANNs) or machine learning algorithms can also be used to create and improve the schedule. They can learn from the interaction with the virtualized resources and the user behavior, e.g., based on incoming and outgoing communication flows, as discussed in Section IV-C.

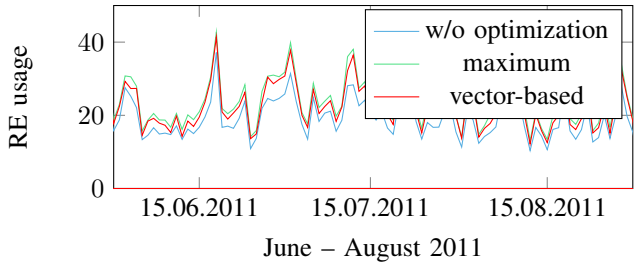
A. Considering Renewable Energies for Virtual Machine Placement Optimization

To evaluate potential benefits from leveraging renewable energies for the placement of VMs in cloud environments, this section presents an optimization of the placement and possible migration of virtual resources across geographically distributed data centers. Primarily, the optimization focuses on the utilization of fluctuating renewable energies at three locations in Germany. The optimization uses weather data from these distant locations, as already presented in [9]. Furthermore, as explained in detail in Section IV-D, affinities and dependencies between virtual resources have to be taken into account for the optimization. Therefore, traffic patterns from VMs running in the virtualization environment of the data center at Fulda University of Applied Sciences and their anonymized incoming and outgoing flows where used as the important secondary criterion for the optimization. The optimization was based on the algorithm introduced in [40]. It uses vector-based approach to iteratively search for virtual resources to migrate while taking the geographical distance and corresponding affinities as well as available renewable energy sources into account. This algorithm tries to maximize the level of utilization of renewable energies and also limits and prevents negative side effects like increased end-user latencies. As mentioned, the algorithm's primary goal of maximizing the usage level of renewable energies, used weather data for three data center locations in northern, central and southern Germany. This data included measurements of global solar radiation and wind speed in ten minute resolution over multiple years. The general feasibility of such an optimization approach was introduced in [9]. Furthermore, network flow data produced by our university data center were used. The data set contains flows between approximately hundred virtual resources, as well as flows identified to either come from or go to physical machines from these virtual resources.

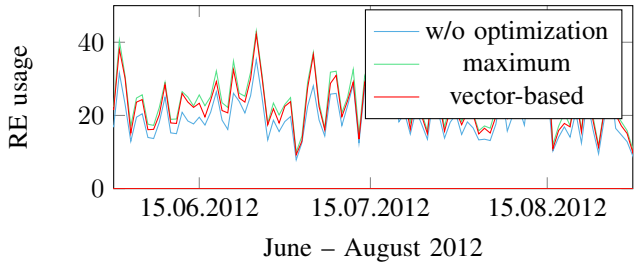
Based on this data, a RE usage optimization was conducted, which moved VMs between data centers in order to move energy consumers near to the energy producers, as discussed for the live migrations in this article in Section IV. Results for the years 2011, 2012 and 2013 are shown in Figure 7a, 7b and 7c. For better legibility and clarification the consecutive summer months June, July and August are chosen for the displayed period in the charts.

B. Optimized Communication Distances for Affinity Groups

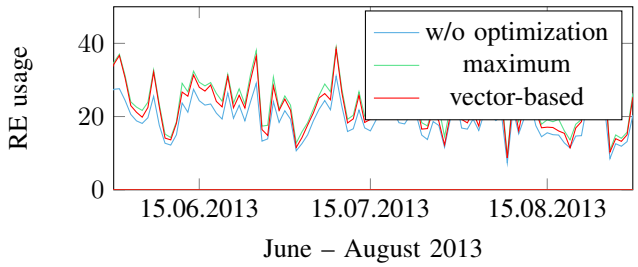
Beside the goal of raising the utilization of RE sources, the algorithm mentioned in the previous section also optimized the communication distances for affinity groups, so that the average distance each packet needs to travel across the network was reduced. This was accomplished by determining the geographical location for flow communication endpoints and by computing the average distance per packet. This not only assures traffic locality for affinity groups, but also reduces the average distance per packet to end-users. This results in a reduced end-user latency. The optimization used real-world network flow and affinity characteristics from the mentioned



(a) Optimized RE usage for the summer period 2011.

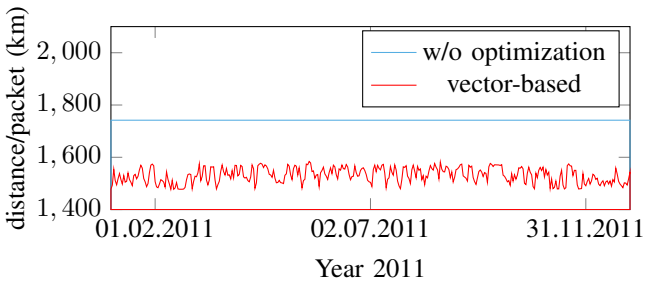


(b) Optimized RE usage for the summer period 2012.

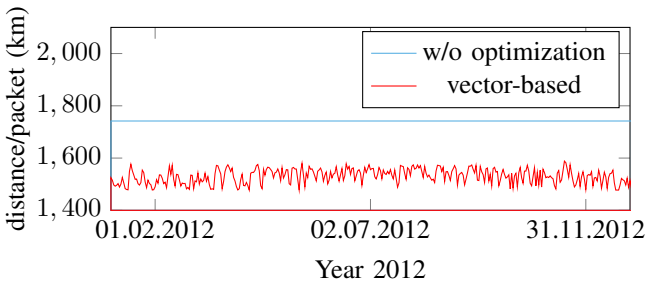


(c) Optimized RE usage for the summer period 2013.

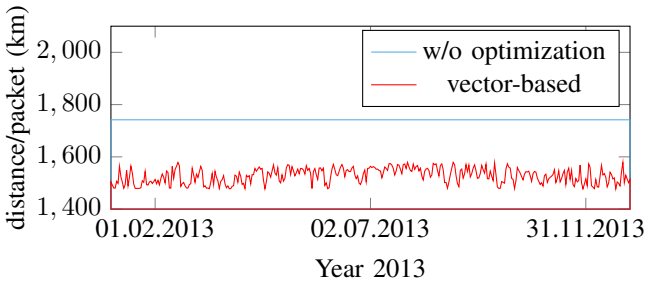
Figure 7. Optimized usage of renewable energy (RE) for the years 2011-2013.



(a) Optimized average distance per packet for the year 2011.



(b) Optimized average distance per packet for the year 2012.



(c) Optimized average distance per packet for the year 2013.

Figure 8. Optimization for the three years 2011-2013.

data set from VMs within a virtualization environment at Fulda University of Applied Sciences.

Figure 8a, 8b, and 8c show the results for the three years 2011, 2012 and 2013. The figures clearly illustrate that the average distance of communication endpoints was minimized. The results are summarized in Table II.

TABLE II. VECTOR-BASED ALGORITHM OPTIMIZED DISTANCE.

Year	Minimum [km]	Maximum [km]	Average [km]
2011	-265.55	-0.26	-211.21
2012	-265.59	-0.30	-211.34
2013	-265.59	-0.26	-214.39

The average communication distance was reduced between 211.21 km and 265.59 km per packet. This could lead to an approximate latency improvement of, e.g., one or two milliseconds for fiber-based networks.

VII. CONCLUSION AND FUTURE WORK

Energy costs are an important factor for today's IT infrastructures, due to rising energy prices and increasing power consumption. The virtualization offered for compute, storage and network resources, e.g., in private clouds, allows for a seamless and transparent migration of virtual resources due to the abstraction from the underlying hardware. These migration techniques can be used to enhance the energy efficiency in data centers and have been constantly evolving over the last decade. This includes adaptive migration, e.g., to consolidate or enhance the utilization of physical resources, as well as long-distance migration, which is not only covered by the related work and research presented in this article, but also by current virtualization and hypervisor products (e.g., the introduction of long-distance vMotion in VMware vSphere 6 that was previously already available in Microsofts Hyper-V). Regarding the energy efficiency, however, additional costs of the migration itself have to be taken into account. These costs can either directly (i.e., higher load on the physical compute, storage and network resources) or indirectly gain energy costs, e.g., if the migrated services and applications cannot provide the same service quality during the migration. Hence, to improve the energy efficiency by using live migration techniques offered in cloud infrastructures, the migration costs need to be minimized. This especially holds true, if the migration is used to benefit from lower energy prices or the availability of RE at distant data center sites.

Based on our previous research projects in this area, this article introduce an evaluation of the migration costs for I/O-intensive VMs in an OpenStack environment. Due to the incoming and outgoing network traffic, especially virtual network services operated in VMs typically have a large I/O footprint in the infrastructure that is typically compensated

by using hardware acceleration (e.g., virtual switch or kernel enhancements, DPDK, SR-IOV, FD.io, XDP etc.). To be able to measure the additional load caused by a live migration of such services, and to quantify the impact on the service quality, additional tools (i.e., *stress*, *openssl*, *dd*, *find*) are used to add artificial I/O load on the machines while migrating them to another physical host in the OpenStack infrastructure. Based on the findings presented in this article, the migration time increases proportionally to the added artificial I/O load. Furthermore, the load on storage and network resources grows accordingly as expected. The burden of the ongoing live migration can especially be measured if more than 80% of the total memory of the VM are continuously utilized and changed. Interestingly, the migration time can be reduced by increasing the number of concurrent live migrations. This is due to the impact of the scheduler and message bus, handling the migrations in OpenStack together with libvirt and KVM. Similar effects can be observed with other hypervisors like vSphere or Hyper-V, though these products typically limit the number of parallel live migrations to smaller values.

The results of the experiments show a significant performance decrease for I/O read operations on the VMs during the migration. This conspicuous effect is likely due to limited available buffer/cache and extensive flush operations during the migration. The impact on the underlying OpenStack infrastructure leveraging libvirt and KVM, can also be observed in a performance decrease during start of VMs with high I/O and memory load, even if the VMs are running on separate hosts using different block storage. Several I/O operations (i.e., using *dd*, *find*, *stress*) were used to evaluate this decrease while constantly monitoring the service quality of the main operation. During the migration, a *find* process across the files on the VMs experienced a significant performance decrease. Also, VMs running on the target machine for the migration, experience a significantly reduced performance during this period. Moreover, for high additional artificial I/O loads, the main operation of the virtual network service was also impacted accordingly. Response times on the migrated service (offering the function of a web proxy) increased from 0.166s to 0.785s during the migration. The high I/O load on the VMs leads expectedly to higher overall response times as more and more VMs are consolidated on a single physical host. However, a previous paper [6] presented an expected increase of the overall energy efficiency due to the higher utilization of the physical host, made possible by this consolidation.

Building on the results presented in this article, we are currently focusing our research on live migration techniques for containers as a lightweight virtualization alternative compared to full-size VMs. Some types of services allow migration and scaling by simply destroying the containers at one site and respawning them at another. The required live migration techniques for containers are still being developed (e.g., in CRIU [31]) and are also within the focus of some related research projects. Initial results of our experiments show that the transferred amount of data during container migrations is expectedly less compared to VMs. Conversely, the migration process itself is more difficult, as the entire state of a process stack in the operating system needs to be stored and transferred. Existing checkpoint and restore techniques need to be extended to support live migration of container-based virtual network services. As virtualization techniques like containers are evolving, the requirement to seamlessly migrate virtual resources is likely to grow. Additionally, virtualization techniques themselves make heavy use of virtual network functions, for example to form overlay networks for distributed container networks, e.g., using virtual routers, switches, load balancers or firewalls in distributed clouds, offering potential for energy-efficient migrations of virtual network services and resources in upcoming cloud infrastructures.

In this article, affinity groups and dependencies between migrated and adaptively placed virtual resources have been identified and considered for the optimization. This way, possible negative side effects of migrations, e.g., leading to high access latencies or higher communication overhead after the migration or separation of highly dependent resources were addressed. Concerning the identification of affinity groups of VMs and virtual network services, an additional classification of communication endpoints could be viable to prioritize and weight traffic for different virtual resources (e.g., staff/customer machines or central services). Furthermore, a comparison of the raised migration costs in relation to the benefits of the optimization would be helpful to rate the quality and effectiveness of the developed algorithms. Moreover, an energetic estimation of communication efforts based on the distance packets need to be sent across, can shed light on energy savings in computer networks based on real world scenarios.

