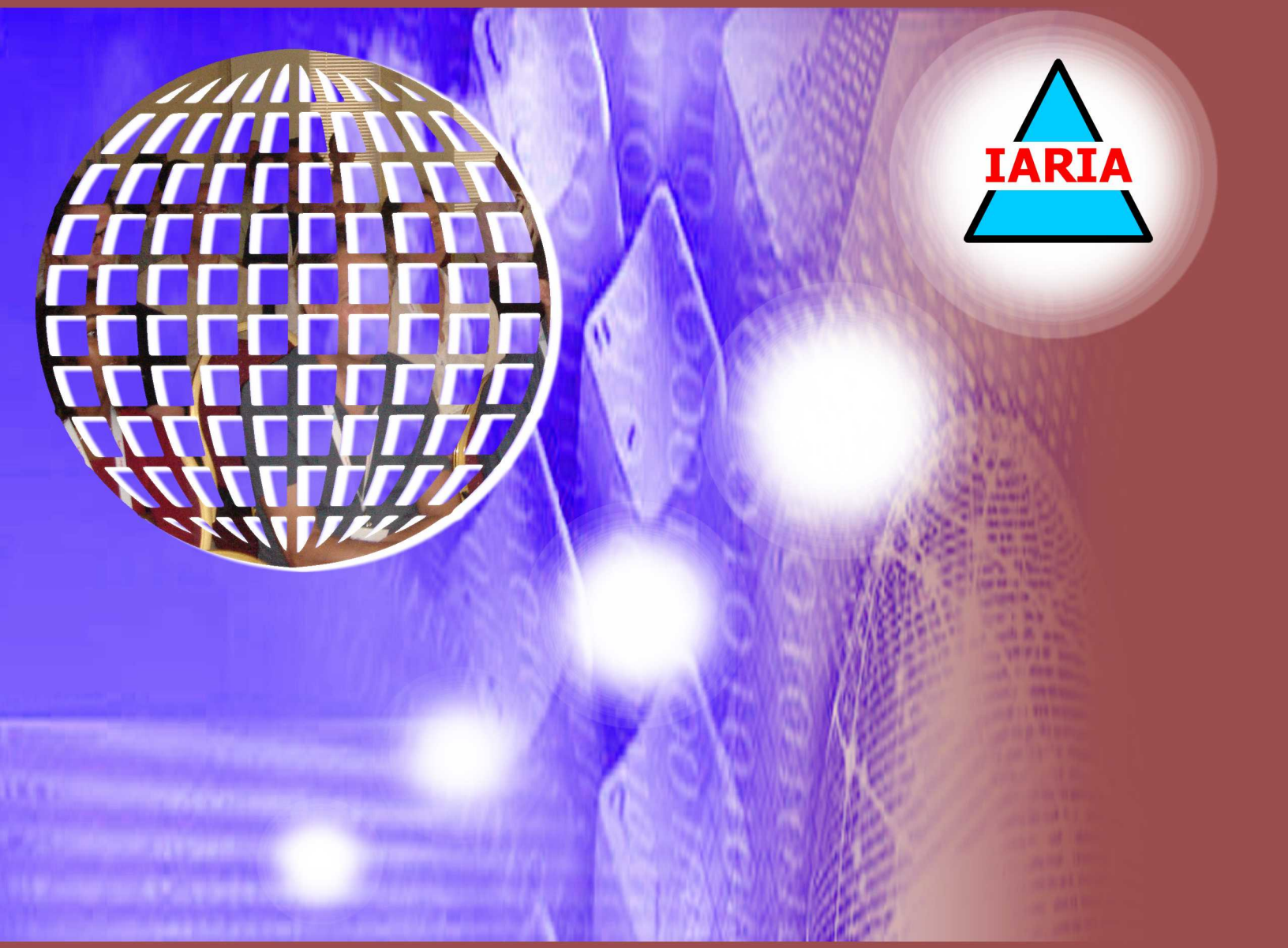
International Journal on Advances in Networks and Services



2023 vol. 16 nr. 3&4

The *International Journal on Advances in Networks and Services* is published by IARIA.
ISSN: 1942-2644
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Responsibility for the contents rests upon the authors and not upon IARIA, nor on IARIA volunteers, staff, or contractors.

IARIA is the owner of the publication and of editorial aspects. IARIA reserves the right to update the content for quality improvements.

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy or print, providing the reference is mentioned and that the resulting material is made available at no cost.

Reference should mention:

International Journal on Advances in Networks and Services, issn 1942-2644
vol. 16, no. 3 & 4, year 2023, http://www.iariajournals.org/networks_and_services/

The copyright for each included paper belongs to the authors. Republishing of same material, by authors or persons or organizations, is not allowed. Reprint rights can be granted by IARIA or by the authors, and must include proper reference.

Reference to an article in the journal is as follows:

<Author list>, "<Article title>"
International Journal on Advances in Networks and Services, issn 1942-2644
vol. 16, no. 3 & 4, year 2023, <start page>:<end page> , http://www.iariajournals.org/networks_and_services/

IARIA journals are made available for free, proving the appropriate references are made when their content is used.

Sponsored by IARIA
www.iaria.org

Copyright © 2023 IARIA

International Journal on Advances in Networks and Services
Volume 16, Number 3 & 4, 2023

Editor-in-Chief

Tibor Gyires, Illinois State University, USA

Editorial Advisory Board

Mario Freire, University of Beira Interior, Portugal
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Rainer Falk, Siemens AG - Corporate Technology, Germany
Cristian Anghel, University Politehnica of Bucharest, Romania
Rui L. Aguiar, Universidade de Aveiro, Portugal
Jemal Abawajy, Deakin University, Australia
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France

Editorial Board

Ryma Abassi, Higher Institute of Communication Studies of Tunis (Iset'Com) / Digital Security Unit, Tunisia
Majid Bayani Abbasy, Universidad Nacional de Costa Rica, Costa Rica
Jemal Abawajy, Deakin University, Australia
Javier M. Aguiar Pérez, Universidad de Valladolid, Spain
Rui L. Aguiar, Universidade de Aveiro, Portugal
Ali H. Al-Bayati, De Montfort Uni. (DMU), UK
Giuseppe Amato, Consiglio Nazionale delle Ricerche, Istituto di Scienza e Tecnologie dell'Informazione (CNR-ISTI), Italy
Mario Anzures-García, Benemérita Universidad Autónoma de Puebla, México
Pedro Andrés Aranda Gutiérrez, Telefónica I+D - Madrid, Spain
Cristian Anghel, University Politehnica of Bucharest, Romania
Miguel Ardid, Universitat Politècnica de València, Spain
Valentina Baljak, National Institute of Informatics & University of Tokyo, Japan
Alvaro Barradas, University of Algarve, Portugal
Mostafa Bassiouni, University of Central Florida, USA
Michael Bauer, The University of Western Ontario, Canada
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Zdenek Becvar, Czech Technical University in Prague, Czech Republic
Francisco J. Bellido Outeiriño, University of Cordoba, Spain
Djamel Benferhat, University Of South Brittany, France
Jalel Ben-Othman, Université de Paris 13, France
Mathilde Benveniste, En-aerion, USA
Luis Bernardo, Universidade Nova of Lisboa, Portugal
Alex Bikfalvi, Universidad Carlos III de Madrid, Spain
Thomas Michael Bohnert, Zurich University of Applied Sciences, Switzerland
Eugen Borgoci, University "Politehnica"of Bucharest (UPB), Romania
Fernando Boronat Seguí, Universidad Politecnica de Valencia, Spain

Christos Bouras, University of Patras, Greece
Mahmoud Brahimi, University of Msila, Algeria
Marco Bruti, Telecom Italia Sparkle S.p.A., Italy
Dumitru Burdescu, University of Craiova, Romania
Diletta Romana Cacciagrano, University of Camerino, Italy
Maria-Dolores Cano, Universidad Politécnica de Cartagena, Spain
Juan-Vicente Capella-Hernández, Universitat Politècnica de València, Spain
Eduardo Cerqueira, Federal University of Para, Brazil
Bruno Chatras, Orange Labs, France
Marc Cheboldaeff, Deloitte Consulting GmbH, Germany
Kong Cheng, Vencore Labs, USA
Dickson Chiu, Dickson Computer Systems, Hong Kong
Andrzej Chydzinski, Silesian University of Technology, Poland
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Noelia Correia, University of the Algarve, Portugal
Noël Crespi, Institut Telecom, Telecom SudParis, France
Paulo da Fonseca Pinto, Universidade Nova de Lisboa, Portugal
Orhan Dagdeviren, International Computer Institute/Ege University, Turkey
Philip Davies, Bournemouth and Poole College / Bournemouth University, UK
Carlton Davis, École Polytechnique de Montréal, Canada
Claudio de Castro Monteiro, Federal Institute of Education, Science and Technology of Tocantins, Brazil
João Henrique de Souza Pereira, University of São Paulo, Brazil
Javier Del Ser, Tecnalia Research & Innovation, Spain
Behnam Dezfouli, Universiti Teknologi Malaysia (UTM), Malaysia
Daniela Dragomirescu, LAAS-CNRS, University of Toulouse, France
Jean-Michel Dricot, Université Libre de Bruxelles, Belgium
Wan Du, Nanyang Technological University (NTU), Singapore
Matthias Ehmann, Universität Bayreuth, Germany
Wael M El-Medany, University Of Bahrain, Bahrain
Imad H. Elhajj, American University of Beirut, Lebanon
Gledson Elias, Federal University of Paraíba, Brazil
Rainer Falk, Siemens AG - Corporate Technology, Germany
Károly Farkas, Budapest University of Technology and Economics, Hungary
Huei-Wen Ferng, National Taiwan University of Science and Technology - Taipei, Taiwan
Gianluigi Ferrari, University of Parma, Italy
Mário F. S. Ferreira, University of Aveiro, Portugal
Bruno Filipe Marques, Polytechnic Institute of Viseu, Portugal
Ulrich Flegel, HFT Stuttgart, Germany
Juan J. Flores, Universidad Michoacana, Mexico
Ingo Friese, Deutsche Telekom AG - Berlin, Germany
Sebastian Fudickar, University of Potsdam, Germany
Stefania Galizia, Innova S.p.A., Italy
Ivan Ganchev, University of Limerick, Ireland / University of Plovdiv “Paisii Hilendarski”, Bulgaria
Miguel Garcia, Universitat Politecnica de Valencia, Spain
Emiliano Garcia-Palacios, Queens University Belfast, UK
Marc Gilg, University of Haute-Alsace, France