As a potential beneficial use case of the optimization presented in this article, the use of renewable energies was discussed. Migrating as well as consolidating virtual resources at sites with currently high renewable energy power or low energy prices was presented. Besides the migration of virtual resources, this could also include an energy-aware placement of virtual resources in the first place. If affinity groups and dependencies of virtual and physical resources are also considered for the optimization, this allows for a better overall utilization of renewable energies for data centers owing the location-independent placement and movability of virtual resources. The importance of affinity groups and dependencies is especially important for the virtual network services. The network not only enables distributed services, but also introduces costs either directly from operating the communication systems and links or indirectly, e.g., resulting from additional latencies when using these services. While these dependencies where addressed in the optimization algorithm used in this article, a further optimization regarding involved costs could be to use forecasts of available renewable energies at each site in distributed cloud infrastructures. Further research needs to be carried out in this area, to profit from the fluctuation of renewable energies. This includes the investigation of the use of machine learning techniques to improve forecasts for available renewable energies as well as dependencies and affinities (e.g., based on requirements of network flows between the virtual services and end-users). The data sets used for weather data at three different locations in Germany and the network flow data of virtual machines at Fulda University of Applied Sciences will be used as a starting point for further research projects in this area.

References

[1] N. K. Truong, C. Pape, S. Rieger, and S. Reißmann, "Energy-efficient live migration of I/O-intensive virtual network services across distributed cloud infrastructures," in ICSNC 2017, The Twelfth Interna-

tional Conference on Systems and Networks Communications, 2017, pp. 6–11.

[2] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," ACM SIGCOMM Computer Communication Review, vol. 39, no. 1, Dec. 2008, pp. 68–73.

[3] T. Cheocherngngarn, J. H. Andrian, D. Pan, and K. Kengskool, "Power efficiency in energy-aware data center network," in Proceedings of the Mid-South Annual Engineering and Sciences Conference, 2012.

[4] A. Andreyev, "Introducing data center fabric, the next-generation Facebook data center network," 2014, URL: https://code.facebook.com/posts/360346274145943/introducing-data-center-fabric-the-next-generation-facebook-data-center-network/, 2018-05-26.

[5] P. Mell and T. Grance, The NIST definition of cloud computing. Washington DC: National Institute of Standards and Technology, 2011, URL: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf, 2018-05-26.

[6] K. Spindler, S. Reissmann, and S. Rieger, "Enhancing the energy efficiency in enterprise clouds using compute and network power management functions," in ICIW 2014, The Ninth International Conference on Internet and Web Applications and Services, 2014, pp. 134–139.

[7] ETSI, "Network Functions Virtualisation (NFV); Infrastructure Overview," 2015, URL: {http://www.etsi.org/deliver/etsi_gs/NFV-INF/001_099/001/01.01.01_60/gs_NFV-INF001v010101p.pdf}, 2018-05-26.

[8] S. Tilkov, "The Modern Cloud-Based Platform." IEEE Software, 2015.

[9] K. Spindler, S. Reissmann, R. Trommer, C. Pape, S. Rieger, and T. Glotzbach, "AEQUO: Enhancing the energy efficiency in private clouds using compute and network power management functions," International Journal On Advances in Internet Technology, vol. 8, no. 1 & 2, 2015, pp. 13–28.

[10] A. Beloglazov and R. Buyya, "Energy efficient resource management in virtualized cloud data centers," in Proceedings of the 2010 10th IEEE/ACM international conference on cluster, cloud and grid computing. IEEE Computer Society, 2010, pp. 826–831.

[11] ——, "OpenStack neat: A framework for dynamic consolidation of virtual machines in OpenStack clouds–a blueprint," Cloud Computing and Distributed Systems (CLOUDS) Laboratory, 2012.

[12] C. Pape, S. Rieger, and H. Richter, "Leveraging Renewable Energies in Distributed Private Clouds," MATEC Web of Conferences, vol. 68, Aug. 2016, p. 14008.

[13] A. Song, W. Fan, W. Wang, J. Luo, and Y. Mo, "Multi-objective virtual machine selection for migrating in virtualized data centers," in Pervasive Computing and the Networked World. Springer, 2013, pp. 426–438.

[14] N. A. Singh and M. Hemalatha, "Reduce Energy Consumption through Virtual Machine Placement in Cloud Data Centre," in Mining Intelligence and Knowledge Exploration. Springer, 2013, pp. 466–474.

[15] A. Verma, P. Ahuja, and A. Neogi, "pMapper: power and migration cost aware application placement in virtualized systems," in Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware. Springer-Verlag New York, Inc., 2008, pp. 243–264.

[16] D. Versick and D. Tavangarian, "CAESARA-Combined Architecture for Energy Saving by Auto-adaptive Resource Allocation." in DFN-Forum Kommunikationstechnologien, 2013, pp. 31–40.

[17] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, "ElasticTree - Saving Energy in Data Center Networks." NSDI, 2010.

[18] T. Huong, D. Schlosser, P. Nam, M. Jarschel, N. Thanh, and R. Pries, "ECODANE-reducing energy consumption in data center networks based on traffic engineering," in 11th Würzburg Workshop on IP: Joint ITG and Euro-NF Workshop Visions of Future Generation Networks (EuroView2011), 2011.

[19] V. Mann, K. Avinash, P. Dutta, and S. Kalyanaraman, "VMFlow: Leveraging VM Mobility to Reduce Network Power Costs in Data Centers." Networking, vol. 6640, no. Chapter 16, 2011, pp. 198–211.

[20] W. Fang, X. Liang, S. Li, L. Chiaraviglio, and N. Xiong, "VMPlanner: Optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers," Computer Networks, vol. 57, no. 1, 2013, pp. 179–196.

[21] X. Wang, Y. Yao, X. Wang, K. Lu, and Q. Cao, "Carpo: Correlation-aware power optimization in data center networks," in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 1125–1133.

[22] Y. Han, J. Li, J. Y. Chung, J.-H. Yoo,, and J. W.-K. Hong, "SAVE: Energy-aware Virtual Data Center embedding and Traffic Engineering using SDN." NetSoft, 2015, pp. 1–9.

[23] T. Kim, T. Koo, and E. Paik, "SDN and NFV benchmarking for performance and reliability." APNOMS, 2015, pp. 600–603.

[24] W. Hu, A. Hicks, L. Zhang, E. M. Dow, V. Soni, H. Jiang, R. Bull, and J. N. Matthews, "A quantitative study of virtual machine live migration," in Proceedings of the 2013 ACM cloud and autonomic computing conference. ACM, 2013, p. 11.

[25] K. Rybina, A. Patni, and A. Schill, "Analysing the migration time of live migration of multiple virtual machines." in CLOSER, 2014, pp. 590–597.

[26] S. Ghorbani, C. Schlesinger, M. Monaco, E. Keller, M. Caesar, J. Rexford, and D. Walker, "Transparent, live migration of a software-defined network," in Proceedings of the ACM Symposium on Cloud Computing, ser. SOCC '14, 2014, pp. 3:1–3:14. [Online]. Available: http://doi.acm.org/10.1145/2670979.2670982

[27] U. Deshpande, X. Wang, and K. Gopalan, "Live gang migration of virtual machines," in Proceedings of the 20th international symposium on High performance distributed computing. ACM, 2011, pp. 135–146.

[28] T. Lu, M. Stuart, K. Tang, and X. He, "Clique migration: Affinity grouping of virtual machines for inter-cloud live migration," in Networking, Architecture, and Storage (NAS), 2014 9th IEEE International Conference on. IEEE, 2014, pp. 216–225.

[29] G. Sun, D. Liao, D. Zhao, Z. Xu, and H. Yu, "Live Migration for Multiple Correlated Virtual Machines in Cloud-based Data Centers," IEEE Transactions on Services Computing, vol. PP, no. 99, 2015, pp. 1–1.

[30] C. Canali and R. Lancellotti, "Improving scalability of cloud monitoring through pca-based clustering of virtual machines," Journal of Computer Science and Technology, vol. 29, no. 1, 2014, pp. 38–52.

[31] K. Kolyshkin, "CRIU: Time and space travel for linux containers," 2015, URL: http://de.slideshare.net/kolyshkin/criu-time-and-space-travel-for-linux-containers, 2018-05-26.

[32] OpenStack Foundation, "OpenStack administration guide," 2017, URL: http://docs.openstack.org/admin-guide-cloud/index.html, 2018-05-26.

[33] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live migration of virtual machines," in Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation-Volume 2. USENIX Association, 2005, pp. 273–286.

[34] M. R. Hines, U. Deshpande, and K. Gopalan, "Post-copy live migration of virtual machines," ACM SIGOPS Operating Systems Review, vol. 43, no. 3, 2009, p. 14.

[35] A. Strunk, "Costs of virtual machine live migration: A survey," 2012 IEEE Eighth World Congress on Services, 2012, pp. 323–329.

[36] Microsoft, "Virtual machine live migration overview," 2015, URL: https://technet.microsoft.com/en-US/library/hh831435.aspx, 2018-05-26.

[37] R. Hofstede, P. Celeda, B. Trammell, I. Drago, R. Sadre, A. Sperotto, and A. Pras, "Flow Monitoring Explained - From Packet Capture to Data Analysis With NetFlow and IPFIX." IEEE Communications Surveys and Tutorials, 2014.

[38] J. Quittek, T. Zseby, B. Claise, and S. Zander, "Requirements for IP Flow Information Export (IPFIX)," Internet Requests for Comments, RFC Editor, RFC 3917, October 2004, 2018-05-26. [Online]. Available: http://www.rfc-editor.org/rfc/rfc3917.txt

[39] ovs-vsctl, Open vswitch manual ed., The Linux Foundation, 1 Letterman Drive, Building D, Suite D4700, San Francisco CA 94129, 2017, 2018-05-26. [Online]. Available: http://openvswitch.org/support/dist-docs/ovs-vsctl.8.pdf

[40] C. Pape, S. Rieger, and H. Richter, "Leveraging Renewable Energies in Distributed Private Clouds," in Proc. International Conference on Advances on Clean Energy Research (ICACER 2016), 16-18 April 2016, pp. 26–30. [Online]. Available: http://www.icacer.com

Using Cisco VIRL and GNS3 to Improve the Scale-out of Large Virtual Network Testbeds in Higher Education

Sven Reißmann*, Sebastian Rieger†, Christoph Seifert†, Christian Pape†
*Datacenter
Fulda University of Applied Sciences, Fulda, Germany
Email: sven.reissmann@rz.hs-fulda.de
†Department of Applied Computer Science
Fulda University of Applied Sciences, Fulda, Germany
Email: {sebastian.rieger, christoph.seifert, christian.pape}@cs.hs-fulda.de

***Abstract*—Over the last years, education paradigms developed from the traditional classroom learning to novel approaches like e-learning and blended learning. Especially blended learning, which combines the traditional approach with e-learning independent of time and place, is an important concept to increase the quality of study programmes. For the creation of laboratory setups in higher education dealing with computer networks, virtualized network environments have been continuously gaining momentum. Depending on the desired practical or theoretical orientation, they can be implemented using different paradigms, as well as corresponding hardware or software solutions. A high practical relevance and functional realism can be achieved using network emulation. However, emulation requires more resources compared to network simulators, due to the complexity of realistic network functions. To offer emulated virtual network environments, e.g., for a large number of participants in higher education courses, scalable virtualization backends and cluster solutions are necessary. In this article, we provide an overview over available paradigms to create networking experiments in higher education classes. We also present a number of requirements, which we identified to be important in the context of the networking laboratory (NetLab) of Fulda University of Applied Sciences. Based on these requirements, the available software solutions were compared and the best matching solutions were selected for the use in our laboratory. The scalability and performance evaluation of virtual network testbeds presented in this article, outlines the suitability of each compared network virtualization and emulation solution for higher education courses, and discusses possibilities for further improvements.**

Keywords–Network Virtualization; Network Emulation; Higher Education; VIRL; GNS3.

I. INTRODUCTION

In recent times, the paradigms for teaching in higher education have been evolving to include concepts like blended learning. This paradigm shift is especially helpful for hands-on laboratory exercises, as they typically include for example virtual testbed setups for larger class sizes as well as extending the use of the experiments and testbeds beyond the laboratory or classrooms. Furthermore, offering e-learning and blended learning content is a necessity today to increase the quality of study, e.g., by including complex and practical examples, and to keep pace with the rapid development of online courses and the mobility of students as well as lifelong learning. This is especially true for computer network labs, where real-world test setups are needed to complement lectures and theoretical models with practically relevant exercises [1].

To provide such environments, various approaches are possible, depending on the intended focus on theoretical or practical relevance. Figure 1 gives an overview of corresponding gradations of possible approaches, including references to some of the appropriate tools available. While the implementation of testbeds in real-world networks, e.g., campus networks of organizations and universities as shown at the left end of the figure, would provide the most realistic testbed, the risk of interference with regular operation and availability of the production network forbids this option in most cases. To overcome this problem, a separate physical testbed can be created apart from the production network. However, bootstrapping such a testbed with realistic characteristics is complex and expensive, hence making fast adaption to varying requirements and ever-changing network environments far from realistic. Virtual testbeds can reduce the setup cost immensely, but besides the additional effort and complexity introduced by the virtualization, still a lot of manual work is required for setting up and providing the required networking components and interfaces, virtual networks and so on.

Theoretical models, as shown on the right of Figure 1, take a completely different approach by abstracting complex network topologies and protocols in the model. This is specifically useful when designing experimental protocols, but the implementation of such formal specifications - or even the transfer of the insights learned - in the real world can be quite challenging due to missing practical orientation. Simulations, though also based on an abstract model of the network and protocols, improve the practical relevance by trying to replicate real-world characteristics. One of the big advantages of simulations is the capability of exact modeling of real-world behavior, such as timing or transmission quality. Another advantage of deterministic simulation is the option of changing the simulation speed. However, only an abstract network is modeled by a simulation, which does not fully reflect the characteristics of a real network.

To resolve these issues, emulation is recommended in [2] as the means of choice to implement virtual network testbeds. The authors come to this conclusion by evaluating the goals and advantages of emulation using the following criteria: *Functional Realism*, *Timing Realism*, *Traffic Realism*, *Topology Flexibility*, *Easy Replication*, and *Low Cost*. In their research, they show that emulation only lacks in the area of *Timing Realism* while fulfilling all other evaluation criteria. Therefore, emulation provides a flexible foundation for experimental network testbeds being positioned in the middle between practical implementations with high relevance for real-world environments and accurate but typically rather

abstract theoretical models. Besides the evaluation in [2], the advantages perceived by using emulation for virtual network testbeds with a high practical relevance is also described in [3], [4] and [5].

In addition, while lectures typically last about 90 minutes, the work in labs needs time for preparation and post-processing, which means that often only about 60 minutes are available for the hands-on exercises. This period is often not enough for in-depth exercises, especially when dealing with an abstraction or simplification of complex setups with high practical relevance. Also, the effort to provide testbeds increases significantly for large numbers of participants in laboratory courses. By choosing an emulation approach, much greater flexibility is gained also in terms of accessing the virtual test environment. Where formerly students had to attend in person, i.e., to be able to use the equipment on-site, they are now enabled to work basically from any location as long as sufficient Internet access is provided. Hence, emulated environments are currently primarily used for higher education courses in the networking laboratory (NetLab) at the Applied Computer Science department of Fulda University of Applied Sciences. These environments support the argument for emulation environments as presented in [2], which were also recently updated in [6]. The advantages of emulation-based environments, described in detail in the remaining sections of this article are also further amplified by the possibility to use them to improve the reproducibility of research experiments as described in [6] and also in [7] and [8]. This way, students in higher education, e.g., master courses, can get an insight into current research in the area of computer networks, use explorative learning to understand and discuss the findings, leading to a deeper understanding and developing the ability to carry out own small research projects and individually contribute to applied sciences, e.g., in papers adjacent to the master thesis. Emulated testbeds and experiments being developed by the students can in turn be used in research projects supporting the students to become junior researchers.

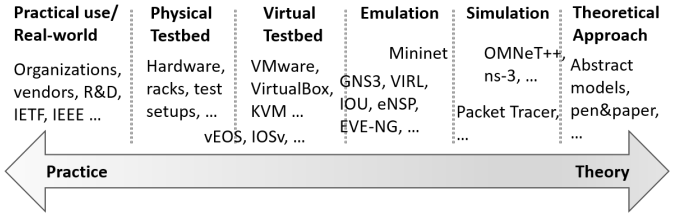


Figure 1. Classification of experimental scientific networking testbeds.

Nevertheless, the emulation of scalable virtual network testbeds results in high resource requirements due to the number of virtual machines needed, especially for large research projects or higher education courses. In this article, we present an analysis of these resource requirements of virtual network testbeds based on experiences from providing such an environment in the NetLab. Solutions that were evaluated and used in the NetLab are shown in Figure 1 (e.g., GNS3, EVE-NG, VIRL). Primarily, currently Cisco's Virtual Internet Routing Lab (VIRL) is used. Advantages of this solution will be presented in this article. Further, we discuss options to improve the scalability of such an environment.

The remaining part of this article is laid out as follows. The following Section II discusses related work in the area of virtualized networking laboratories and the effectiveness of such labs in higher education courses. Section III gives examples for emulated network topologies used in higher education courses in the NetLab. Based on these use-cases, requirements for virtual networking testbeds, and software products available to implement such testbeds, based on the classification and examples shown in Figure 1, are presented. Our experiences during the implementation of such a solution are discussed in Section IV, where we present our selection of tools for creating a scalable platform for virtual networking testbeds. Next, an evaluation of the scalability and performance of the VIRL environment is presented in Section V. As an alternative to VIRL, Section VI presents a comparison to the scalability and performance of GNS3 using the same virtualization environment and methodology as for the VIRL benchmark. Finally, a conclusion and future work can be found in Section VII.

II. Related Work

Services in today's cloud-driven infrastructures are based on sophisticated network topologies, which interconnect various network and server components, and build the foundation for scalability, redundancy and high availability of IT services. Given this practical relevance, it seems obvious that training in the lab is essential for computer science students to gain practically relevant knowledge of the theoretical concepts taught in lectures. Therefore, the simulation and emulation of such network topologies for teaching higher education classes as well as employing them in scientific projects is subject of current research. In [4], a virtual environment based on VIRL is compared to physical setups using Cisco Certified Network Associate (CCNA) pods and network simulation software like Cisco Packet Tracer. Furthermore, the cost, setup requirements and limitations are estimated and reviewed. The solutions discussed in this article address these limitations. A similar comparison is presented in [9], where the network emulator GNS3 is compared to Cisco Packet Tracer. The use of VIRL and GNS3 in the area of research and education are also discussed in [5] and [10], respectively. However, these papers do not address the scalability for VIRL and GNS3 in large higher education courses. Also, they do not discuss their didactical suitability and characteristics for the use in higher education. An extensible and scalable emulation/simulation framework based on a declarative XML-based language (as also used in VIRL) for modeling and evaluation of large network topologies is explained in [11]. This open-source toolset is also compared to other well-known simulation and emulation environments like ns-3 or PlanetLab. A brief introduction to model and simulate interconnected autonomous systems using the Boson Network Simulator is described in [12]. An open-source tool for simulating protocol behavior in congested networks by throttling network links and controlling delay and packet loss is described in [13]. This software is widespread and used in other emulation products to alter link transmission properties in network topologies. The didactics-oriented software called Netkit to model and emulate a wide-range of real-world devices is outlined in [3]. The goal of this emulation environment software is to setup networking experiments at low cost and with little effort. While these papers focus on large topologies and didactical suitability, they do not address the performance and scaling for large higher

education courses. A framework for reproducible container-based emulation of networking experiments is presented in [2]. In [14], network topologies are emulated by using virtual routers on Linux-based hosts. The real-time network emulator EmuNET used for testing protocol and application behavior in network topologies, is introduced in [15]. However, these solutions do not allow the use of real-world network operating systems and networks components (e.g., virtualized Cisco or Juniper equipment). The application of a cloud-based virtual environment for security related education is outlined in [16]. It is based on a platform called V-Lab, which utilizes open-source virtualization technologies, and software defined networking (SDN) solutions. Finally, an evaluation of the effectiveness of virtual laboratory environments for student learning is performed in [17]. In the paper, a course taught at the University of Massachusetts Amherst, which consisted of multiple lectures, homework assignments, and lab assignments, is evaluated. The paper tries to quantify student learning in lectures and labs, and the author concludes that learning indeed occurs almost equally as much during lab sessions as in lectures. This also coincides with the results of a study conducted by Stanford University researchers, who assigned the task of reproducing results from over 40 papers in the research area of computer networks to over 200 students [18]. They found that this kind of practical training is interesting for the students, and gives them the opportunity to understand topics of current research, or even interact with researchers. While these papers are useful for the evaluation of didactical suitability and characteristics of virtual laboratory network testbeds, they did not address the scalability and performance for the use in higher education courses discussed in this article.

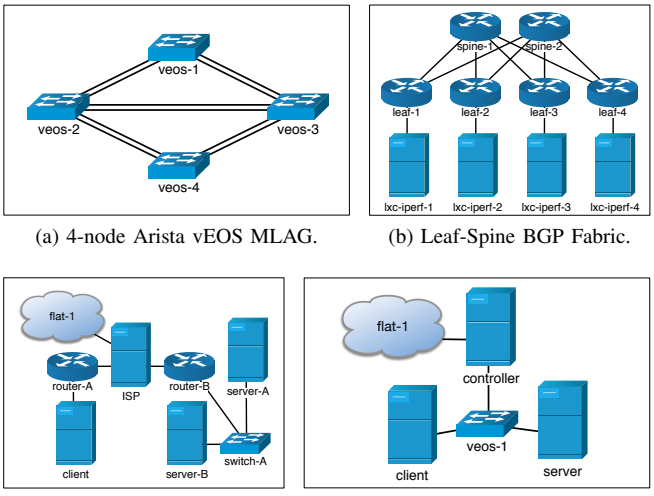
III. Virtual Environments for Network Testbeds

Following on from the consideration of emulation testbeds in [2], various possibilities of assisting lectures with practical exercises in our networking laboratory (NetLab), being described in this article, were evaluated. The following sections will first provide an overview of the approaches and the laboratory scenarios used in some of our courses, to present concepts and ideas behind the experiments carried out by the students. Afterwards, requirements for an implementation of a virtual environment are derived from the characteristics of these approaches and concepts. Finally, an overview of the candidates for a concrete implementation, currently used in the NetLab, are presented.

A. Examples for Exercises in the NetLab at Fulda University

The networking laboratory at the Applied Computer Science department of Fulda University of Applied Sciences provides practical relevant exercises in the field of computer network development, configuration and operation. Besides computer networks themselves, exercises also include distributed systems, e.g., internet, cloud or multimedia services, as well as network security or network management and monitoring. Thus, the laboratory supports lectures and further allows students to perform independent experiments to improve their knowledge and expertise in areas related to the indicated topics.