Debasis Giri, Haldia Institute of Technology, India
Markus Goldstein, Kyushu University, Japan
Luis Gomes, Universidade Nova Lisboa, Portugal
Anahita Gouya, Solution Architect, France
Mohamed Graiet, Institut Supérieur d'Informatique et de Mathématique de Monastir, Tunisie
Christos Grecos, University of West of Scotland, UK
Vic Grout, Glyndwr University, UK
Yi Gu, Middle Tennessee State University, USA
Angela Guercio, Kent State University, USA
Xiang Gui, Massey University, New Zealand
Mina S. Guirguis, Texas State University - San Marcos, USA
Tibor Gyires, School of Information Technology, Illinois State University, USA
Keijo Haataja, University of Eastern Finland, Finland
Gerhard Hancke, Royal Holloway / University of London, UK
R. Hariprakash, Arulmigu Meenakshi Amman College of Engineering, Chennai, India
Eva Hladká, CESNET & Masaryk University, Czech Republic
Hans-Joachim Hof, Munich University of Applied Sciences, Germany
Razib Iqbal, Amdocs, Canada
Abhaya Induruwa, Canterbury Christ Church University, UK
Muhammad Ismail, University of Waterloo, Canada
Vasanth Iyer, Florida International University, Miami, USA
Imad Jawhar, United Arab Emirates University, UAE
Aravind Kailas, University of North Carolina at Charlotte, USA
Mohamed Abd rabou Ahmed Kalil, Ilmenau University of Technology, Germany
Kyoung-Don Kang, State University of New York at Binghamton, USA
Sarfraz Khokhar, Cisco Systems Inc., USA
Vitaly Klyuev, University of Aizu, Japan
Jarkko Kneckt, Nokia Research Center, Finland
Dan Komosny, Brno University of Technology, Czech Republic
Ilker Korkmaz, Izmir University of Economics, Turkey
Tomas Koutny, University of West Bohemia, Czech Republic
Evangelos Kranakis, Carleton University - Ottawa, Canada
Lars Krueger, T-Systems International GmbH, Germany
Kae Hsiang Kwong, MIMOS Berhad, Malaysia
KP Lam, University of Keele, UK
Birger Lantow, University of Rostock, Germany
Hadi Larijani, Glasgow Caledonian Univ., UK
Annett Laube-Rosenpflanzer, Bern University of Applied Sciences, Switzerland
Gyu Myoung Lee, Institut Telecom, Telecom SudParis, France
Shiguo Lian, Orange Labs Beijing, China
Chiu-Kuo Liang, Chung Hua University, Hsinchu, Taiwan
Wei-Ming Lin, University of Texas at San Antonio, USA
David Lizcano, Universidad a Distancia de Madrid, Spain
Chengnian Long, Shanghai Jiao Tong University, China
Jonathan Loo, Middlesex University, UK
Pascal Lorenz, University of Haute Alsace, France

Albert A. Lysko, Council for Scientific and Industrial Research (CSIR), South Africa
Pavel Mach, Czech Technical University in Prague, Czech Republic
Elsa María Macías López, University of Las Palmas de Gran Canaria, Spain
Damien Magoni, University of Bordeaux, France
Ahmed Mahdy, Texas A&M University-Corpus Christi, USA
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Gianfranco Manes, University of Florence, Italy
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Moshe Timothy Masonta, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa
Hamid Menouar, QU Wireless Innovations Center - Doha, Qatar
Guowang Miao, KTH, The Royal Institute of Technology, Sweden
Mohssen Mohammed, University of Cape Town, South Africa
Miklos Molnar, University Montpellier 2, France
Lorenzo Mossucca, Istituto Superiore Mario Boella, Italy
Jogesh K. Muppala, The Hong Kong University of Science and Technology, Hong Kong
Katsuhiro Naito, Mie University, Japan
Deok Hee Nam, Wilberforce University, USA
Sarmistha Neogy, Jadavpur University- Kolkata, India
Rui Neto Marinheiro, Instituto Universitário de Lisboa (ISCTE-IUL), Instituto de Telecomunicações, Portugal
David Newell, Bournemouth University - Bournemouth, UK
Ngoc Tu Nguyen, Missouri University of Science and Technology - Rolla, USA
Armando Nolasco Pinto, Universidade de Aveiro / Instituto de Telecomunicações, Portugal
Jason R.C. Nurse, University of Oxford, UK
Kazuya Odagiri, Sugiyama Jyogakuen University, Japan
Máirtín O'Droma, University of Limerick, Ireland
Jose Oscar Fajardo, University of the Basque Country, Spain
Constantin Paleologu, University Politehnica of Bucharest, Romania
Eleni Patouni, National & Kapodistrian University of Athens, Greece
Harry Perros, NC State University, USA
Miodrag Potkonjak, University of California - Los Angeles, USA
Yusnita Rahayu, Universiti Malaysia Pahang (UMP), Malaysia
Yenumula B. Reddy, Grambling State University, USA
Oliviero Riganelli, University of Milano Bicocca, Italy
Antonio Ruiz Martinez, University of Murcia, Spain
George S. Oreku, TIRDO / North West University, Tanzania/ South Africa
Sattar B. Sadkhan, Chairman of IEEE IRAQ Section, Iraq
Husnain Saeed, National University of Sciences & Technology (NUST), Pakistan
Addisson Salazar, Universidad Politecnica de Valencia, Spain
Sébastien Salva, University of Auvergne, France
Ioakeim Samaras, Aristotle University of Thessaloniki, Greece
Luz A. Sánchez-Gálvez, Benemérita Universidad Autónoma de Puebla, México
Teerapat Sanguankotchakorn, Asian Institute of Technology, Thailand
José Santa, University Centre of Defence at the Spanish Air Force Academy, Spain
Rajarshi Sanyal, Belgacom International Carrier Services, Belgium
Mohamad Sayed Hassan, Orange Labs, France
Thomas C. Schmidt, HAW Hamburg, Germany

Véronique Sebastien, University of Reunion Island, France
Jean-Pierre Seifert, Technische Universität Berlin & Telekom Innovation Laboratories, Germany
Dimitrios Serpanos, Univ. of Patras and ISI/RC ATHENA, Greece
Roman Y. Shtykh, Rakuten, Inc., Japan
Salman Ijaz Institute of Systems and Robotics, University of Algarve, Portugal
Adão Silva, University of Aveiro / Institute of Telecommunications, Portugal
Florian Skopik, AIT Austrian Institute of Technology, Austria
Karel Slavicek, Masaryk University, Czech Republic
Vahid Solouk, Urmia University of Technology, Iran
Peter Soreanu, ORT Braude College, Israel
Pedro Sousa, University of Minho, Portugal
Cristian Stanciu, University Politehnica of Bucharest, Romania
Vladimir Stantchev, SRH University Berlin, Germany
Radu Stoleru, Texas A&M University - College Station, USA
Lars Strand, Nofas, Norway
Stefan Strauβ, Austrian Academy of Sciences, Austria
Álvaro Suárez Sarmiento, University of Las Palmas de Gran Canaria, Spain
Masashi Sugano, School of Knowledge and Information Systems, Osaka Prefecture University, Japan
Young-Joo Suh, POSTECH (Pohang University of Science and Technology), Korea
Junzhao Sun, University of Oulu, Finland
David R. Surma, Indiana University South Bend, USA
Yongning Tang, School of Information Technology, Illinois State University, USA
Yoshiaki Taniguchi, Kindai University, Japan
Anel Tanovic, BH Telecom d.d. Sarajevo, Bosnia and Herzegovina
Rui Teng, Advanced Telecommunications Research Institute International, Japan
Olivier Terzo, Istituto Superiore Mario Boella - Torino, Italy
Tzu-Chieh Tsai, National Chengchi University, Taiwan
Samyr Vale, Federal University of Maranhão - UFMA, Brazil
Dario Vieira, EFREI, France
Lukas Vojtech, Czech Technical University in Prague, Czech Republic
Michael von Riegen, University of Hamburg, Germany
You-Chiun Wang, National Sun Yat-Sen University, Taiwan
Gary R. Weckman, Ohio University, USA
Chih-Yu Wen, National Chung Hsing University, Taichung, Taiwan
Michelle Wetterwald, HeNetBot, France
Feng Xia, Dalian University of Technology, China
Kaiping Xue, USTC - Hefei, China
Mark Yampolskiy, Vanderbilt University, USA
Dongfang Yang, National Research Council, Canada
Qimin Yang, Harvey Mudd College, USA
Beytullah Yildiz, TOBB Economics and Technology University, Turkey
Anastasiya Yurchyshyna, University of Geneva, Switzerland
Sergey Y. Yurish, IFSA, Spain
Jelena Zdravkovic, Stockholm University, Sweden
Yuanyuan Zeng, Wuhan University, China
Weiliang Zhao, Macquarie University, Australia

Wenbing Zhao, Cleveland State University, USA
Zibin Zheng, The Chinese University of Hong Kong, China
Yongxin Zhu, Shanghai Jiao Tong University, China
Zuqing Zhu, University of Science and Technology of China, China
Martin Zimmermann, University of Applied Sciences Offenburg, Germany

International Journal on Advances in Networks and Services
Volume 16, Numbers 3 & 4, 2023

CONTENTS

6G Architecture: New Use Cases, New Needs and New Challenges

Francine Cássia de Oliveira
Integration and Testing Department
Eldorado Research Institute
Campinas, Brazil
francine.oliveira@eldorado.org.br

Gustavo Iervolino de Morais
Electronic Devices Department
Eldorado Research Institute
Campinas, Brazil
gustavo.iervolino@eldorado.org.br

João Victor Menino e Silva
Integration and Testing Department
Eldorado Research Institute
Campinas, Brazil
joao.menino@eldorado.org.br

Ramon Magalhães Nogueira
Integration and Testing Department
Eldorado Research Institute
Campinas, Brazil
ramon.nogueira@eldorado.org.br

***Abstract*—Reliable data communication is essential for connections that are increasingly intelligent and automated. The fifth generation of mobile networks offers major improvements over the previous generation, 4G, but is still not capable of offering a ubiquitous connection to meet the demands of the new applications and needs that are emerging. On the basis of this, the cellular network standard will require a new communications network and this will come in the new generation of mobile networks, called 6G. This new standard has been considered a key enabler for the smart information society of 2030. The 6G networks are expected to deliver superior performance over 5G and satisfy new emerging services and applications that integrate space, air, ground, and underwater networks to provide ubiquitous and unlimited wireless connectivity. There is a huge number of use cases that pose varying requirements, which include extreme mobility, extreme low latency, ultra-high data rates, high energy efficiency, enhanced security, as well as high reliability. From this perspective, it is necessary to consider some of the key features that can be fundamental for the construction of the 6G network architecture. In this article, we will list some main use cases like Digital Twins, Global Ubiquitous Connectivity, Remote Communications, and others, that will need the main 6G functionalities to work correctly and meet the expectations of the main 6G requirements so that it is possible to identify relevant points in the construction of a robust and flexible network architecture.**

***Index Terms*—6G network(s); use cases; architecture; application; THz communication; energy efficiency; fog computing**

I. Introduction

A. Background

The growing demand for greater data traffic capacity, the staggered growth in the number of users, technological advances, and new services drive the mobile communication systems and thus the development of the 5G system of International Mobile Telecommunications-2020 (IMT-2020) [1] was initiated. In International Telecommunication Union-Radiocommunication Sector (ITU-R), the Working Party 5D (WP5D), is responsible for the radio system that includes the IMT-2000, IMT-Advanced, IMT-2020, and IMT-2030. For IMT-2020, the WP5D created a process to be followed from the beginning of the study of trends until the end of the work on standards. The capabilities of IMT-2020 are identified such that IMT-2020 is more flexible, reliable, and secure than previous IMT and provides diverse services. IMT-2020 can be considered from multiple perspectives, including the users, manufacturers, application developers, network operators, service and content providers. The WP5D commenced its work on the recommendation "IMT Vision for 2030 and beyond" in March 2021. The IMT Vision for 2030 and beyond is being developed with the aim to drive the industries and administrations to encourage further development of IMT by defining the objectives of the future of the IMT, including the role IMT could play to meet the needs of future societies. Some of the objectives of the vision towards IMT for 2030 and beyond are: focus on the continued need for increased coverage, capacity and extremely high user data rates, focus on the continued need for lower latency and both high and low speed of the mobile terminals, full support to the development of an Ubiquitous Intelligent Mobile Society, focus on delivering on digital inclusion and connection with the rural and remote communities, among others [2]. To meet the diverse requirements of the upcoming decade, a robust, scalable, and efficient network is thus necessary to be the key enabler for achieving this objective; it will connect everything, provide full dimensional wireless coverage, and integrate all functions, including sensing, communication, computing, caching, control, positioning, radar, navigation, and imaging, to support full-vertical applications.

B. Motivation

In fifth-generation networks, one of the main pillars in their development was the interconnection of everything, but applications involved with the Internet of Vehicles and Industrial Internet, for example, may be far from being met with such technology. Some questions are still unanswered, such

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

as: What will be the problems of 5G for application in the industrial area? What will a green industry look like? Perhaps, these and other questions challenge the capacity of 5G and, probably, only 6G can solve.