Refer to Figure 2 for some examples of network topologies used in our NetLab environment for student laboratory exercises. Figure 2a shows a topology consisting of four Arista vEOS (4.16.9) nodes with redundant links between them. The topology is one out of many used by master's students to understand and troubleshoot real-world networks. In this specific case, students team up to investigate the impact of the Spanning Tree Protocol (STP) in various data center networking scenarios using technologies like MSTP, LACP and MLAG. Based on similar exercises, the master's students also work on Leaf-Spine-based topologies including BGP fabrics, which are widely used in web-scale data centers by companies like Facebook or Microsoft (Figure 2b). Here, the endpoint nodes are based on Ubuntu 14.04 GNU/Linux containers (LXC), while the spine and leaf switches are driven by Cisco IOSv (15.6(2)T). In the past, we also implemented similar topologies using vEOS (BGP) and NX-OSv (FabricPath). A great advantage of this setup is the flexibility we provide for the students. LXC containers can be managed using standard Linux commands via SSH, nodes can be started and stopped in the middle of a running emulation, network links can be connected or disconnected, and it is also possible to configure various QoS parameters like delay, jitter or packet loss on the links. Beyond that, traffic entering or leaving a specific interface can be collected and investigated using Wireshark. SDN-based scenarios, including the OpenFlow controller *OpenDaylight* and OpenFlow 1.3 capable *Arista vEOS* switches, can be explored using the topology shown in Figure 2d. Since vEOS behaves identical compared to the EOS-based Arista hardware switches, by choosing VIRL over the previously used Mininet, students can test OpenFlow deployment in a near real-world environment.



(a) 4-node Arista vEOS MLAG. (b) Leaf-Spine BGP Fabric.

(c) Simple Toubleshooting. (d) vEOS/OpenDaylight SDN topology.

Figure 2. Examples of emulated network topologies for student exercises.

An example of a much simpler network topology is shown in Figure 2c. Students in bachelor programmes troubleshoot network misconfiguration (i.e., ARP, routing, delay, packet loss, port status) and discover the underlying network topology by using tools like ping, traceroute/mtr and Wireshark, before they establish connections to an Apache web server running on node server-B. Again, all server and client nodes are based on Ubuntu 14.04 LXC, while switches in this scenarios are based on IOSvL2 (15.2(4.0.55)E). For realistic WAN emulation we use the standard Linux network emulation *netem* on the ISP node to inject delay and add packet loss. The node named ISP

acts as a default gateway for the emulated topology, which is connected to the physical local area network for NetLab projects (flat-1). Thus, the invocation of commands like ping or traceroute targeting hosts on the Internet (i.e., google.com) is possible from inside the emulated environment, enhancing the practical relevance of the exercises. Furthermore, from within the NetLab, students can connect to their nodes in the emulated network using OpenVPN, for instance to configure the web server.

All examples and topologies mentioned in this section are under continuous development and are actively used in the NetLab. Corresponding topology configurations for VIRL are managed using Git and can be downloaded from [19].

B. Requirements for Virtual Networking Testbeds

The NetLab is currently equipped with 20 PCs and additional workspaces for notebooks, allowing students to bring their own devices to the laboratory to extend the lab equipment or to extend the experiments over the teaching time as described for e-learning and blended learning approaches in Section I. These workspaces are arranged in so-called ”islands”, so that students can work together in groups of four, and share a pre-packed experimental rack filled with networking equipment, such as routers, switches, and firewalls.

As mentioned in the introduction, working in the laboratory requires additional effort for preparation of the test setups (i.e., cabling the experimental racks and setting up an initial configuration) before the actual experiments can begin, as well as for cleaning up after the experiments are finished. Hence, often only 60 minutes are left for the hands-on exercises, which is not enough time for in-depth practical training, and the lab's state can not easily be preserved to span over multiple lessons when physical network components are being used. For this reason, in our networking lab we aim to provide a combination of classroom teaching and novel approaches like e-learning and blended learning to the students. Each of these approaches introduces demands on the learning environment, the time for preparation in the lab, or possibilities of external access.

- **Classroom teaching** — The traditional on-site learning in a classroom. In order to fulfill the time restrictions of classroom learning with a typical duration of 90 minutes for each session, the preparation of exercises to a predefined state must be considered. Further, the current state of an exercise should be storable and resumable at a later time to allow students to continue, discuss, share and present their work.
- **E-learning** — A teaching method that enables high flexibility in learning regardless of the times and locations of conventional classroom sessions. It must be possible to attend a course without ever visiting a traditional classroom. A fully virtual lab is required to allow students to login from home at any time and conduct an experiment using their own or individual resources.
- **Blended learning** — This can be seen as a combination of the previous paradigms, where students attend classroom sessions, but have the opportunity to finish exercises at any time by accessing the virtual lab from the NetLab outside of teaching time, but also from at home or any place providing sufficient Internet access. In addition, students are enabled to conduct their own experiments in order to deepen their teaching content.

The paradigm of blended learning offers continuing learning and collaboration between students together with lecturers and tutors, combining the advantages of classroom teaching and e-learning. Therefore, we followed this idea for the higher education computer network courses discussed in this article. We experienced an increase in student groups using the labs outside of regular class hours throughout the last semesters, supporting our decision.

Regardless of the paradigm chosen, the implementation must meet some of the requirements already outlined in [2], which are essential to ensure that the behavior of real networks and protocol peculiarities can be observed by students.

- **Functional realism** — Each system must provide the same functionality as its pendant in real hardware. Ideally, this can be achieved by executing exactly the same program code that is also used on real-world systems, i.e., by using real hardware or virtualizing the operating system images of real routers, switches and client systems.
- **Timing realism** — The timing behavior of all systems in the test bed should be indistinguishable from real physical systems. This is especially important for exercises where traffic behavior is inspected with tools like ping or tcpdump.
- **Traffic realism** — It should be possible to inspect real network traffic in the networking testbed. Client systems should be capable of generating and receiving network traffic from users and systems on the local network, or even on the Internet.
- **Topology flexibility** — It should be possible to create new topologies of various kinds without great effort, in order to be able to optimally map the requirements of the various courses.
- **Easy replication** — It should be possible to duplicate existing topologies without great effort, so that they can be used by different groups of students independently.
- **Scalability at low cost** — The environment must scale for large numbers of students, while at the same time minimize administrative effort and financial costs.
- **Didactical reduction** — The implementation approach must allow students to focus on the essential learning materials, as there is not enough time in classroom to understand complex simulation software before the actual experiment can begin.

Figure 3 shows the characteristics of different networking testbed approaches as introduced in Figure 1 and defined as well as evaluated in [2]. In the lower part of the figure, the didactical suitability of the approaches as perceived in several computer networking related courses in the NetLab was added. These didactical aspects are discussed in more detail in [20]. However, from this perception, it can be seen by evaluating positive (+) and negative (-) values while ignoring a neutral value (˜) that emulation is still the best option. Simulation (6 points) is offering better suitability for blended learning and e-learning, since it is lightweight, provides exact timing and the state can be prepared and restored even when using the

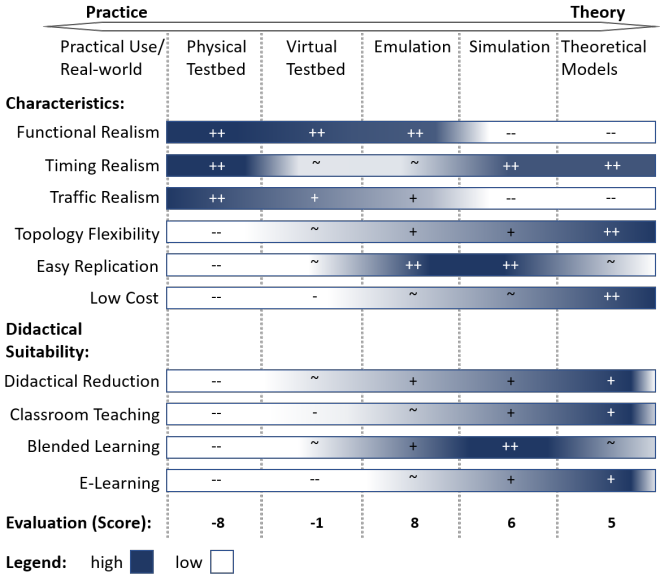


Figure 3. A comparison of networking testbed approaches.

course material remotely. This also holds advantages when using simulations in the classroom. Theoretical models (5 points), however, can be difficult to understand for students leading to a worse suitability for blended learning. As already discussed in this article, physical (-8 points) and virtual (-1 point) testbeds require much effort not only to build them, but also to prepare their use during lectures and laboratory courses. Hence, these options also received a low score in their didactical suitability observed in the NetLab. Regarding didactical reduction, emulation, simulation and theoretical models can use abstraction to hide the complexity, e.g., by using prepared experiments, models or formulas. Overall, emulation (8 points) still offers the best overall score, when taking the didactical suitability identified from courses in the NetLab into account.

C. Approaches for Implementing Virtual Network Testbeds

Blended learning techniques require ubiquitous access, as well as the possibility to easily comprehend and modify the experimental environment. Hence, integrated environments like simulators and emulators are the method of choice to meet the requirements of high realism and practical use.

Emulation can be seen as a compromise between theory and practice, especially when it is used to implement virtual network testbeds. It allows to deploy real-world operating systems (i.e., GNU/Linux servers, Windows clients) and utilize typical network management tools, like Wireshark or iperf. In the NetLab *physical and virtual testbeds*, as well as *emulation* and *simulation* have been used over the past years to support higher education courses and research projects. In accordance with the results discussed in [2], emulation has proven to be a particularly flexible solution. Physical testbeds are provided in the lab in form of pre-packed experimental racks to allow realistic student projects and Cisco certifications (i.e., CCNA, CCNP) [21]. However, due to the complex and time-intense preparation of the environment, these physical testbeds are not suitable for short-term exercises and lab sessions, in which students should carry out experiments, e.g., to see the practical use of theoretical concepts presented in a corresponding lecture.

Yet, the realization of *virtual testbeds* (i.e., using a distributed approach with VMware Workstation or a central approach with VMware vSphere ESXi) doesn't require less effort, as the preparation and maintenance of the virtual machines and networks is time-consuming. For this reason, in the NetLab *virtual testbeds* are mostly used for practically relevant client-server applications and experiments in the IT security area. The constant need to install software updates in the virtual machines used for these testbeds and adapt them to changes in the surrounding laboratory environment throughout the semester, requires additional effort. Simulation software like ns-3 [22] or OMNeT++ [23] is mentioned in some lectures in master's programmes, but not currently used for practically relevant experiments in the lab.

Practical training for the previously mentioned CCNA certification includes the use of Cisco Packet Tracer [24]. However, in some lecture exercises students criticized the missing practical relevance. For example, network clients (i.e., PCs) in Packet Tracer are simulated and do not provide feature-complete implementations of common network tools, such as arp, ping or traceroute. Another drawback of the simulation is that there are peculiarities, which only appear in the simulation and need to be specifically explained to students. One example of such behavior is that in case of an ICMP PING, the first packet will be dropped at the router in the destination network until the destinations MAC address has been determined using ARP. While this behavior is correct, it typically can't be observed in a real network, where almost any client starts to send packets to the router immediately after booting the operating system, hence its MAC address is already in the routers ARP table, when sending ICMP PING packets. Besides this lack in *Traffic Realism* and the stated lack in *Functional Realism*, e.g., due to missing common tools in the simulated clients and network components, there is also no *Timing Realism* achieved within Packet Tracer, which is an even bigger problem for research projects.

The software Mininet [25], mentioned in [26] and [2], is also provided in our NetLab, where it is mainly used for experiments in the SDN area. The resource requirements of Mininet are extremely low, due to a container-based approach, which allows to emulate huge topologies with hundreds or thousands of individual nodes. Therefore, Mininet is typically the best choice for large topologies and complex network management and automation implementations, e.g., in research projects. Still, the creation of complex topologies in Mininet requires decent knowledge of the underlying Python-API, which again is time-consuming in short-term courses. Further, it is not possible to use or connect arbitrary real-world network components in Mininet topologies, limiting the practical relevance of the experiments carried out in Mininet. Also, images of real-world network equipment, e.g., in form of virtual machines, cannot be used in Mininet. Hence, though the scalability and performance of Mininet is excellent, and its actively used in the NetLab for research projects and master's courses, it is not specifically considered in the evaluation of this article, due to the limitations for the use cases and virtual network topologies discussed in the previous sections.

IV. Rationale for Selecting VIRL and GNS3

Over the last years, Mininet [25], eNSP [27], GNS3 [28], EVE-NG [29] (formerly UNetLab) and VIRL [30] have been

deployed and evaluated as a solution for realistic emulation of networking environments in the NetLab. By the time of publishing our initial research [1][20], VIRL was the most promising option, and was already used in several courses. It was compared to other emulation software alternatives based on criteria that are directly derived from administrative and educational requirements presented in Section III-B. In the meantime, new major versions of GNS3 (v.2.1) have been released, which implement features like QoS properties that we missed so far. Figure 4 depicts an updated revision of the initial comparison table, which is simplified compared to our initial paper and now includes the new version of GNS3. It clearly shows VIRL as the most promising approach when compared to GNS3 (v.2.0). However, with the new software release, GNS3 (v.2.1) even gets a slightly higher score.

Cisco Modeling Lab (CML), which is able to scale for a large number of nodes providing centralized management and automatic load balancing, is mainly depreciated due to the high licensing costs. CML is using the same technical basis as VIRL, and can be considered as the multiuser version of VIRL. In the case of EVE-NG, we were only able to test the free version, which lacks in some required functionality, such as centralized management, automatic load-balancing, but more importantly, the possibility to connect to a physical network, and to allow incoming VPN connections. These functions, however, might be available in the upcoming Premium or Learning Center release, which was announced to be released in 2018.

The poor performance of Mininet in our comparison is mainly due to the fact that no custom images are supported and connections to the console are very limited, hence the demands for functional realism and ubiquitous learning are not fulfilled for the use cases described in Section III-A. In addition, due to the implementation of Mininet as a command line tool and the missing GUI, the collaboration of students as well as the required training time in the laboratory is worse compared to the considered alternatives GNS3 and VIRL. However, even though the application of Mininet is not suitable for our virtual lab, we provide it locally installed on the lab computers, where it is especially used for thesis work in the field of SDN and NFV as well as for master's courses. By using LXC containers and OpenVSwitch as the basis for virtual hosts and networks, Mininet allows large network topologies to be launched on single-user computers even with limited resources. Due to its lightweight approach and its excellent suitability as an SDN environment, Mininet remains the best choice for research-oriented experiments in the master's field despite its lower rating compared to GNS3 and VIRL.

The advantages of VIRL and GNS3 are strongly related to the findings in [2] and [5]. The functional realism of VIRL is extended compared to the other alternatives, as it allows to use officially licensed network operating system images of Cisco components, like IOS or NX-OS, while at the same time, images from other vendors (i.e., Arista vEOS, HP VSR, Juniper vSRX/vMX, Cumulus VX) are supported as well. However, the most important advantage over the other alternatives is the underlying scale-out architecture based on OpenStack. This allows a central and scalable installation and enables the users to access the emulated network topologies location-independently (i.e., from within the NetLab or from at home using private PCs and laptops. Multiple VIRL hosts in the NetLab enable a scale-out of the testbed, which allows to emulate much bigger topologies than possible on a single PC in the laboratory or on a student notebook. In addition, the open architecture of OpenStack, as well as the open components used by VIRL (i.e., Ubuntu 14.04, LXC, linux-bridge, VXLAN) make it possible to extend the environment with in-house developed components to build specifically tailored testbeds for the use in research and education. Still, for the operation of VIRL in our environment, a few extensions were implemented, including customizations to the Arista vEOS and CumulusVX operating system images for our university's environment and for deployment in VIRL [31]. For example, to allow the operation of MLAG, it was necessary to modify the *base_mac* in order to prevent clashes of MAC addresses generated by vEOS with locally generated KVM MAC addresses [31].

In the latest version, GNS3 offers a lot of the functions required in our lab environment. Regarding network operating system choices it is as flexible as VIRL, with most vendors providing software images that are free to use in GNS3. However, especially Cisco requires users to buy a regular software maintenance contract to legally use their software images in this virtual environment. Beyond network operating system images, GNS3 supports the use of a wide range of server, desktop, and mobile operating systems by QEMU virtual machines. Although, cluster support with automatic scheduling of nodes in a topology (or project in GNS3) across multiple compute nodes is still not supported, the abandoning of Ciscos academic license for VIRL raised the importance to consider GNS3 as an alternative for our virtual network testbeds.

Improvements in terms of scalability (i.e., the size and amount of emulated testbeds), performance (i.e., the time to bootstrap a complete emulated topology) and usability (i.e., initial configuration) of VIRL and GNS3 will be discussed in the following sections.

V. Performance and Scalability Evaluation

Thanks to using virtual networks for the exercises shown in Figure 2, students can especially benefit from the advantages of the emulation as discussed in Section I. However, for complex topologies and a large number of students in the class, the benefits of the emulation places enormous demands on the virtual infrastructure it is running on. Even for smaller topologies it can take more than 5 minutes to start the emulations. Therefore, in the following sections, we describe a way to benchmark and optimize the waiting time until the emulations are ready to be used by the students in our laboratory.

A. Implementation of a VIRL Benchmarking Environment

The hardware we use for evaluating the performance and scalability of our VIRL environment was described in detail in [32]. For the evaluation presented in this article, initially VIRL 1.2.83 (October 2016) was used, since we kindly received an extended node count license in the Cisco dev/innovate research program for this version. Later, we repeated the evaluation with VIRL version 1.3.296 (August 2017) leading to slightly better, but similar results, as shown in the following sections of this article. All VIRL hosts are based on Ubuntu 14.04 VMs, each configured with 32 vCPUs and 64 GB of RAM. These VMs build a nested virtualization environment inside a

		Mininet		VIRL		CML		GNS 3 (v.2.0)		GNS 3 (v.2.1)		EVE-NG	
Criteria	**Weight 1-3**	**Description**	**#**	**Description**	**#**	**Description**	**#**	**Description**	**#**	**Description**	**#**	**Description**	**#**
Administrative Effort													
- License (Cost)	3	None (Open Source)	1	Cheap (199 € per 20 Cisco nodes per year)	0	Expensive (> $ 2.500 per 25 nodes per year)	-1	None (Open Source)	1	None (Open Source)	1	Free (Premium/Learning Center Edition?)	1
- Centralized Management	2	Isolated installation (Scripting support)	0	Central managed cluster with multiple compute nodes	1	Central managed cluster with multiple compute nodes	1	Multiple isolated servers	0	Multiple isolated servers	0	Perhaps with Premium/Learning Center Edition	0
- Compatibility and Accessibility	2	GNU/Linux (Virtual machine for Windows and MacOS)	0	VMMaestro-Client, and VMs for GNU/Linux, Windows, MacOS, vSphere	1	VMMaestro-Client, and VMs for GNU/Linux, Windows, MacOS, vSphere	1	GNU/Linux, Windows, MacOS	1	GNU/Linux, Windows, MacOS	1	Web-UI	1
- Custom Images	2	No support for real router and switch firmware images	-1	Extendable (node restriction only for included Cisco nodes), third-party images importable	1	Extendable (node restriction only for included Cisco nodes), third-party images importable	1	Images for routers and switches needed (Cisco Images not included)	1	Images for routers and switches needed (Cisco Images not included)	1	Images for routers and switches needed (Cisco Images not included)	1
- Load Balancing	2	Manually	0	Yes	1	Yes	1	Manually	0	Manually	0	Perhaps with Premium/Learning Center Edition	0
- Required Compute Resources	1	Low (LXC containers)	1	High (device dependent)	0	High (device dependent)	0	Medium (device dependent)	1	Medium (device dependent)	1	Medium (device dependent)	1
Educational Requirements													
- Connect to Physical Network	1	Yes	1	Yes	1	Yes	1	Yes (NAT with GUI, Bridged requires CLI configuration)	1	Yes (NAT with GUI, Bridged requires CLI configuration)	1	Only manually on CLI	0
- Multiple Users and Sessions	2	Manually	0	Yes	1	Yes	1	Single project per user, multiple user sessions from different clients possible	0	Single project per user, multiple user sessions from different clients possible	0	Yes, single project per user, multiple user sessions perhaps with Premium Edition	0
- Console Session	1	Limited (LXC console)	-1	Yes (singleuser)	0	Yes (singleuser)	0	Yes (multiuser support)	1	Yes (multiuser support)	1	Yes (multiuser support, but topology only shown in one browser)	1
- QoS properties for links	2	Yes	1	Yes (Only delay and packet loss)	0	Yes (Only delay and packet loss)	0	Only manually on CLI	0	Yes	1	Only manually on CLI	0
- VPN Connection to Virtual Networks	2	Manual configuration	0	Yes	1	Yes	1	Yes, with VPN server in the virtual network	1	Yes, with VPN server in the virtual network	1	No	-1
Weighted Score			**4**		**13**		**10**		**12**		**14**		**7**

Figure 4. A comparison of network emulation software based on out requirements.

VMware vSphere 6.5 cluster, in which each of the four VMs is bound to a separate physical ESXi host by DRS constraints to limit fluctuations in available resources in the virtualization environment. Each underlying ESXi host is equipped with two 8-core Intel(R) Xeon(R) E5-2650v2 2.60 GHz CPUs, 256 GB RAM and uses two NetApp E2700 over a redundant 16 Gbit/s Fibre Channel connection as a storage back end. The nodes are connected via 1 Gbit/s Ethernet to a Cisco Catalyst 3850 switch and with two 10 Gbit/s links to an Arista 7150S-24 and Arista 7050S-52 leveraging MLAG. As discussed in [2], the *Timing Realism* regarding emulation with regard to virtualization particularly depends on the isolation level of the virtualization environment. When strict isolation is not guaranteed, concurrently running VMs can have a negative performance impact. Therefore, we defined a separate resource pool with static resource allocation for our VIRL environment. All benchmark scenarios were repeatedly performed at night in the semester break to ensure minimum load and interference of the workload on the ESXi cluster. By monitoring the overall performance and capacity of our VMware vSphere cluster, we were able to verify that VMs related to the VIRL benchmark were the only systems that produced considerable load in our VMware environment during the tests. The topology shown in Figure 2a was used to evaluate the performance and scalability of our environment. Due to its small size and node count, it can be scaled fine-grained up to the full capacity of our cluster. For the evaluation of the scalability of virtual testbeds and their application in higher education courses with a large number of participants, the following four metrics are of special interest:

- *Start Time*, the time until the start of all submitted topologies is processed by VIRL's REST API
- *Active Time*, the time until all VMs start to boot
- *Usable Time*, the time until booting has finished and the virtual console of all vEOS nodes is accessible
- *Console delay*, the latency of the virtual console

To measure these metrics and in order to minimize outliers, we developed a script to run our performance tests in a reliable and reproducible manner. Each test run first starts up the network topologies using VIRL's REST API and measures the time until the start is confirmed (*Start Time*) and all nodes become active (*Active Time*). In our terminology, *Active Time* means that the VM is deployed by the OpenStack Nova scheduler on a VIRL host system, all virtual networks and ports are up and accessible, the vEOS image is provided and its boot sequence starts. Next, we measure the time until the VM really becomes responsive by connecting to the virtual console (*Usable Time*) and finally measure the interaction delay of keyboard inputs (*Console delay*). For this purpose, a Python script was developed, which establishes WebSocket connections to the serial consoles of the nodes running on the VIRL hosts.

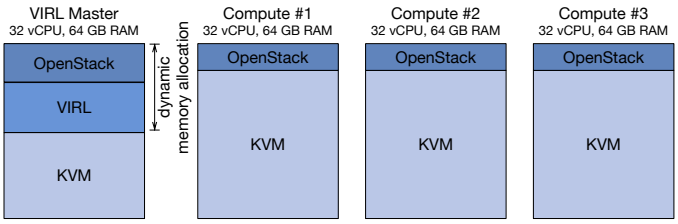


Figure 5. Schematic view of the environment's memory usage.

A schematic representation of the VIRL environment is depicted in Figure 5. At first, we performed all tests on only a single VIRL node, meaning that the node not only executes the

VMs needed for the emulated topology, but also acts as control node, which provides the OpenStack and VIRL environment. In Section V-B, we will share some negative observations we made, when including the control node in VM execution. Each individual test run for an increasing number of simultaneously emulated network topologies was executed ten times in a row. We started with only one topology and scaled in steps of five, until the VIRL host was working to capacity. Next, all tests were repeated on a 2-node and finally on a 4-node VIRL cluster to draw conclusions concerning scalability of the environment. The benchmark script and related toolset is available at [33].