Many white papers have addressed some aspects of the 6G network. For example, new 6G applications and requirements are discussed in [3], 6G enabling technologies are mentioned in [4] and 6G enablers to drive Industry 5.0 are discussed in [5]. However, it is still too early to say exactly what the 6G network architecture will look like as the network and corresponding technologies are still under development.

Therefore, the main objective of this study is to analyze the main use cases that will require a network such as the sixth generation, and the indicators related to them that may directly influence the construction of the 6G network architecture.

C. Paper Organization

The subsequent sections of this paper are organized as follows: A synopsis of related work is given in Section II, along with an analysis of relevant studies and literature. Potential use cases in the context of 6G are examined in Section III, providing insight into a range of applications. In Section IV, the difficulties of Remote Communications are examined, along with their complexities and obstacles. In Section V, target indicators associated with the identified use cases are analyzed with respect to their possible impact on 6G network architecture. Section VI, which concludes the study, provides a thorough summary of the structure and contributions of the work by synthesizing the findings and suggesting possible avenues for further research.

II. Related Work

Several studies involving 6G architecture have been carried out to meet the demands of a fully connected, intelligent, and digital world. In [6], Huda Mahmood et al. propose an architecture composed of seven functions that have functionalities for essential enabling technologies. The objective of this architecture is to allow the optimization of such functionalities through dedicated network components.

According to Purbita Mitra et al., [7], 6G networks aim for ubiquitous intelligence and high-speed wireless connectivity in air, space and sea. This will require a super fast service with data speeds close to around 1000 Mbps. Marco Giordani et al. [8] have an analysis that suggests that meeting these high demands will require new communication technologies, network architecture, and deployment models. Finally, Bariah et al. gave a comprehensive overview of 6G in [9], identifying seven disruptive technologies, associated requirements, challenges, and open research questions.

So far, a considerable number of papers have explored possible applications and solutions for the architecture of 6G networks. Therefore, the present related work analyzes factors that may influence the evolution of the 5G network to 6G or the construction of a new network architecture with the purpose of fulfilling the requirements specified by the IMT-2030 for the next decade of technological evolution.

III. Use Cases

The 5G, through the Massive Machine-Type Communications (mMTC) and Ultra Reliable Low Latency Communications (URLLC) use cases, has resulted in a significant increase in the number of connected devices. New applications in vertical industries emerge every day, bringing a significant impact on people's daily lives. Internet of Things (IoT) solutions will continue to emerge and there are several use cases whose strict requirements 5G can not meet, such as Augmented Reality (AR), Virtual Reality (VR), haptic internet, and telemedicine, among others. The sixth generation mobile network, 6G, should support and improve the connectivity and operation of such applications.

Therefore, many use cases will require requirements that can only be met with sixth-generation technology. Some of these cases are listed below.

A. Digital Twins

With the increasing number of connected "things", in 6G, a self-sustainable system should be proposed, which can be intelligent and operate with minimal human intervention. One technology, that presents itself as a strong candidate for such a requirement, and has received great attention, is the Digital Twins. It is a virtual representation of the elements and dynamics of a physical system [10]. In an ideal scenario, a Digital Twin will be indistinguishable from the physical asset, both in terms of appearance and behavior, with the added benefit of making predictions [11]. Figure 1 illustrates this representation of the virtual elements in relation to a physical system. Advances in other technologies make Digital Twins a powerful solution and contribute to its advancement.

For example, recent advances in Machine Learning enable Digital Twins to analyze data and make decisions to be applied to the physical entity. This data can come from a network of sensors, from historical data or even from other Digital Twins (through a twin-to-twin interface). In other words, automation and intelligence will be created in the cyber world and delivered to the physical world through 6G wireless networks [12].

Another enabler for the Digital Twins has been the significant advances in cloud solutions. The transformation of a physical system into a Digital Twin is mainly based on the concept of decoupling. To enable Digital Twin for 6G with decoupling, Software Defined Networking (SDN) and Network Function Virtualization (NFV) could be promising candidates [9], which are heavily dependent on cloud solutions.

Digital Twins consist of three parts - the physical part, the digital part and the connection between the two for two-way communication. For this two-way communication, there is a unanimous opinion in the research community that the sixth generation (6G) mobile network will play a significant role [13], given that the Digital Twins technology requires a fast and reliable communication network.

In addition to the contribution of 6G, many benefits can be achieved through the technology of Digital Twins. Since the Digital Twin "mimics" the real physical environment and

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

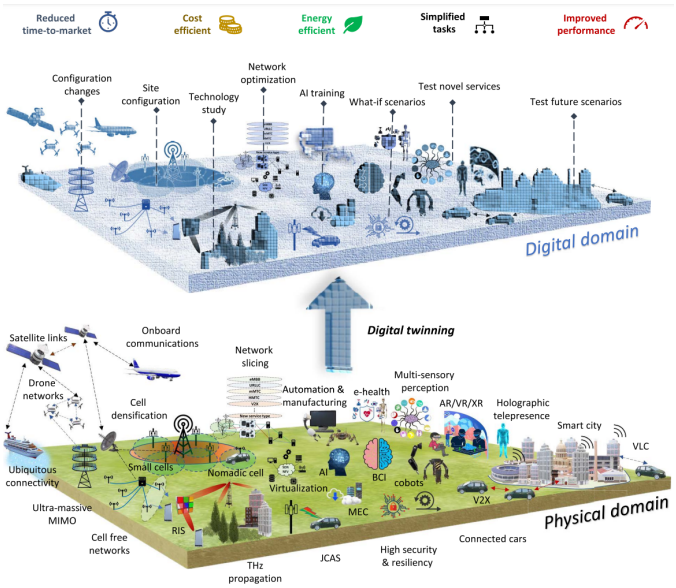


Fig. 1. Representation of the virtual elements in relation to a physical system [12]

can learn and make decisions through artificial intelligence algorithms, there are several aspects in the research and development of 6G communication systems that could benefit from the application of this technology.

There are several network domains, such as Radio Access Network (RAN), Network Edge, Radio Resource Management (RRM), Edge Computing, Network Slicing, etc., that can significantly improve their performance using Digital Twins technology [14].

B. Human-Centric Immersive Communications

Through the ages, human beings have evolved their cognitive capacity through the use of all the senses in relationships with other individuals and with nature, therefore, the search for a better communication experience has been constant since the invention of the first communication systems. In smartphones, every year, the screen resolution is improved to the limit of human perception, which is quite interesting, but it has the limiting factor of having to enter data through only touches on the screen. Therefore, in order to provide an immersive experience, in which the human being can use senses in a more accurate way, new technologies such as AR and VR, as well as holographic communications have been emerging in recent times.

Through them, it will be possible to offer new forms of interaction between human beings and their devices and, consequently, new forms of human-to-human interaction. Communication that until then was carried out strictly through a smartphone, mostly with touches on the screen, can evolve so that it is possible to enter data through gestures and even through nerve impulses generated by the brain. Obtaining data will also be improved and synesthesia becomes even more present through, for example, the combination of sounds and three-dimensional elements that can be inserted and merged with the user's perception of the real world through glasses, ocular lenses, and devices in-ear audio. For such technologies to be offered as good user experiences through the 6G network, ultra-high data rates are required, in the order of Tbits/s, which is currently impossible to achieve with the 5G network. In addition to the very high rate, another fundamental requirement for such teleoperations involving the senses and human perception is very low latency. This parameter is necessary in order to avoid dizziness and fatigue when obtaining tactile and visual feedback in real time [12].

C. Industry 5.0

Industry 5.0 is the enhancement of Industry 4.0 and brings new goals with resilient, sustainable, and human-centric approaches in a variety of emerging applications, for example, factories of the future and digital society. It is a quest to leverage human intelligence and creativity in connection with intelligent, efficient systems, the use of cognitive collaborative robots to achieve zero waste, zero defects, and mass customization-based manufacturing solutions.

The enabling technologies of Industry 5.0 are multiple systems resulting from the continuous convergence of technologies and paradigms that unite physical spaces and cyberspaces. Successfully working the symbiotic relationship between multiple complex systems and supporting technological frameworks together can only enable the true multidimensional potential of Industry 5.0 functions [4]. These Industry 5.0 technology enablers are Human-Machine Interaction, Real-time Virtual Simulation and Digital Twin, Artificial Intelligence-native Smart Systems, Data Infrastructure, Sharing and Analytics, and Bio-inspired Technologies, among others.

The relationship between 6G and Industry 5.0 is expected to meet with the intelligent information standard that provides high energy efficiency, very low latency, high reliability, plus capacity of traffic.

D. Global Ubiquitous Connectivity

As it is known, legacy mobile communication systems aimed to provide connectivity with a focus on dense urban areas, resulting in many sparsely populated regions lacking adequate connectivity and basic Information and Communication Technology (ICT) services. However, especially in countries with vast territorial expanses, a significant portion of the population resides in remote areas. This is particularly intensified in countries that are major agricultural and agribusiness producers, where a large part of the population chooses to develop their production in more suitable locations, generally distant from major urban centers, as is the case in Brazil, for example.

Besides the extensive terrestrial territories, it is essential to remember that over 70% of the planet's surface is covered by water, making the development of communication systems in these areas equally crucial. However, achieving total global coverage with adequate capacity, high-quality service (QoS), and affordable cost is still far from reality. Nevertheless, it

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

is of maximum importance for mobile systems to develop in these areas to avoid significant digital divides among people worldwide. This development is necessary not only for enhancing security with improved geolocation and emergency response methods, but also for enhancing consumer goods production through the use of IoT devices, for instance.

In summary, providing means of global connectivity through resilient infrastructure is essential for enhancing security, production, and overall quality of life. These purposes align with the Sustainable Development Goals defined by the United Nations (UN) [15], which aim to provide ubiquitous Internet access for anyone or any device anywhere. However, it is technically impossible for terrestrial networks to cover remote areas such as oceans, deserts, and high mountainous regions, and furthermore, providing communication services to sparsely populated areas is not attractive to major players in the industry.

Attaining ubiquitous global coverage necessitates overcoming challenges spanning political, market, and chiefly technical issues. Technically, the development of 5G in Releases 15, 16, and 17 initiates the addressing of concerns regarding architecture interoperability with various technologies. However, the call for global coverage surpasses the current definitions of these Releases and can only be tackled with the establishment of 6G Networks. Simultaneously, the analysis of 3GPP Release 18 is focused on enhancing coverage for handheld devices in the sub-6 GHz band with added antenna gain losses in the device. Combining the attenuation due to long propagation distances with the reduction in antenna gain within the device yields a diminished signal-to-noise ratio (SNR) in both Downlink (DL) and Uplink (UL) directions. Increasing transmission power, whether in the device or satellite, stands as a solution. However, the utilization of a legacy waveform design from 3GPP NTN yields an inefficient solution with increased complexity due to peak-to-average power ratio (PAPR) and out-of-band (OOB) power leakage [15].

Hence, currently, the best alternative for achieving total coverage of the planet Earth is through the use of satellites. Geostationary satellites (GEO), despite being expensive to deploy and having a capacity of only a few gigabits per second (Gbps) per satellite [16], may not be suitable for common uses, but they can be harnessed for critical applications, such as in the maritime and aviation sectors. Medium Earth Orbit (MEO) and Low Earth Orbit (LEO) satellites can be a viable option for everyday uses, given the possibility of building low-cost satellite constellations and providing highly profitable global communication services [17].

In other words, satellites will not only collaborate with existing terrestrial networks but will also be integral to the architecture of the 6G network, with common management besides other network elements. Satellite constellations can be formed, and furthermore, other elements can contribute to ubiquitous coverage, such as Unmanned Aerial Vehicle (UAVs), drones, balloons, and aircraft, each serving different roles within the system, such as gateways, relays, or even radio base stations [18].

E. Pervasive Intelligence

The dissemination of mobile devices and the emergence of new intelligent devices such as cars, drones, IoT devices, and robots, for example, lead to significant growth in over-the-air intelligent services. As the use of these devices continues to advance, the need for artificial intelligence technologies to collaborate in providing fundamental functions like Simultaneous Localization and Mapping (SLAM), facial, speech, and image recognition, natural language processing, and motion detection, among many others, becomes increasingly essential. However, AI services require high computational capabilities that may not always be available to the devices intending to use them. Therefore, 6G presents itself as an excellent alternative to offer generalized AI services, falling under the AI-as-a-Service model [19].

A particularly promising scenario for the use of pervasive intelligence can be identified in the utilization of humanoid robots as cooperative partners. These robots aim to physically resemble human beings to perform risky functions, arduous tasks, and other daily life activities. An example where pervasive intelligence can be observed is in humanoid robots like Atlas, developed by Boston Dynamics [20].

Such robots and other devices can utilize computational resources offered through the 6G network to spare their own resources and optimize computational load, thereby increasing energy efficiency. By receiving processed instructions from a central core provided via 6G, they can save their own resources, prolong battery life, and preserve computational capacity for more critical functions.

In addition to handling intensive computational tasks, pervasive intelligence also enables the execution of real-time AI operations. This is particularly advantageous as it overcomes the latency limitations associated with cloud computing, enabling quick decision-making and immediate responses to real-time conditions.

IV. Remote Communications

Connectivity in remote areas has been a challenge for many years. However, the COVID-19 pandemic has highlighted the importance of connectivity more than ever before. The pandemic accelerated the transition to remote work and learning, but unfortunately, it left many people out of this digital age. according to the state of broadband 2022 report [21], in 2019, 54% of the world's population was using the internet, with this number growing to 66% in 2022. However, there are still many people around the world who do not have access to the Internet, especially in rural and remote areas.

The numbers from the mobile economy 2023 report [22] show that there are still 3.5 billion disconnected people, and thus excluded from the digital age. The majority of these individuals live in developing countries, especially in Sub-Saharan Africa and India.

The exclusion of these individuals from the digital age has a significant impact on their lives, as connectivity plays a crucial role in a wide range of activities, spanning sectors such as education, health, business, and public administration.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

One field that has shown remarkable advancements is the e-health sector, as documented in successful cases in the future of virtual health and care reports [23]. A notable case in India, where there was a significant 300% increase in the number of teleconsultations between March and May 2020. Consequently, these initiatives contribute to reducing disparities and enhancing the quality of life in rural and remote regions.

It is believed that 6G will be developed taking these needs into account, in order to enable the population in rural and remote regions to be integrated into the digital age, facilitating their participation in the opportunities offered by this new era. However, for this to become a reality, it is necessary to delve into the analysis and research of a set of solutions, in order to address the challenges that contribute to limited internet connectivity in these remote areas.

A. Lack of Energy Sources

Many remote or rural regions face challenges in accessing the electrical grid, making the provision of energy for telecommunications networks is a difficult task. In the context of 6G, potential solutions may involve the use of generators or, preferably, the adoption of renewable energy sources such as solar or wind. However, it is important to note that this approach may lead to an increase in deployment costs and make the network more susceptible to failures. For this reason, other lines of research include the development of energy-efficient equipment to mitigate these challenges.

B. Spectrum Availability

One of the biggest barriers to network deployment in rural areas is spectrum licensing, as participating in spectrum auctions is difficult for small ISPs [24]. Additionally, spectrum frequency regulation in remote areas adheres to national standards, although the possibility of flexibility could be considered, given that many frequency bands remain underutilized (unallocated) in these isolated locations [25]. Therefore, it might be necessary to have two sets of regulations: one for urban areas and another for remote and rural areas.

C. Maintenance and Operation, Access Difficulty, and Qualified Workforce

Understanding the inherent complexities of maintaining and operating telecommunications networks in remote regions presents a significant challenge, both in practical and financial terms. This complexity arises from the difficulty of accessing such areas due to the often-present topographical adversities in remote locations, which lack proper road infrastructure, making transportation to these points challenging.

Furthermore, this scenario is exacerbated by the scarcity of qualified workforce, as the regions in question often face financial constraints inherent to their situation, frequently being situated in developing nations. Possible solutions can stem from advances in Self-Organizing Networks (SONs), which can be employed to automate as many resources as possible, encompassing various network components and engineering phases [26].

By automating network management tasks, SONs can contribute to improving network performance and user experience in remote and rural areas, while also reducing the need for manual intervention and maintenance. Moreover, the digitization of these remote areas could enhance the level of skills in that region, bringing new opportunities to the population and possibly encouraging the government to invest in education in that area.

D. Critical Infrastructure

Another challenging issue concerns the lack of infrastructure in these areas. The deployment of cables, fibers, and even communication towers faces obstacles due to terrain characteristics. Consider, for example, the complexity involved in deploying fiber optic networks in the Amazon region, aiming to provide connectivity to an isolated indigenous community.

Various alternatives have been discussed as potential solutions, among which the utilization of existing infrastructure stands out, such as those used in TV and radio transmissions [24]. Another strategy involves the adoption of Integrated Access And Backhaul (IAB) technology to replace the use of fiber optics, which proves to be a cost-effective solution when compared to fiber optics. IAB offers a more flexible implementation approach. Based on a wired connection to the core network, the IAB donor can provide communication access to mobile users and act as a wireless backhaul for IAB nodes. These IAB nodes are capable of providing network service access to the mobile user as well as backhaul traffic [27], as illustrated in Figure 2.

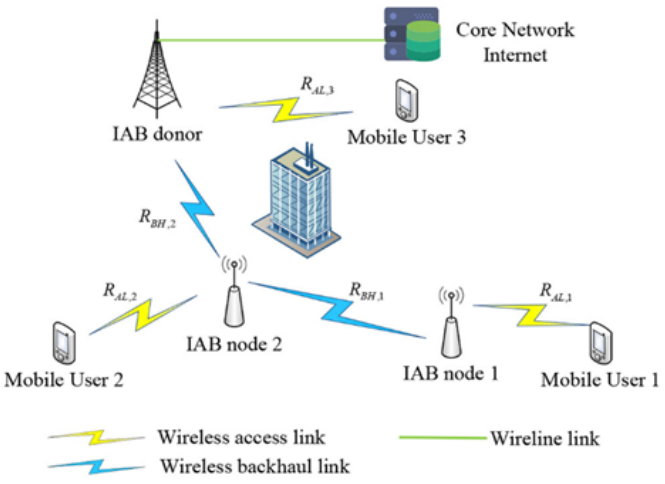


Fig. 2. IAB structure [27]

E. Low Return on Investment/High Cost/Low Income of the Target Population

Each of the mentioned points above requires a considerable allocation of resources for their resolution, resulting in a relatively reduced return on investment (RoI), as the target

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

population often consists of individuals with less favorable socioeconomic conditions. As a result, national operators often lack interest in investing in such areas, therefore, new business plans and mechanisms that facilitate the entry of local and micro-operators will be necessary. To reduce costs, integrating various solutions is feasible.

The discussion about the use of Non-Terrestrial Networks (NTNs) has gained prominence, as well as Device-to-Device (D2D) connections, which ensure coverage in the network's peripheral regions. Non-Terrestrial Networks (NTNs) encompass Unmanned Aerial Vehicles (UAVs), High Altitude Platform Stations (HAPSs), and satellites (such as a Low Earth Orbit (LEO) constellation). These solutions could provide connectivity both in the front (fronthaul) over vast geographical areas and in the back (backhaul), potentially replacing the use of fiber optic networks.

V. Target Indicators For 6G

Each new use case presents highly specialized and demanding requirements that the 5G network lacks the capacity to meet and work with. Figure 3 illustrates the comparison between the requirements of 5G and 6G, where the vertices of the inner polygon represent the Key Performance Indicators (KPIs) of 5G, while the vertices of the outer polygon represent the KPIs of 6G. In this image, we can see a noticeable improvement in all the majorly considered KPIs.

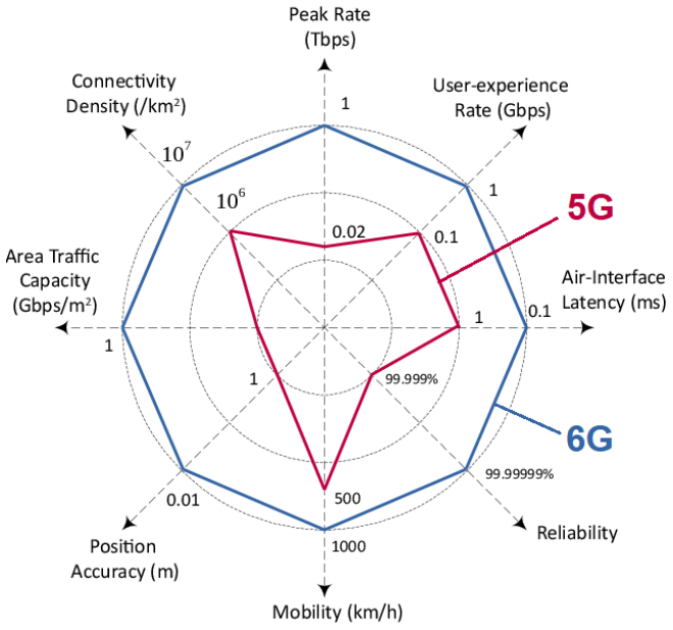


Fig. 3. Comparison between 5G and 6G requirements [21]

Nonetheless, different use cases demand distinct KPIs. For example, Ultra-Reliable Low Latency Communications (URLLC) applications require the lowest possible latency, with the other KPIs not being as important. In contrast, Massive Machine Type Communications (mMTC) applications demand high connection density and energy efficiency, with the other KPIs being less relevant. Figure 4 shows the significance of each KPI for different use scenarios.

Furthermore, Figure 4 shows a fresh use case that hasn't been discussed yet. As we've seen, the deployment of infrastructure faces significant challenges in remote and rural regions due to the high costs involved. This situation could give rise to a new usage scenario in 6G, in addition to the well-known eMBB, URLLC, and mMTC. This new element, represented as the fourth pillar, would involve "basic internet connectivity" [28]. This approach would provide inferior performance in various KPIs but would still ensure minimum connectivity for users in remote and rural areas.

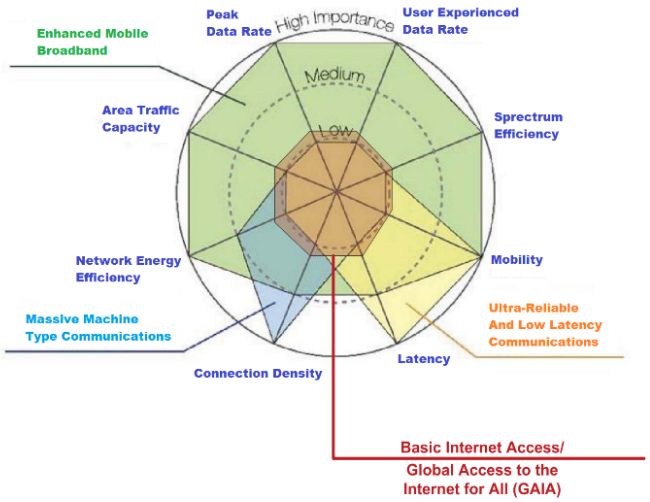


Fig. 4. Four pillars for 6G [28]

In order to understand how the new 6G network should be designed, some target indicators will be presented that exemplify the needs of this new generation of networks.