B. Scalability Evaluation

The results from our previously explained test case are illustrated in Table I and Figure 6. When using a single VIRL host, the maximum number of simultaneously emulated network topologies is mainly limited by the resources (primarily the amount of main memory) available to the host. Each vEOS node uses 1 vCPU and 2 GB of RAM. Therefore, a test run with ten concurrently emulated network topologies requires 80 GB of main memory (10 * 4 vEOS nodes * 2 GB RAM) to be available, which is more than a single VIRL host in our test system can provide (see Figure 5). Hence, a stable execution was not possible with a single host, even with memory overprovisioning enabled (Figure 6a).

TABLE I. RELATIVE CHANGE IN *Start Time*, *Active Time* AND *Usable Time* DEPENDENT ON THE NUMBER OF TOPOLOGIES.

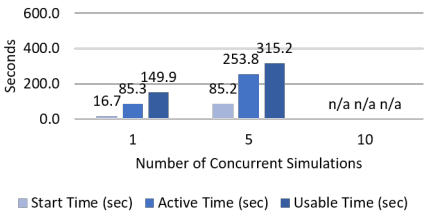
Number of Topologies			*Start Time*	*Active Time*	*Usable Time*
1	→	5	5.0617	2.5242	1.7440
5	→	10	2.0378	1.8517	1.6705
10	→	15	1.5105	1.4693	1.4686
15	→	20	1.3597	1.4548	1.4420
20	→	25	1.2334	1.2797	1.2793
25	→	30	1.2081	1.2304	1.2349

Looking at the effects of the number of parallel topologies in respect of performance, Figure 6b clearly depicts our expectation of a linear increase of the *Start Time*, while with an increasing number of topologies the *Active Time* and *Usable Time* ascends non-linear. The effect can best be observed when performing the test on a cluster with four or more nodes (Figure 6c). Up to a number of 10 simultaneously started topologies, the difference between *Active Time* and *Usable Time* gets smaller. This can be explained by the overhead the OpenStack-based resource scheduling and management introduces, which decreases with the number of simultaneously started topologies. For higher numbers of concurrent simulations, the system load generated from setting up virtual networks and interfaces causes an increased difference between *Active Time* and *Usable Time*. The curve for *Usable Time* has a comparatively steep slope as expected, as for an increasing number of virtual nodes, the time until all nodes are usable increases due to the limited resources.

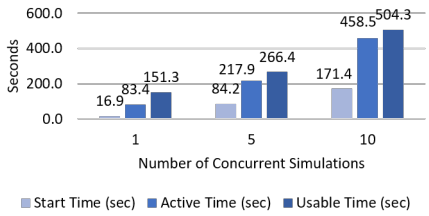
Alongside with the overhead introduced by the scheduling, the comparatively large difference between *Start Time* and *Active Time* results from the expensive process of creating all required virtual networks for connecting the devices. All links of the topology (Figure 2a) are redundant, which requires the creation of two VXLAN segments and its associated ports per pair of devices on the GNU/Linux bridge interface of the OpenStack nodes. The overhead of this was clearly visible by the CPU load produced by the Neutron process on the OpenStack controller node. The main limitation here is the fact that neutron-server and nova-conductor are single-threaded in VIRL's OpenStack Kilo setup, which limits the maximum performance of virtual network creation to a single CPU core. In most of our test cases, the CPU core neutron-server was executed on, was working to capacity. To overcome this limitation, we increased the number of neutron-server, nova-api and nova-conductor worker processes to ten. However, due to the fact that memory gets reserved on the VIRL master for these processes, the additional resource requirements resulted in a decrease of the maximum number of simultaneously started topologies to only 25 (Figure 5). Even more problematic was the observation that some of the vEOS nodes were not successfully started at the end of the test run, which is most likely explained by the dynamic memory allocation. When starting all topologies nearly at the same time, nova-scheduler is not able to determine the truly remaining amount of main memory and schedules too many VMs to the control node. At the same time, Figure 6d depicts that the *Usable Time* of the 20 topologies decreased by 16% and *Active Time* by 18%. While we assume room for improvement by carefully optimizing the OpenStack and KVM configuration, our recommendation is rather the use of a dedicated VIRL master node, which is currently not possible in VIRL, but can be manually achieved by deactivating nova-compute on the controller. The average console delay of the emulated vEOS nodes stays nearly constant with a growing number of simultaneously active topologies, as shown in Figure 6g. As a result, even if the time to start the concurrent simulations increases, a smooth use of the individually usable emulations is guaranteed despite the increased CPU load of the hypervisors. Limiting factors regarding our benchmark are more related to the amount of main memory and I/O performance, rather than the CPU load. What accounts for the latter is primarily the OpenStack and VIRL management processes, as well as the boot process of the vEOS instances, which utilizes the assigned vCPU to its maximum capacity for about 60 seconds in case of our test setup. As a performance improvement, Cisco recommends the use of a ramdisk for running Nova VMs, as well as an SSD for the VIRL hosts. We implemented both recommendations in our test environment to compare the impact on performance. First, a ramdisk was created, which is regularly only supported on the controller in VIRL, hence we needed to manually configure it also on the compute nodes. Figure 6e depicts the measurement results, clearly showing only a minor performance improvement, which is obviously attributable to the small amount of required ephemeral storage of only about 213 MB for a vEOS image. Second, we added two local solid state drives (Samsung 850 PRO) to each of the servers. Figure 6f shows no significant improvement of the performance, which is attributable to the previously used storage back end (NetApp E2700) already offering about 650 MB/s read/write performance. Due to the higher number of IOPS of the SSD, we assume that an improvement is likely to be observable when the I/O load of the VMs increases as a result of more complex topologies.

VI. EVALUATION OF GNS3 AS AN ALTERNATIVE

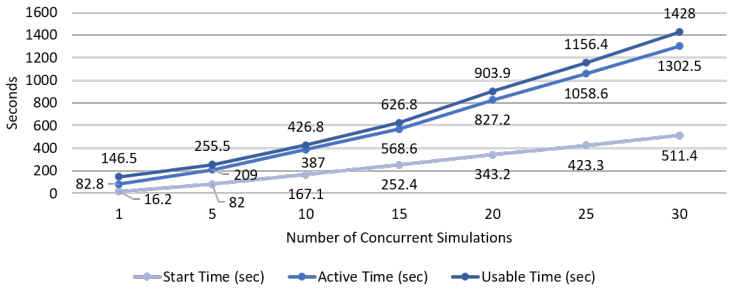
During winter term, we ported the example topologies for the advanced computer networks masters' course, as introduced in Section III-A, to a GNS3 testbed. We experimented with GNS3 v.2.1.3. The production environment described in



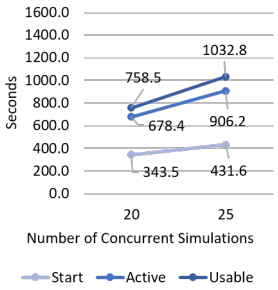
(a) Single VIRL Host.



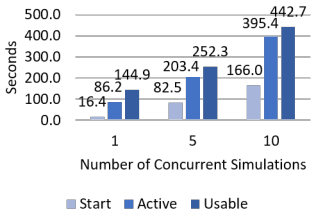
(b) 2-node Cluster.



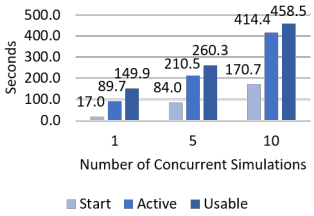
(c) 4-node Cluster.



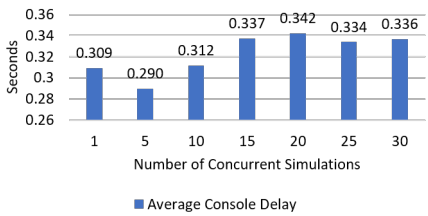
(d) 4-node, More Workers.



(e) 4-node, Ramdisk.

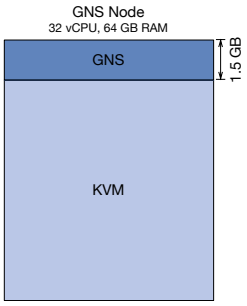


(f) 4-node, SSD.

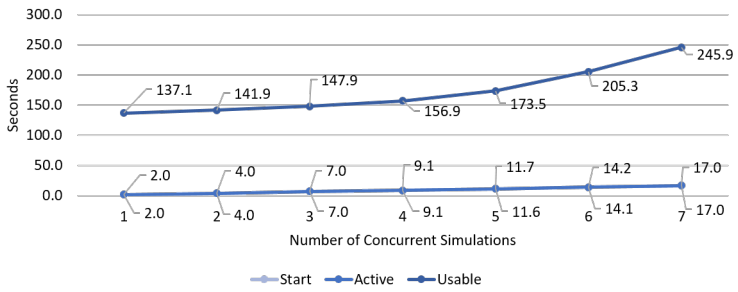


(g) 4-node, Avg Delay.

Figure 6. Results from measuring the time it takes to start multiple instances of the topology shown in Figure 2a.



(a) RAM usage.



(b) Measured time to start multiple instances of the topology from Figure 2a in GNS3.

Figure 7. Evaluation of GNS3.

this section, however, currently still uses v.2.0.3. The GNS3 testbed consists of five VMs, using an identical configuration as described for the VIRL testbed in Section V-A. Due to the missing automatic clustering and scheduling of started topologies in GNS3, we only used one GNS3 VM as a single node for the initial evaluation of the GNS3 performance described in this section. We started the same diamond-shaped topology with four Arista vEOS nodes, shown in Figure 2a, as for the VIRL evaluation discussed in the previous section. The entire process to start the topologies and the handling of so-called projects in GNS3 is different, compared to VIRL. Also, not only the code quality (e.g., regarding documentation and comments as well as the stability of the GNS3 GUI) still seems to be significantly lower on the GNS3 side. However, the progress of the project is impressing and not only the unattractive license model of VIRL for the use in academia strengthens the potential of GNS3 compared to VIRL. Since the code for GNS3 is available as open-source (under GPL-3.0 license) in a GitHub repository, extensions as well as fixes to the environment are possible. Although, GNS3 does not seem to focus classroom environments, as described in the requirements for a cluster environment in Section III-B. Instead, GNS3 seems to focus on a single user environment (e.g., for network consultants). However, features like the

simultaneous shared access to the console of emulated devices by multiple students, hold advantages and unique functions offered by GNS3 compared to VIRL when being used in classroom environments.

To get reproducible evaluation results for the performance of GNS3 in the same way and using the same requirements as described for VIRL in Section V-B, a Python-based benchmark solution was implemented to automatically run the experiment. Again, benchmarks were run ten times and GNS3 results in Figure 7 and Table II show the average time of these runs. Furthermore, benchmarks were run in different timeframes, to reduce possible side effects of the underlying virtualization infrastructure, which was not experiencing any other workload peaks during the tests. As for VIRL, the source of the benchmark process can be downloaded from our Git repository [34]. The Python script uses the GNS3 REST API [35] to create benchmark projects (as copies of the original topology) and measuring the time for this process (analog to the *Start Time* described in this article for VIRL). Afterwards, the benchmark projects are started and the time until all nodes in the topologies are active is measured (as for *Active Time* described in previous sections of this article). Since GNS3 is not dependent on another scheduling solution, as for example OpenStack in the VIRL setup, *Active Time* was reached only a fraction of a second after *Start Time*. This is due to the fact that GNS3 directly created the QEMU/KVM processes after topologies were started without the overhead of scheduling and messaging between the OpenStack components as described in Section V-B.