A. Latency

As shown, several new end-user and vertical industry applications tend to emerge with the advancement of technology, for example, autonomous vehicles, Virtual Reality, Augmented Reality, and holographic communication should be common applications in the future. These new use cases tend to require the same Key Performance Indication (KPI) as seen in 5G, but with new target values, for example, higher throughput, lower latency and better reliability.

Latency was a critical KPI in 5G and is expected to continue to be a concern in 6G networks, given that many applications are dependent on this KPI. On 5G, the minimum user plan latency requirement is 4ms for enhanced Mobile Broadband (eMBB) and 1ms for Ultra-Reliable Low Latency Communications (URLLC). This value is expected to be further reduced in 6G, to 100 µs or even 10 µs. In addition to air interface latency, 6G must also consider End To End (E2E) latency [28]. E2E latency is trickier to manage due to the myriad network elements involved, but 6G should overcome this challenge.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

B. Reliability

As with 5G, ultra-reliable, low-latency communications requirements will continue to guide the future 6G network. Although the 5G system has created an environment for a more secure system, its reliability mechanisms are strictly connectivity-oriented, therefore, the handling of failures in the application layer is left to the application itself. From the point of view of mobile networks, any instance outside its domain is considered outside the scope of treatment, but with 6G this should change.

In addition to enhancements to existing 5G security mechanisms, one of the most promising mechanisms for the sixth-generation system is Make-Before-Break-Reliability (MBBR). With it, it is possible to promote an interaction between the application servers and the mobile network, in order to detect failures. In short, MBBR gives the mobile network the possibility of previously detecting problems and security flaws in the application servers and transferring a problem-free copy to a redundant application server. In this way, the communication sections between the end device and the application will receive treatment from the 6G network, which will surely promote another layer of reliability for the system, making it a truly ultra-reliable network [12].

C. Terahertz Communications

Communications in Terahertz work between 100GHz and 10THz and, compared to millimeter waves, they bring great potential for high-frequency connectivity, enabling high data rates, in the order of hundreds of Gbps, which is what is expected from 6G.

On the other hand, the main problems in adopting this type of communication are directly linked to problems of propagation, molecular absorption, high penetration loss and major challenges related to antennas and Radio Frequency (RF) circuits [30].

In the case of millimeter waves, the propagation loss can be compensated using antenna arrays and spatial multiplexing with interference limitation.

Terahertz communications can be maximized by operating in frequency bands that are not severely affected by molecular absorption. And, finally, because these are very high frequencies, for indoor scenarios, it will be necessary to enable new types of RF solutions and ultra-small scale antennas.

Based on the characteristics of this type of transmission, the 6G network architecture will be directly impacted. For example, density and high data rates will increase demands on the capacities of the transport network, which must provide more fiber access points and greater capacity than current network backhauls. Furthermore, the wide range of different communication media available will increase the heterogeneity of the network, which will have to be managed [7].

To overcome these challenges, most of the conventional resource allocation algorithms are designed using high-speed fiber backhaul links, which are not applicable due to geographic limitations in historic buildings.

Fortunately, the very short wavelength in the THz band allows the use of an ultramassive array of antennas, i.e. containing 256, 512 or even 1024 antennas in the transmitter, which can provide a high beamforming gain to compensate for the loss of propagation. Meanwhile, precoding with multiple data streams can be used to provide multiplexing gain to further improve the spectral efficiency of THz systems. In the THz band, hybrid precoding that combines digital and analog domain signal processing is promising, as the number of RF chains is substantially less than that of full digital precoding, while achieving superior performance [31].

A good comparison of the key THz propagation characteristics and their impact on THz systems, is depicted in Table 1 [32].

TABLE I
THz WAVE PROPAGATION CHARACTERISTICS AND IMPACT ON THz SYSTEMS

Parameter	Impact on THz Systems
Free-Space Pathloss	Distances are limited to tens of meters at most
Atmospheric Loss	Significant absorption loss Useful spectra limited between low loss windows
Diffuse Scattering & Specular Reflections	Limited multipath & high sparsity
Diffraction, Shadowing and LOS Probability	Limited multipath & high sparsity Dense spectral reuse
Weather Influences	Attenuation caused by the rain

D. User-Experienced Data Rate

The user-experienced data rate, as the name suggests, is the throughput that users will perceive in the vast majority of their interactions with the system. This indicator is important because the majority of revenue for operators still comes from regular users and their smartphones, making it essential to offer high data rates.

In the 5G context, in a dense urban scenario, the user-experienced data rate is 100 Mbps for downlink and 50 Mbps for uplink. For 6G, it is expected to provide 1 Gbps or more in the downlink, which is ten times faster than 5G.

It is also important to note that users have a 95% chance of receiving this data rate at any time and in any location within the coverage area. The remaining 5% is allocated to moments of network overload or regions where the signal level is not as favorable, such as at the cell edge, for example. Furthermore, measuring the perceived performance by the user at the cell edge is also crucial as it can reflect factors such as appropriate site density, system architecture, and optimizations, among others. These indicators are valuable for operators, as they can provide adequate coverage and optimize operational costs [33].

E. Energy Efficiency

One of the most discussed aspects currently is the reduction of carbon emissions in the atmosphere, and although the implementation of a mobile system does not generate a direct impact in this scenario, the energy efficiency of the system is

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

certainly a factor to be taken into account, as part of the energy used to power mobile systems may come from non-renewable and polluting sources.

During the deployment of 5G, energy consumption was a closely observed aspect by the industry and standardization organizations, resulting in a significant reduction in energy required per bit compared to previous systems. As for 6G, this indicator suggests that there will likely be an increase in energy efficiency by an order of 10 to 100 times compared to its predecessor [33].

F. Fog Computing

Related to the topic mentioned above, there is another Target Indicator, which is called Fog Computing.

The Fog Computing has great importance in relation to energy efficiency, especially in cases of Massive IoT type communications where the majority of devices in IoT networks are battery powered with limited computational and communication resources.

This technology can provide storage and computational services for 6G networks and allows edge devices to perform computing and storage operations closer to the edge [33]. One of the great strengths of Fog Computing is the inclusion of decentralized computing services compared to the centralized computing offered by the traditional cloud.

Additionally, improved latency can be achieved as data and tasks are accessed and analyzed closer to end devices. The Fog Computing also improves the use of the frequency spectrum and increases network capacity.

Still on energy efficiency, in order to maintain the energy supply of any device, it is necessary to have the collaboration of other devices that can be called helper nodes. The function of helper nodes is to perform tasks on behalf of other devices. This process of sharing resources is called task offloading. Tasks are divided into tasks executed by any device and offloaded tasks that are executed by auxiliary devices. In offloaded tasks, there is an important point, which is the time restriction or maximum time limit for completing the task. Task offloading can improve latency through parallel processing of tasks, but it can also, unlike, increase latency through uplink transmission of the task to task helper nodes and downlink transmission of the result to the local device [35].

IoT networks that are based on Fog Computing do not have a standard architecture and can be represented through layers. Although the three-layer architecture is the most commonly used, there are proposals for architecture based on four and six layers.

In the three-layer architecture, there is a cloud layer, fog layer, and edge layer, and the resources are distributed in a hierarchical manner, that is, servers in the cloud layer with more resources, nodes in the fog layer with mid-features, and edge devices with fewer features.

In addition to layer-based architectures, it is possible to consider three other types: a) Clustered Architecture, b) Centralized Architecture, and c) Distributed Architecture. The Clustered Architecture is widely used in Wireless Sensor Networks (WSNs) as sensor nodes are grouped together in clusters of different sizes. In each cluster, there is a master node that controls the flow of information. All other nodes forward their data to the master node, which aggregates all the data and sends it to the fog node. In this type of architecture, there is a reduction in cluster transmissions, improving the energy efficiency of the system. Figure 5 illustrates this architecture.

In the centralized architecture, there is also the existence of the master node, called Fog Cluster Head (FCH), and member nodes, called Fog Cluster Members (FCM). Master nodes and member nodes are selected according to the geolocation between the FCH and FCMs and through internal policies.

In a distributed architecture, there is no formation of clusters, and nodes can be selected according to their availability and also according to internal policies.

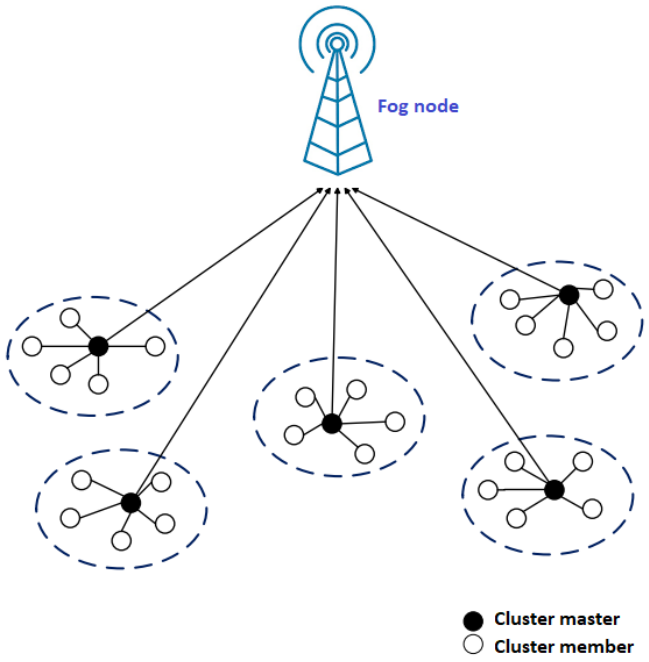


Fig. 5. Clustered Architecture to Fog Network [34]

G. Communication on Smart Surfaces

With the need to increase spectral and energy efficiency, increase data rates and higher frequencies, the use of massive MIMO, which is already a reality in fifth generation networks, will continue to exist in 6G. However, for sixth-generation networks, massive MIMO must work with smart surfaces that are matrices capable of controlling and amplifying wireless signals in targeted environments. These surfaces allow innovative forms of communication, as the use of radio frequency and holographic MIMO is possible [36].

H. Edge AI

In sixth-generation networks, the use of Artificial Intelligence will be imminent considering, for example, the possible creation of a Self-Sustaining Network (SSN) that can manage

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

resources, control the network, and maintain, autonomously, the high KPIs of the network. These and other functions enabled in 6G through the use of AI will be complemented by the use of AI at the edge, by running AI and learning algorithms on devices to provide distributed autonomy [36].

VI. Conclusions

In this article, an overview of 6G was presented, the expectations of society as a whole for the coming years in relation to this new technology, the preparation of the ITU in the construction of IMT-2030, and also a comparison of requirements with 5G. The study was directed towards researching some of the new use cases that will be introduced with the arrival of 6G, in order to present its objectives, characteristics, and necessary requirements for its operation. The Digital Twins use case makes it clear that machine learning, cloud solutions, and fast and reliable communication will be some of your key requirements. The Human-centric immersive communications use case presents needs such as bit rates in the order of Tbits/s and very low latency. The Industry 5.0 use case presents the requirements already mentioned in the previous use cases as a basic need. The challenge seen in Global Ubiquitous Connectivity has already begun to be addressed in 5G networks, however, to achieve truly ubiquitous global coverage, the insertion of both medium and low-orbit satellites into the 6G network architecture must occur. In the case of Pervasive Intelligence, the use of humanoid robots that come physically close to a human being will be one of the most promising scenarios. The use of computational resources offered by 6G will be a great opportunity for these robots to be more energy efficient. Finally, in the case of remote communications, the great expectation is that 6G will allow the digital era to be introduced into rural and remote populations. After examining these use cases, the requirements, also known as target indicators, were discussed, confirming the need for a revised, rather than entirely new, network architecture. This is particularly evident in the case of remote communications use, where infrastructure implementation will face significant challenges. This could lead to the emergence of a new pillar for 6G dedicated to guaranteeing minimum connectivity for remote and rural users. With ultra-low latencies, in the order of 10 micros, ultra-reliable networks, and transmissions in the order of Terahertz, it is expected that new network elements will be introduced, as well as the communication structure between them will be modified. An indicator that will also require a stable and consolidated network architecture is the data rate experienced by the user, as even with new applications and services, use cases will continue to demand increasingly higher rates so that the user experience is unique. Finally, energy consumption is one of the most important aspects observed since 5G, therefore, energy efficiency must be considered as a point of extreme attention when building sixth-generation networks. To this end, one of the technologies that is already expected to contribute to this end is Fog Computing, which can provide storage and computational services for 6G networks and allow edge devices to perform computing and storage operations closer to the edge. The 6G is expected to have intelligent and distributed network management in such a way that it can handle all demands privately and securely. All this must occur so that the success of the 6G deployment is possible and that all the desired objectives are achieved.

References

[1] O. Francine Cássia, M. Gustavo Iervolino, S. João Victor Menino, N. Ramon Magalhães, "Use Cases and 6G Architecture: New Needs and Challenges,"International Conference on Networks, Venice, Italy, 2023.

[2] ITU-R, "IMT traffic estimates for the years 2020 to 2030," Report ITU-R M.2370-0, July 2015.

[3] L. Zhang, Y. Liang and D. Niyato, "6G Visions: Mobile ultrabroadband, super internet-of-things, and artificial intelligence," China Communications, vol. 16, no. 8, 2019, pp. 1-14.

[4] J. Zhu, M. Zhao, S. Zhang and W. Zhou, "Exploring the Road to 6G: ABC - Foundation for Intelligent Mobile Networks," China Communications, vol. 17, no. 6, 2020, pp. 51-67.

[5] S. Zeb et al., "Industry 5.0 is Coming: A Survey on Intelligent NextG Wireless Networks as Technological Enablers," Journal of Network and Computer Applications, vol. 200, 2022.

[6] N. H. Mahmood et al., "A Functional Architecture for 6G Special-Purpose Industrial IoT Networks," IEEE Trans. Ind. Informatics, vol. 19, no. 3, 2023, pp. 2530-2540.

[7] P. Mitra et al., "Towards 6G Communications: Architecture, Challenges, and Future Directions," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021.

[8] M Giordani, P. Michele, M. Marco, R. Sundeep and Z. Michele, "Towards 6G Networks: Use Cases and Technologies," IEEE Communications Magazine, vol. 58, no. 3, 2020.

[9] L. Bariah et al., "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," IEEE Access, vol. 8, pp. 174792–174820, 2020.

[10] L. U. Khan et al., "Digital Twin Enabled 6G: Vision, Architecture Trends, and Future Directions," IEEE Communications Magazine, vol. 60, no. 1, 2022.

[11] A. Rasheed, O. San and T. Kvamsdal, "Digital Twin Values, Challenges and Enablers from a Modeling Perspective," IEEE Access, vol. 8, 2020.

[12] W. Tong and P. Zhu, "6G The Next Horizon," Cambridge University Press, 2021.

[13] N. P. Kuruvatti, A. H. Mohammad, H. Bin, F. Amina and D. S. Hans, "Empowering 6G Communication Systems With Digital Twin Technology: A Comprehensive Survey," IEEE Access, vol. 10, pp. 112158–112186, 2022.

[14] Z. Zhang et al., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," IEEE Vehicular Technology Magazine, vol. 14, pp. 28-41, 2019.

[15] M. Matinmikko-Blue et al., "White Paper on 6G Drivers and the UN SDGs," arXiv:2004.14695 [cs, eess], Apr. 2020, arXiv: 2004.14695. [Online]. Available: http://arxiv.org/abs/2004.14695

[16] S. Chen, Y. Liang, S. Sun, S. Kang, W. Cheng and M. Peng, "Vision, Requirements, and Technology Trend of 6G: How to Tackle the Challenges of System Coverage, Capacity, User Data-Rate and Movement Speed," IEEE Wireless Communications, vol. 27, no. 2, pp. 218–228, Apr. 2020.

[17] Z. Qu, G. Zhang, H. Cao and J. Xie, "LEO satellite constellation for Internet of Things," IEEE Access, vol. 5, pp. 18391–18401, 2017.

[18] Y. Hu and V. O. K. Li, "Satellite-based Internet: A tutorial," IEEE Commun. Mag., vol. 39, no. 3, pp. 154–162, Mar. 2001.

[19] K. B. Letaief et al., "The roadmap to 6G: AI empowered wireless networks," IEEE Commun. Mag., vol. 57, no. 8, pp. 84–90, Aug. 2019.

[20] Boston Dynamics. Accessed: Jul. 29, 2023. [Online]. Available: https://www.bostondynamics.com/atlas

[21] ITU - International Telecommunication Union. The State of Broadband. Geneva: ITU, 2022.

[22] GSMA Intelligence. The Mobile Economy 2023.

[23] ITU - International Telecommunication Union. The Future of Virtual Health and Care.

[24] A. Chaoub, "6G for Bridging the Digital Divide: Wireless Connectivity to Remote Areas," IEEE Wireless Communications, vol. 29, no. 1, pp. 160-168, Feb. 2022.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[25] H. Saarnisaari et al., ”A 6G White Paper on Connectivity for Remote Areas,” ArXiv abs/2004.14699, 2020.
[26] A. Chaoub et al., ”Self-Organizing Networks in the 6G Era: State-of-the-Art, Opportunities, Challenges, and Future Trends,” 2021.
[27] Y. Zhang et al., ”A Survey on Integrated Access and Backhaul Networks”. 2021.
[28] E. Yaacoub, E. M. S. Alouini. ”A Key 6G Challenge and Opportunity - Connecting the Base of the Pyramid: A Survey on Rural Connectivity,” Proceedings of the IEEE, vol. 108, pp. 533-582, 2020.
[29] W. Jiang et al., “The Road Towards 6G: A Comprehensive Survey,” IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021.
[30] Q. Zhang, W. Ma, Z. Feng and Z. Han., “Backhaul-Capacity Aware Interference Mitigation Framework in 6G Cellular Internet of Things,” IEEE Internet of Things Journal, vol. 8, pp. 10071-10084, 2021.
[31] L. Yan, C. Han and J. Yuan., “Hybrid Precoding for 6G Terahertz Communications Performance Evaluation and Open Problems,” 2nd 6G Wireless Summit, 2020.
[32] H. Tataria et al., “6G Wireless Systems: Vision, Requirements, challenges, Insights, and Opportunities,” Proceedings of the IEEE, 109.7, pp. 1166-1199, 2021.
[33] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, ”The Road Towards 6G: A Comprehensive Survey,” IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366.
[34] U. M. Malik et al., ”Energy-efficient fog computing for 6G-enabled massive IoT: Recent trends and future opportunities,” IEEE Internet of Things Journal, vol. 9, no. 16, pp. 14572-14594, Aug. 2022.
[35] S. Luo et al., “Incentiveaware micro computing cluster formation for cooperative fog computing,” IEEE Transactions on Wireless Communications, vol. 19, no. 4, pp. 2643–2657, 2020.
[36] W. Saad and M. Chen, ”A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” IEEE Network, vol. 34, pp. 134-142, 2019.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

On the Study of Internet Ossification, Impacts, and Solutions

Lin Han, Richard Li
Futurewei Technologies, Inc.
Santa Clara, California, U.S.A
email: lin.han@futurewei.com; richard.li@futurewei.com

***Abstract*— The current Internet is based on IPv4 and IPv6. It has been in service for many years and is very successful. However, it is facing challenges in protocol ossification, security, and service quality. Recently, the geographical tension, trading confrontation, digital asset and digital sovereignty, the regulation for data protection and localization have raised decentralization requirements for the Internet. This paper analyses the factors for the Internet ossification and its impacts, it proposes a new architecture that is distributed based on region or country. It can maintain the support of the current IPv4/IPv6 and existing applications, and provide more flexibility for the protocol, thus mitigating the ossification of the Internet. With the new architecture, the Internet will be decentralized based on regional governance and provide more space for more diversities within different regions. Meanwhile, the global connectivity, accessibility and integrity of the Internet are kept.**

Keywords- Future Internet; Ossification; Decentralization; Distributed; Fragmentation.

I. INTRODUCTION

This paper is an extended version of [1], which investigates the Internet ossification, proposes a new architecture and protocol to solve the problem.

The Internet has penetrated everywhere in our life and has provided tremendous momentum to the development and progress in communication, technology, culture, and economy. The current Internet is based on IPv4 [2] and IPv6 [3] protocols, and consists of many other protocols for different areas, such as address assignment, domain name service, routing and switching, security, transport. All these protocols are governed by the Internet Engineering Task Force (IETF). In the document thereafter, the name IP represents both IPv4 and IPv6.

However, the Internet's deficiency and ossification are also noticed. This includes slow evolution, protocol ossification, resource allocation unfairness, security and privacy concerns. Digital asset [4] and digital sovereignty [5] are also debated in different countries and regions. All these problems are not easy to be solved under the current Internet architecture since those factors were never considered in the time of the Internet was born.

The paper briefs our research on a new architecture for the Internet and associated protocol structures. It can provide extra flexibility for the Internet while maintaining the current IP based technologies and services. Internet ossification can be mitigated by a new architecture including distributed Internet resource management and domain name service, free choice of address type, and heterogeneous communications.

The rest of the paper is structured as follows. In Section II, we present an overview of the Internet architecture and protocols. Section III discusses the Internet ossification and analyzes the root causes. The technical factors are analyzed in Section IV. Our new network protocol is proposed in Section V. Section VI presents the detailed design. Section VII illustrates the new Internet architecture with the new protocol. The compatibility issues are discussed in Section VIII. Sections IX and X summarize the advantages and disadvantages of the new proposal, respectively. Section XI concludes the paper and gives further research directions.

II. OVERVIEW OF THE INTERNET

The Internet is the global system of interconnected computer networks that uses the Internet protocol suite to communicate between networks and devices [6]. Recently, with the growth of 5G [7], Internet of Things (IOT) [8], Non-Terrestrial-Network (NTN) integration [9], the Internet has become the communication infrastructure that almost every person, every device and everything can be connected to. The Internet scope is very broad and has a couple of key fundamental blocks:

- The definition of IP address, the mechanism to allocate and assign the IP addresses. There are two types of IP addresses, one in IPv4 and another is IPv6. Currently, IPv4 is in the process of becoming obsolete from the perspective of IETF, and IPv6 is the only supported address. The IP address (except the local address and non-routed address) is globally significant and unique in the world. It is allocated by the Internet Assigned Numbers Authority (IANA) [10] to each region and country. There are five Regional Internet Registries (RIRs). Each RIR has a couple of Local Internet Registries (LIRs) or National Internet Registries (NIRs). They are responsible for the allocation of the IP addresses block on their authorized areas. Figure 1 and Figure 2 show the hierarchical architecture of IANA [11].
- The definition of Asynchronous System Number (ASN) [12], and the mechanism to assign ASN. ASN is used for BGP [13] to represent autonomous systems across the Internet. Similar to IP address, the public ASN is also globally significant, it is managed by IANA. ASN is key to BGP that is critical protocol for the inter-connection and inter-working of different networks distributed globally. BGP will exchange the global IP address of different networks, thus making every global IP address reachable from anywhere around the world.
- The definition of Domain Name, the mechanism to manage Domain Name Servers and provide the Domain Name System (DNS) [14] Service. Similar to IP address, Domain Name is also globally significant. The DNS root zone management [15] and DNS root servers [16] are managed by IANA as well. Domain Name and Domain

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Name Servers are distributed globally. There are thirteen DNS root server located in U.S.A. Different leaf servers belonging to different region and country are deployed globally. In addition to this, some countries may have mirror root servers in their own region to back up the root server and speed up the DNS services.