Also, as shown in Figure 7b *Usable Time* was reached quicker compared to VIRL on a single node. This is influenced by the smaller CPU and RAM footprint of GNS3, shown in Figure 7a compared to VIRL as depicted in Figure 5. This smaller resource footprint and corresponding shorter duration of benchmark runs, allowed for more fine-grained steps of scaling the number of concurrent simulations. While *Start* and *Active Time* increased linearly, as already described in this section, similar to VIRL, *Usable Time* increased rapidly, as soon as half (4 concurrent simulations * 4 vEOS nodes * 2 GB RAM = 32 GB RAM) of the memory was filled by the vEOS VMs. As expected, we were not able to run a significant amount of benchmarks with 8 concurrent simulations (filling up the entire 64 GB RAM), although Kernel Samepage Merging (KSM) was used to compress identical memory pages of the VMs. Since KSM's memory deduplication also takes resources and especially time to scan allocated pages, only a few runs were successful using 8 concurrent simulations. Their *Start Time* varied from 13 to 20 seconds, *Active Time* was between 14 and 20 seconds, *Usable Time* between 291 and 331 seconds, fitting into the trend of Figure 7b. The relative change in *Start Time*, *Active Time* and *Usable Time* for increasing numbers of concurrent simulations is listed in Table II. The last line of Table II can be indirectly compared to the same increase in concurrent simulations for VIRL from Table I, though the table for VIRL is based on a setup with 4 nodes while the GNS3 benchmarks were run on a single host. For this reason, the comparably large increase of *Start Time* and *Active Time* for GNS3 is plausible. However, the lightweight implementation and smaller resource footprint of GNS3 results in a lower increase of *Usable Time* even when GNS3 running on a single node is compared to a 4-node VIRL cluster.

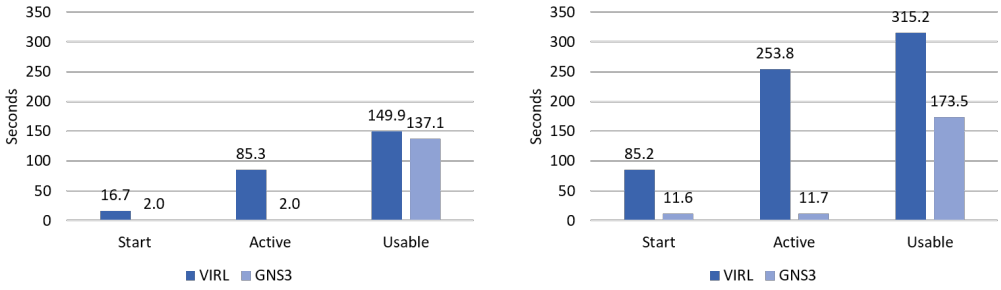
TABLE II. RELATIVE CHANGE IN *Start Time*, *Active Time* AND *Usable Time* DEPENDENT ON THE NUMBER OF TOPOLOGIES USING GNS3.

Number of Topologies			*Start Time*	*Active Time*	*Usable Time*
1	→	2	2.0000	2.0000	1.0350
2	→	3	1.7500	1.7500	1.0423
3	→	4	1.3000	1.3000	1.0609
4	→	5	1.2747	1.2857	1.1058
5	→	6	1.2155	1.2137	1.1833
6	→	7	1.2057	1.1972	1.1978
1	→	5	7.3000	7.3000	1.2416

Figure 8 compares the results of one (Figure 8a) as well as five (Figure 8b) concurrently started topologies between GNS3 and VIRL. To allow a comparison, the results shown in the figures are taken from a single VIRL node, which was shown in Figure 6a in Section V-B, so that GNS3 and VIRL both used only one node with an identical setup and the same environment as described in Section V-A. It can be seen from the figures that not only the *Start Time* and *Active Time* is smaller for GNS3, due to the fact that GNS3 starts all QEMU/KVM processes for the vEOS nodes directly, but also and more importantly, the *Usable Time* decreases significantly especially for large numbers of concurrent simulations. For example, as shown in Figure 8b, the same topology that was started five times in VIRL was usable after 315.2 seconds, while being available for the students after 173.5 seconds (55% of the time taken in the VIRL setup) when using GNS3 running in the same virtualization environment with the same dedicated resources.

Given the smaller resource requirements of GNS3, we tried to identity the limiting factors for the results we measured in further experiments, where CPU and RAM resources were systematically reduced. Figure 9 shows CPU and RAM usage of the GNS3 VM. Additionally, in the middle of the figure, the CPU load on the underlying host is depicted to check that there were no significant additional workloads in the host while running the benchmarks. In the timeframe from 07.01.18 10:40 to about 13:45, a benchmark using the standard value of 32 vCPUs and 64 GB, as for the VIRL tests, was used. It can be seen in the figure that the VM experienced high CPU ready time and less active time. Investigating this effect further, led to the finding that this impact was caused by a hyperthreading overhead combined with the NUMA architecture of the underlying host (2 CPU sockets with 8 physical cores each). This means that during the start of a large number of concurrent simulations, hyperthreaded cores competed against each other and also led to the effect that they needed to access RAM on another NUMA node, further increasing ready time and access delay. We were able to mitigate this effect, by decreasing the virtual CPUs available to the VM from 32 to 16, so that the cores used by the VM could be placed mostly on the same NUMA node. The result can be seen in Figure 9. Two benchmark runs, one from ca. 13:45 to 19:45 and one from ca. 19:45 to 1:45 are graphed in the figure.

The results measured for these runs are depicted in Figure 10. As visible, the results are similar to the runs with 32 vCPUs, shown in Figure 7b. However, as shown in the graph for the CPU load on the VM in Figure 9, the CPU of the underlying host is nearly used up to its capacity for these runs, with the ready time of the CPU being reduced to a value close to zero. Hence, the overall load of the benchmark on the underlying virtualization infrastructure was less, though



(a) 1 Simulation. (b) 5 Concurrent Simulations.

Figure 8. Comparison of starting the topology shown in Figure 2a on a single VIRL and GNS3 node.

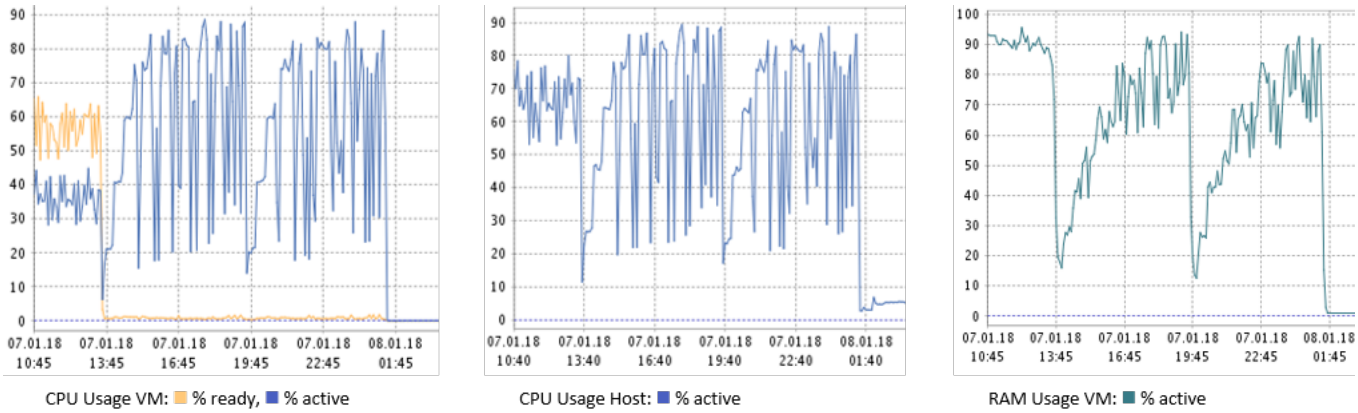


Figure 9. CPU and RAM usage of the GNS3 VM and the underlying host during benchmarks shown in Figure 7b and 10.

the results for our measurements did not change significantly. From the two full benchmark runs shown in Figure 9, the CPU and RAM resources consumed by the tests are also visible. The benchmark run started around 13:45 first shows steps that result from first 10 times repeatedly starting a single topology, then 10 times 2, followed by 10 times 3, and afterwards 10 times 4 concurrent simulations. As stated before, a single topology uses 4 x 1 = 4 vCPUs and 4 x 2 GB = 8 GB of RAM. Hence, 4 concurrent simulations used 16 vCPUs and 32 GB of RAM. Therefore, vCPUs were the most significant limiting factor for our benchmarks, even after the issue described for concurrent hyperthreaded vCPUs across NUMA nodes was resolved. As shown runs with more than 4 concurrent simulations, starting at around 15:15 (1.5 hours after the benchmark was started), account for ca. 75% of the entire 6 hours that the benchmark took to complete. In the right part of Figure 9, the RAM consumed by the VM is depicted. Along with the CPU usage, also the RAM consumption increases with the amount of concurrently run simulations, as expected. It can be seen in the picture that the RAM of the VM is slowly starting to saturate, after the maximum capacity of concurrent simulation is reached.

Booting a single vEOS 4.17.5M instance in our environment already takes ca. 135 seconds until the prompt is usable. Therefore, the results measured for the start of a single instance of our topology with four vEOS nodes is still close to the minimum time that a single vEOS switch needs to boot. As discussed, the amount of vCPUs is the main limitation in our scenario. To further investigate the influence of the vCPUs on the start of the concurrent topologies, we reduced the number of vCPUs available to the GNS3 VM further down from 16 to 8. Results after this degradation are shown in Figure 11. As expected, the further reduction of vCPUs moves the point where *Usable Time* starts to rise, increasing the slope of the curve, from 4 to 2 concurrent simulations, as 2 x 4 x 1 = 8 vCPUs of the vEOS instances fill up all virtual CPU cores available for the VM.

To crosscheck a possible influence of the RAM limits of the VM, additionally, the experiment was again repeated 10 times in different timeframes using only 32 instead of 64 GB RAM for the VM while changing the vCPUs back to 16. Results of this experiment are shown in Figure 12. By comparing Figure 10 and Figure 12, it can be seen, as expected, that neither *Start*, *Active* nor *Usable Time* were significantly impacted by limiting the RAM to 32 GB, as long as not more than 3 concurrent simulations were started. Starting 4 concurrent simulations already experienced a slightly higher *Usable Time*. Benchmarks with more than 4 concurrent simulations were, as expected, not possible, due to the fact that 4 concurrent simulations already consume 4 x 4 x 2 GB = 32 GB RAM.

Evaluating the comparison running the same benchmark process for GNS3 as for VIRL, GNS3 not only has the advantage of being open-source and hence highly customizable as well as offering an attractive licensing model compared to VIRL and Cisco Modeling Lab (CML) for the use in higher education, it also offers better scalability due to its significantly smaller resource footprint. However, the lack of a cluster solution as well as classroom features (e.g., user management) for GNS3 is still leaving some advantages on VIRL's side. Nonetheless, manually distributing nodes of a single topology

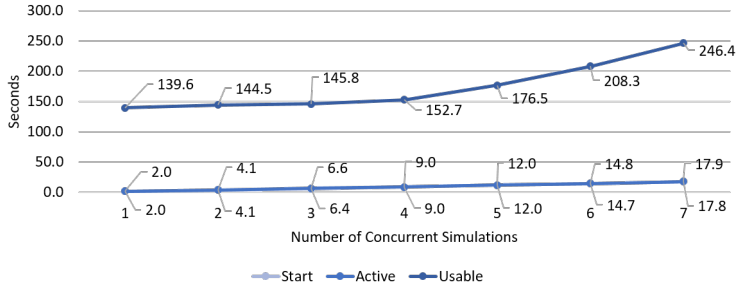


Figure 10. Results with 16 instead of 32 VCPUs for the GNS3 VM.

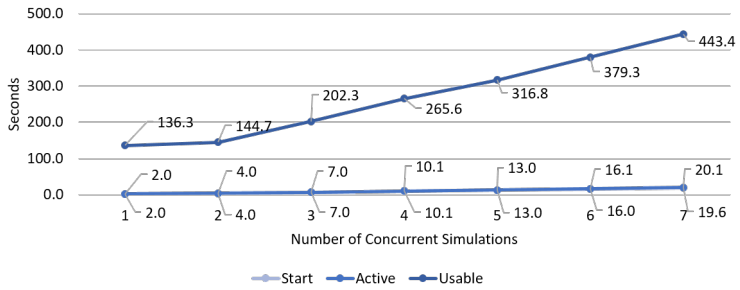


Figure 11. Results with 8 instead of 32 VCPUs for the GNS3 VM.

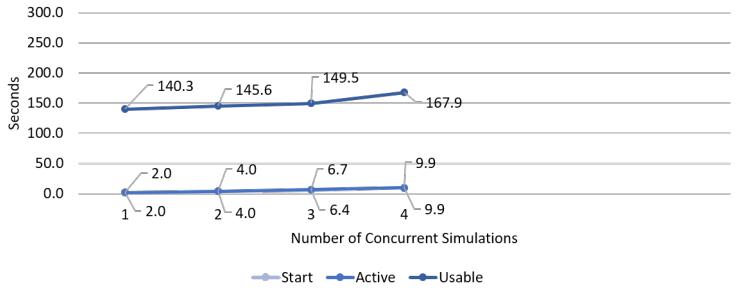


Figure 12. Results with 16 instead of 32 VCPUs and 32 instead of 64 GB RAM for the GNS3 VM.

run in GNS3 across multiple servers is possible. While not being able to dynamically spread the load evenly across all nodes, this already allows for fine grained scaling of GNS3 environments as back ends for higher education networking courses. Also, clustering and collaboration support seems to be on the road map for future GNS3 releases. Comparing GNS3 and VIRL regarding their features, as for example given in [20], recent versions of GNS3 already introduced previously missing features for the use in education. For example, since version 2.1.0 controlling the delay, jitter and packet loss of links, e.g., for WAN emulation, which was previously already available in VIRL, is now also possible in GNS3 projects. Version 2.1.1 introduced the display of the server individual nodes of the topology are placed on. Further tracking of the RAM and CPU usage of the started topology in GNS3 is planned for the upcoming GNS3 web interface [36]. Overall, despite some glitches and the difference in code and documentation quality, GNS3 looks like a promising alternative for future virtual testbeds for computer networks in our NetLab.

VII. Conclusion and Future Work

Cisco VIRL provides a platform, which is capable of creating realistic and scalable virtual network testbeds for education and research projects. In comparison with alternatives, such as GNS3 or EVE-NG, a clear advantage is that it offers to use original Cisco operating system images in conformance with license requirements. Beyond that, an even more important feature is the foundation of VIRL, which is based on the open-source project OpenStack. This enabled us to modify and extend the environment as shown in this article, and to build a well-scaling multi-node VIRL cluster, which supports a sufficiently large number of simultaneous emulations for application in education. Further, by allowing the utilization of standard network management applications (i.e., ping, traceroute) and operating systems (i.e., Ubuntu VMs) inside the emulated network testbeds, as well as the connection to real physical networks, a great flexibility and functional realism can be achieved in comparison with other simulation approaches. The increased start time introduced by the emulation, especially for complex topologies, can be compensated by using a VIRL scheduler that we developed to specifically address the requirements of our NetLab [37]. It offers to pre-load topologies based on a schedule, e.g., in advance of an upcoming seminar, hence minimizing delays for the students. Additionally, we are actively developing a management application for VIRL and GNS3 labs. When finished, it will provide a self-service system enabling students

to subscribe to courses and to start working on topologies and reserving virtual lab time. To increase performance even further, the most promising approaches are given by increasing the number of cluster nodes and optimizing to the OpenStack resource and network management.

By the time of writing, new major versions of GNS3 were released. Still, GNS3 does not provide a way to use Cisco operating system images in conformance with the license requirements. However, a large number of third-party virtual network equipment images is available. Sadly, a load balancing of started projects across multiple hosts still is not possible. In this article, we present a comparison of the scalability and performance of GNS3 as an alternative to VIRL. Though the overall quality of GNS3 seems to be lower compared to VIRL, performance and features in recent versions have surpassed the performance as well as educational suitability of VIRL for the use cases described in this article. Since Cisco seems to have changed the strategy for VIRL and tries to move universities to the expensive Cisco Modeling Lab, the new version of GNS3 could become an interesting alternate candidate for our environment. We are currently looking into the possibility to develop appropriate extensions to GNS3 or EVE-NG to fix the current lack in cluster-based load balancing and centralized management and real-time collaboration options on running simulations compared to VIRL.

ACKNOWLEDGMENT

We thank Cisco for providing us with a research license within the context of the Cisco dev/innovate research program for our Virtual Internet Routing Lab (VIRL) cluster. Also, we thank VMware for providing academic licenses to run our private cloud environment, being used as the virtualization backend for our GNS3 and VIRL installation, as well as Arista for providing firmware images and support for the switches used in our network.

REFERENCES

[1] S. Reißmann, S. Rieger, and C. Pape, "Using VIRL to improve the scale-out of large virtual network testbeds in higher education," in SIMUL 2017, The Ninth International Conference on Advances in System Simulation, 2017, pp. 29–34.

[2] N. Handigol, B. Heller, V. Jeyakumar, B. Lantz, and N. McKeown, "Reproducible network experiments using container-based emulation," in Proceedings of the 8th international conference on Emerging networking experiments and technologies. ACM, 2012, pp. 253–264.

[3] M. Pizzonia and M. Rimondini, "Netkit: network emulation for education," Software: Practice and Experience, vol. 46, no. 2, Feb. 2016, pp. 133–165.

[4] S. V. Tagliacane, P. W. C. Prasad, G. Zajko, A. Elchouemi, and A. K. Singh, "Network simulations and future technologies in teaching networking courses: Development of a laboratory model with Cisco Virtual Internet Routing Lab (Virl)," in 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2016, pp. 644–649.

[5] J. Obstfeld et al., "VIRL: The Virtual Internet Routing Lab," SIGCOMM Comput. Commun. Rev., vol. 44, no. 4, Aug. 2014, pp. 577–578. [Online]. Available: http://doi.acm.org/10.1145/2740070.2631463

[6] L. Yan and N. McKeown, "Learning networking by reproducing research results," ACM SIGCOMM Computer Communication Review, vol. 47, no. 2, 2017, pp. 19–26.

[7] V. Bajpai et al., "Challenges with reproducibility," in Proceedings of the Reproducibility Workshop, ser. Reproducibility '17. New York, NY, USA: ACM, 2017, pp. 1–4. [Online]. Available: http://doi.acm.org/10.1145/3097766.3097767

[8] M. Flittner et al., "Taming the complexity of artifact reproducibility," in Proceedings of the Reproducibility Workshop, ser. Reproducibility '17. New York, NY, USA: ACM, 2017, pp. 14–16. [Online]. Available: http://doi.acm.org/10.1145/3097766.3097770

[9] W. Makasiranondh, S. P. Maj, and D. Veal, "Pedagogical evaluation of simulation tools usage in network technology education," World Transactions on Engineering and Technology Education, vol. 8, no. 3, 2010, pp. 321–326.

[10] P. Gil et al., "Computer networks virtualization with GNS3: Evaluating a solution to optimize resources and achieve a distance learning," in 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, Oct 2014, pp. 1–4.

[11] B. Momeni and M. Kharrazi, "Partov: a network simulation and emulation tool," Journal of Simulation, vol. 10, no. 4, 2016, pp. 237–250.

[12] M. A. Qadeer, P. Varshney, and N. H. Khan, "Design and Simulation of Interconnected Autonomous Systems," in 2009 International Conference on Computer Engineering and Technology (ICCET). IEEE, 2009, pp. 270–275.

[13] S. Hemminger, "Network emulation with NetEm," in Linux conf au, 2005, pp. 18–23.

[14] F. Baumgartner, T. Braun, E. Kurt, and A. Weyland, "Virtual Routers: A Tool for Networking Research and Education," SIGCOMM Comput. Commun. Rev., vol. 33, no. 3, Jul. 2003, pp. 127–135.

[15] A. Kayssi and A. El-Haj-Mahmoud, "EmuNET: A Real-time Network Emulator," in Proceedings of the 2004 ACM Symposium on Applied Computing, ser. SAC '04. New York, NY, USA: ACM, 2004.

[16] L. Xu, D. Huang, and W.-T. Tsai, "Cloud-based virtual laboratory for network security education," IEEE Transactions on Education, vol. 57, no. 3, Aug. 2014, pp. 145–150.

[17] T. Wolf, "Assessing student learning in a virtual laboratory environment," IEEE Transactions on Education, vol. 53, no. 2, May 2010, pp. 216–222.

[18] L. Yan and N. McKeown, "Learning networking by reproducing research results," ACM SIGCOMM Computer Communication Review, vol. 47, no. 2, 2017.

[19] HS-Fulda NetLab VIRL topologies. URL: https://gogs.informatik.hs-fulda.de/srieger/git-virl-hs-fulda, 2018-05-20. (2018)

[20] C. Seifert, S. Rieger, and C. Pape, "Realization possibilities for virtual networking labs in higher education courses," in 13 th Annual International Conference on Computer Science and Education in Computer Science 2017 (CSECS 2017), 2017.

[21] C. Pape and C. Seifert, "Adaption and improvement of an industry-developed IP Telephony curriculum," in 7th Annual International Conference on Computer Science and Education in Computer Science, Sofia/Dobrinishte, Jul. 2011, pp. 199–210.

[22] ns-3. URL: https://www.nsnam.org, 2018-05-20. (2018)

[23] OMNeT++ Discrete Event Simulator. URL: https://omnetpp.org, 2018-05-20. (2018)

[24] Packet Tracer - A free network simulation and visualization tool for the IoT era. URL: https://www.netacad.com/courses/packet-tracer, 2018-05-20. (2018)

[25] Mininet - An Instant Virtual Network on your Laptop (or other PC). URL: http://mininet.org, 2018-05-20. (2018)

[26] B. Lantz, B. Heller, and N. McKeown, "A network in a laptop: rapid prototyping for software-defined networks," in Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks. ACM, 2010, p. 19.

[27] eNSP - Enterprise Network Simulator. URL: http://support.huawei.com/enterprise/en/network-management/ensp-pid-9017384, 2018-05-20. (2018)

[28] GNS3 - The software that empowers network professionals. URL: https://www.gns3.com, 2018-05-20. (2018)

[29] Emulated Virtual Environment Next Generation (EVE-NG) / Unified Networking Lab (UNL). URL: http://www.unetlab.com, 2018-05-20. (2018)

[30] VIRL - Virtual Internet Routing Lab. URL: http://virl.cisco.com, 2018-05-20. (2018)

[31] S. Rieger. Arista vEOS image on VIRL. URL: https://learningnetwork.cisco.com/thread/99040, 2018-05-20. (2018)

[32] K. Spindler et al., "AEQUO: Enhancing the energy efficiency in private clouds using compute and network power management functions," International Journal on Advances in Internet Technology Volume 8, Number 1 & 2, 2015, vol. 8, no. 1 & 2, 2015, pp. 13–28.

[33] HS-Fulda NetLab VIRL utilities. URL: https://gogs.informatik.hs-fulda.de/srieger/virl-utils-hs-fulda, 2018-05-20. (2018)

[34] S. Rieger. GNS3 benchmark script. URL: https://gogs.informatik.hs-fulda.de/srieger/gns3bench, 2018-05-20. (2018)

[35] GNS3 2.0.1 API Documentation - Sample session using curl. URL: http://pythonhosted.org/gns3-server/curl.html, 2018-05-20. (2018)

[36] GNS3. Enhancement: Show where a device is installed. URL: https://github.com/GNS3/gns3-gui/issues/2279#issuecomment-345429116, 2018-05-20. (2018)

[37] P. Bug. virl-scheduler. URL: https://gogs.informatik.hs-fulda.de/pbug/virl-scheduler, 2018-05-20. (2018)

www.iariajournals.org

International Journal On Advances in Intelligent Systems
issn: 1942-2679

International Journal On Advances in Internet Technology
issn: 1942-2652

International Journal On Advances in Life Sciences
issn: 1942-2660

International Journal On Advances in Networks and Services
issn: 1942-2644

International Journal On Advances in Security
issn: 1942-2636

International Journal On Advances in Software
issn: 1942-2628

International Journal On Advances in Systems and Measurements
issn: 1942-261x

International Journal On Advances in Telecommunications
issn: 1942-2601