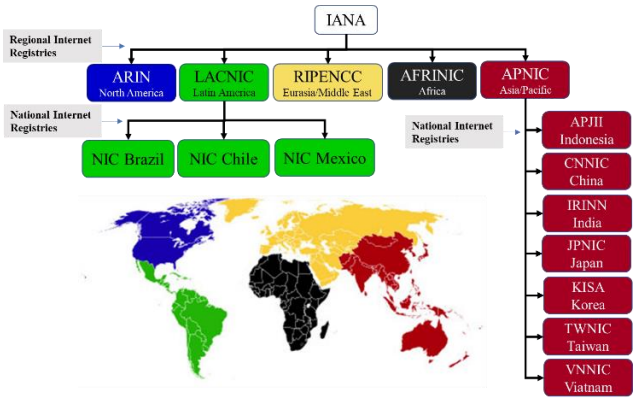


Figure 1. The hierachy of IANA architecture

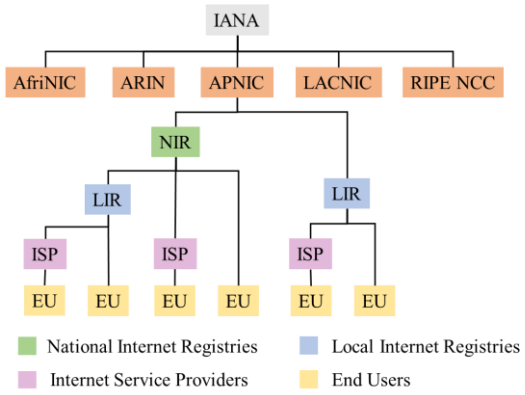


Figure 2. Understanding address management hierarchy [10]

- The protocols to control the Internet. The fundamental protocols are IPv4, IPv6 and many other protocols on top of IPv4 and IPv6. Excluding protocols on L2 that are controlled by the Institute of Electrical and Electronics Engineers (IEEE) and the International Telecommunication Union (ITU), the protocols for Internet include layers from L3 to L7 that are controlled by IETF. There are thousands of protocols related standards that are called RFC (Request for Comments) documents, e.g., more than 500 RFC for IPv6 has been published. Below just lists a very small portion of RFCs and very typical protocols:
  1. Host configuration related protocols (ND[17], DHCPv6[18], etc.)
  2. L3 or routing protocols (BGP, IS-IS [19], OSPF [20], etc.),
  3. Traffic Engineering (MPLS [21], RSVP-TE [22], SRv6 [23], etc.)
  4. L4 or transport protocols (TCP [24], UDP [25], etc.),
  5. Upper layer protocols (QUIC [26], TLS [27], HTTP [28], etc.),

III. INTERNET OSSIFICATION

A. Root Cause

The Internet was essentially designed with simplicity and scalability. [29] has detailed analysis of how this is achieved and lists the important timeline for Internet evolution. After the Internet becomes available to the public in the 1990s, it experienced more than 40-years' development of technology. Gradually, the evolution of the Internet becomes slower and slower. There are less and less new technologies and services coming up for the Internet, especially for the parts of infrastructure and fundamentals. The structure of the internet becomes more rigid and difficult to change over time, and this sometime is called Internet ossification. For example, IPv6 was designed to replace IPv4, but this has not been accomplished since the first IPv6 standard RFC 2460 [30] was introduced in 1998. Even right now, there are still arguments that IPv4 should not be obsoleted [31], and the adoption of IPv6 in Service Provider is still slow.

There are couple of research that proposed new or enhanced architecture for Internet, such as RINA [32], SCION [33], New IP [34], IPv10 [35], and Extensible Internet (EI) [36][37]. Detailed analysis and comparison of proposals of RINA, SCION and New IP can be found in [38]. IPv10 is to allow the communication between IPv6 and IPv4. EI introduces Layer 3.5 between L3 and L4 to provide services that were not available in the current Internet architecture.

Two categories of factors associated with management and technical solutions can contribute to the Internet ossification:

- Consensus challenges:
  The Internet is a huge global network. Many technical definitions, solutions, and changes are globally significant. Any decisions or changes about its development, operation and deployment involve a wide range of stakeholders, including governments, organizations, operators, and individual users. Reaching consensus on changes can be very difficult and slow, especially when there are competing interests or different priorities. As a comparison in the standardization in wireless area, 3GPP has finished the 5G (the fifth generations of wireless technology) in almost the same period that IETF has not completed the IPv4 to IPv6 transition.
- Technical solutions:
  Due to the vast number of users, devices and applications, the Internet has accumulated many technical feedbacks and problem reports. Completely fixing those problems or enhancing the existing solutions are always slow. Some quick fixes that are implemented in a short term, but may need to be addressed or replaced later on. The slow global consensus on any problem fixing, new enhancements or features, can make it more difficult to change any piece of the internet's infrastructure. The Internet is a complex system that involves many different networks, technologies, and standards. How to drive the Internet moving forward but maintain the previous investment is

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

not only a business objective but also a technical challenge. Ensuring compatibility between these different elements can be difficult, and changes to one part of the system may have unintended consequences elsewhere. Due to this reason, people are always conservative and hesitate to adopt new technologies.

B. Consequence of Internet Ossification

The Internet ossification has impacted the internet's ability to continue evolving and progressing. It contributes more or less to the slow solution for following issues and requirements:

- Privacy and Security: These two contradictory requirements have never been solved with satisfaction from different parties. To solve the privacy issues, IETF has had the Working Group for "Host Identity Protocol" [39], HIP [40] provides a cryptographic namespace to applications, and the associated protocol layer, thus provide the best privacy protection. But since many nations do not want such information invisible to the law enforcement for the sake of security, this protocol was never widely deployed. TLS [27], HTTPS [28], IPSec [41] are all security protocols at different layers and are widely used in Internet, but the Internet security issues never disappear even many security events are not associated with the technologies used. Distributed Denial-of-service (DDoS) attack [42] is one of the most notorious security issues for many years. It caused lots of business losses and may lead to international conflict if the DDoS source and victim are in different nations. The current technologies to stop DDoS attacks need to have protection mechanism at different places from connected service provider network to the cloud the application is running [43], the solution is quite extensive and needs coordination between different organizations. To eliminate such attacks, without some Internet infrastructure changes, it is quite difficult.
- Digital Asset and Digital Sovereignty: Bitcoin has been very succeeded in its security, value growth and become a hot trading target, but it has never been recognized as legal currency for the legal business. Non-Fungible Token (NFT) is another type of digital identifier for any digital asset, its recognition is also doubtful due to no endorsement from any government or authorization. The Internet is a network with unified address, protocol, and centralized resource management. The failed acceptance of Bitcoin and NFT have driven us to think whether we should consider the requirements from the sovereignty at the original design of the Internet. Since none of the basic Internet resources (IP address, ASN and Domain name) is controlled and managed by a government or authorized administration for a country, it will naturally cause concerns. Digital Sovereignty is a controversial topic in the European Union and other countries recently. Even though its scope, target and method are still to be decided, it has raised a question how the Internet can be designed to consider such factors.
- Fragility of Internet Architecture
  Even though the Internet architecture is claimed to be distributed and resistant to failure of partial network, it has never been tested for large scale failure due to unexpected incidents like nature disasters or war. The current Internet only relies on the BGP to establish new routes whenever some global network is not reachable. However, since the Internet scalability is super large now, any failure of some links crossing small regions may lead to unexpected and large scale of consequences. The research in [44] has indicated that the Internet in non-relevant countries will be severely degraded if some links between China and Taiwan are cut. [44] has also given the detailed analysis for the reason why such small scale of link failure can lead to large scale of impacts to the Internet, it also proposes to study "Wartime BGP routes" as a short-term solution to handle such scenario.

IV. Design Factors for Internet Ossification

Even though there are many factors, technical or non-technical, contributing to the Internet ossification, we think some short-term design of Internet has made Internet less flexible at the beginning, thus is one of the most important factors we need to consider when thinking about the future architecture. The following are some technical perspectives that contribute to the Internet ossification.

- The Internet resource (IP address, ASN and Domain Name) assignment and management are essentially a centralized hierarchical architecture. The problem of this centralized architecture is that (1) IANA and Regional Internet Registries are both non-profit organizations that do not have any jurisdiction. (2) The Internet resources are hardly allocated fairly, for example, IPv4 address block is not enough in some countries but more than required in other countries. (3) Address preference is not the same in different regions, countries, operators, users, and applications. For example, IPv4 is still preferred by many service providers and enterprise network. That is one reason that IPv6 deployment is so slow. (4) Centralized architecture makes the Internet fragile when the geopolitical tensions are high. In the recent events of war and trading confrontation, some voices to stop the Internet service to specific area is around and has put the threat to the integration of Internet.
- Since IP address is globally significant, it requires that all end-user devices and network devices use IP as unique format for the data packet header, all L3 devices should follow the same principle to process IP packet and provide the services to upper layer. This design is called "narrow waist". Obviously, it has benefits in simplicity and scalability, but it becomes one factor contributing to the Internet ossification, since any changes in IP header will have global impact and hard to get consensus in IETF.
- From the IP packet forwarding perspective, the IP based Internet is flat. All internet packets are forwarded based on IP address lookup; thus, all globally reachable IP addresses must be stored in every network device (even in MPLS network, the Provider Edge (PE) Routers also must store all reachable IP prefix). This can result in two problems: (1) huge amount of IP addresses or prefixes storage leads to huge lookup table size. (2) BGP, the only protocol to exchange the global IP reachability between

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

different networks in different regions or countries, must process huge number of global IP prefixes. Any small internet state changes may lead to BGP re-route huge amount of traffic as described in [44].

V. CONSIDERATION OF NETWORK LAYER

A. Tecnology Progress Considerations

From the analysis in Section IV, we can see that one of the major factors for Internet ossification is the IP design is too rigid. Such rigid design was partially because the hardware or semiconductor performance was limited in the 80s and 90s in the last century. To achieve the line rate of packet processing, it is hard to give too many flexibilities in the address and functions in the packet header, e.g., the address type and size, the extensions, and options. After many years' development, the semiconductor industry has progressed a lot. Recently, high-performance chips with programmability have been commercialized. It is time to think about what we can do from a technical perspective that can mitigate the Internet ossification.

B. Requirements of Internet Decentralization

1) Compared with other system

As a global data communication network, the Internet is supposed to be only responsible for the inter-connection between different networks in the world. The networks could be for enterprises, ISP (Internet Service Providers), a country or a region. Let us compare the similar situation in phone network and mail system. For those two global communication systems, there is no restriction on how to define a local phone number, and local address format. The international community only needs to get consensus on the country code for international calls, or the country names for global mail delivery. Each country will manage and design its own structure of phone numbers, mail addressing system and delivery infrastructure. We think the Internet should take the same approach.

2) Regulation requirements

Recently, more and more countries or regions have new legal requirements for international ISP to provide the service in the country. For example, the internet service provider's infrastructure, including cloud, computers, storages, etc., that is associated with the locally provided services, must be deployed within the territories of the country. All provided services (applications, contents, accounting, etc.) should comply with local regulations for security, privacy, etc. These regulations naturally require ISP to have a decentralized Internet infrastructure and a decentralized Internet service. From this perspective, the major international ISPs already deployed their infrastructure and services in a distributed manner crossing different countries or regions.

3) Trendes for the content localization

To achieve better service (higher bandwidth, shorter latency, less probability of congestion) for content delivery, the content servers or data centers are moving closer to data consumers. This trend has been accelerated after 5G introduced the Mobile Edge Computing (MEC) technologies. Moving closer to data consumer needs to have the localization in Content Delivery Network (CDN), associated APP (Applications), Name resolving, Content searching, etc. All these trends lead to the Internet traffic to be grouped on the base of population and sovereignty.

C. Design Principals for Ideal Internet

Considering all above analysis for Internet history, the current requirement and trends happened for Internet, if we have a chance to redesign an ideal Internet, we may have following principals:

- The Internet should have more flexibility, less restrictions and centralization. Keeping the technology diversity for the Internet will not only reduce the ossification but also satisfy different requirements easier.
- The Internet should be distributed globally based on region or country. All regions are equal and there is no central control. No region can impact other's decision in address selection, peering and service.
- Small countries can decide to form a region if the countries do not want to be independent in internet resource and DNS management due to economy and other constraints.
- Each region has the freedom and authorization to manage the Internet resources used locally, such as address selection, address allocation, ASN allocation, domain name registration, DNS root server, etc.
- The internet should support heterogeneous address types and communications.

VI. DESIGN DETAILS

The key aspects of the new architecture are as follows:

- The Internet for each country or region is connected by a separate protocol. We have two options for this protocol. One is to design a new protocol (described in the sub-section A), and another is using the current IP technology (described in the sub-section B). The comparison of two options is discussed in sub-section C. The paper focuses on the discussion of using the new protocol.
- Each country or region will have independent internet resources including IP addresses, ASN number, DNS, etc. All these resources are managed by the country or region. Since the details of these architecture changes for two options (described in sub-section A and B) are the same, the paper will only focus on the discussion of the architecture changes, compatibility issue and benefits (in Sections VII to IX) for the 1st option or using new protocol.

A. Using a New Protocol

The new network protocol packet header for the Internet as shown in Figure 3. The packet format is preliminary and only for illustration. Final design will decide the detailed coding. This new packet is on top of Layer 2, thus, a new EtherType assignment from IANA is required.

Below is the explanation for each field in the Figure 3:

- Declaration: This field defines the basic info about the packet, it may contain following essential info:
  1. HL: Hop limit, this value is decremented by one at each forwarding node and the packet is discarded if it becomes 0 (except on the last node).

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

2. Prot: The protocol number for payload, it could be a protocol number defined currently by IANA, e.g., IPv4 or IPv6, TCP or UDP, or a new protocol number defined in the future.
3. Len: Total length of the packet including the Pay Load. The unit can be defined in standardization.
4. Other definitions: other definitions for the packet header, it will be defined later.

- Regional codes: This field may contain the "Src (Source) Region Code" and "Dst (Destination) Region Code" for source and destination. The size, code structure and detailed coding should be standardized by an international organization. It could contain region or country code that was defined by ITU E.164 [45], and have its own hierarchy, e.g., region, sub-region, and more granular definitions. See Figure 4 as an example. Only the 8-bit "Region Code" needs to be standardized by an international organization, "Sub-region code" will be managed locally in the region.
- Service: This field contains information about the service and is to be defined. Its length is variable.
- Payload: This part contains the payload which type is specified by the protocol number defined in Declaration. The Payload could be IP type or any other types for L2 to L4.

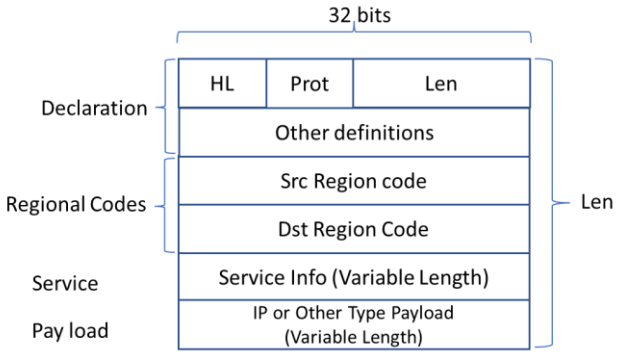


Figure 3. New Internet protocol packet header

Figure 4. The Region Code Example

B. Using Current IP

This option will use the existing IPv4 or IPv6 technologies to interconnect the networks in different countries and regions. By this option, the architecture for the internet is the same as by using a new protocol (sub-Section A). Following works must be done:

- IANA should permanently reserve some un-used IPv4 or IPv6 addresses, then each country or region will have a permanent IP address assigned by an international organization. This address is similar to the area code for telephone system and can only be used to connect different countries. Whether each country will be assigned multiple IP address will be decided by the international community.
- The IPv4 or IPv6 tunnels between countries and regions are established. These tunnels are only used for the traffic crossing border.
- Each country or region will develop its own address assignment, management, and DNS server system. After all these systems are set up, the country can switch those management from the current to local.
- An international organization is responsible for the DNS root connection and traffic distribution between countries and regions.

C. Comparisons of Two Options

- Using the new protocol can give us chance to go through all possible design aspects, make it possible to fix the problems of the current Internet and to satisfy future requirements, thus, it should have longer term benefits.
- Using the existing IPv4 or IPv6 is simpler than using a new protocol, but it will not have the benefits of the new protocol, e.g., it may not support the services that can be introduced by the new protocol. Additionally, it will overload the original IPv4 or IPv6 address definition (prefix plus length) for the use of Point-to-Point interconnection between countries, some existing address aggregation, forwarding, and protocols have to be re-examined to make it not conflicting to the existing IP network.

VII. Architecture for Internet Based on New Protocol

A. Internet Resource Management

The internet resources will include region code, IP address space or other type of address space, ASN, and protocol number. The management of those resource are based on following rules:

International organization managed items:

- The Region code structure and Region code assignment are responsible by international organization, ITU or IANA.
- For the protocols that the interconnection between different region or country are supported, e.g., the new protocol defined by this paper (new EtherType), IPv4, IPv6, Ethernet, MPLS, etc., the protocol numbers are still managed by international organization IANA.

Regional authority managed items:

- Each region or country will be responsible for the sub-region code assignment and management.
- Each region or country will be responsible for the IPv4/IPv6 address and ASN number allocation and management for its own jurisdiction area. Different regions or countries may have different policies and schemes to manage the resource.
- Each region or country can use the whole IPv4/IPv6 address and ASN space. All addresses only have local significance in the region or country, thus different regions or countries may have the same address.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

- Each region or country can define new protocol numbers that are only used locally within the region or country.

B. Scope of New Protocol

The new protocol applies to the internet connection between different regions and countries as shown in Figure 5. It does not restrict communication within the region or country. The current IPv4 and IPv6 can still work. A region or country can define and run a new version of IP without any interruption or interference to the whole Internet. For example, IPv10 to support communication between IPv4 and IPv6 was proposed in IETF but was not accepted. With the new protocol, one region only needs to get consensus on IPv10 in its own sovereignty and then use it within the region.

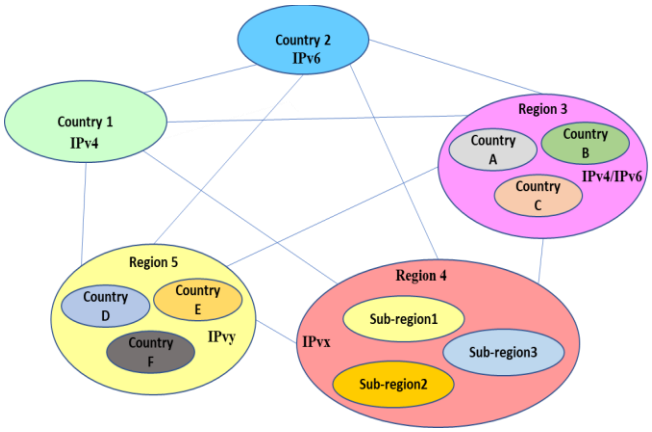


Figure 5. Internet based on new network protocol

It is important to note that a region can also use the new region-based protocol for communication within its own territory (see the communications between sub-regions in Region 4 in Figure 5).

C. Domain Name Service

The Domain Name Service architecture is similar to the current DNS hierarchy architecture, Figure 6 illustrates the new DNS architecture and Figure 7 demonstrates a DNS request and response crossing different regions or countries. The major difference with the current architecture is that the current centralized DNS root zone and root servers are removed, thus is a distributed architecture. Following are details:

- Each region or country will have its own DNS root server and different root servers from different regions or countries are fully equal and there is no central control, thus the current DNS root zone and root servers not needed.
- All DNS root servers are connected virtually to form a DNS network. The addresses of all root servers can be based on the new protocol, thus are unified for different regions. The network may run a dedicated protocol to exchange DNS information for all root servers. This network will be overlay on top of either existing IP or the new network protocol proposed in this paper.

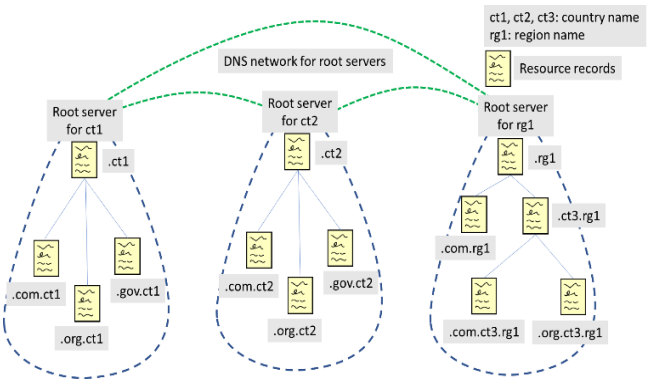


Figure 6. Domain Name System architecture

- The connection between all DNS root servers are fully meshed virtually. Any connection between two servers are voluntary and only managed by two servers' regions or countries. When a new root server for a region or country joins the network, it should have agreement and then connection with existing root servers.
- The ".region" or ".country" domain is the only Top Level Domain (TLD) for the region or country. All other domain names are lower-level domains.
- The ".region" or ".country" suffix is needed when the DNS requester and real domain name are in the different region or country. The suffix can only be omitted when the DNS requester and the real domain name are in the same region or country.
- A domain name with a ".region" or ".country" suffix is always associated with an address physically located within the region or country.

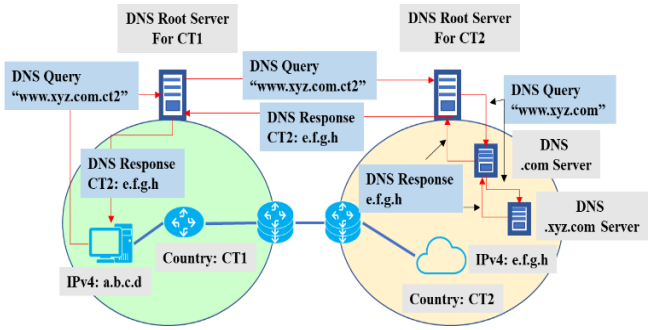


Figure 7. DNS service crossing different regions or countries

The DNS service will have some corresponding implementation changes with the new architecture. Also, there are some regulation or legal issues involved, e.g., a company name in a "domain name" in a different region must be approved by the local authority.

Here is an example: An international company xyz has the header quarter in the country named as "ct1", then the domain name "www.xyz.com.ct1" always points to an address assigned by the DNS authorization in the country ct1. In another country ct2, if there is a branch or service from the company xyz, the DNS request of "www.xyz.com" from ct2 will return an address info found in the name server ".com" in

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

the country ct2. If there is no registration for the company in ct2, DNS request of "www.xyz.com" from ct2 will return null.

Due to the bonding of a name and IP address in every region physically, the new DNS mechanism will make the internet service localization more transparent and easier to be compliant to the local regulation or laws.

D. Communication Between Region or Country

To provide interconnection between different regions or countries using new network protocol, proper control plane and data plane must be defined.

1) *Control Plane*

- The border devices connecting different regions need to support the new control protocol.
- The new control protocol will exchange information about the interconnected border devices, the associated links, the region code, and the reachable end-user's address details, etc.
- The new control protocol could be link-state routing protocol like IGP, or path-vector protocol like BGP.
- New control protocol also must be running within a region or a country to populate the information learnt from border devices about the outside interconnected networks of other regions or countries, e.g., the links that can reach other regions or countries, the associated remote reginal code, the remote reachable address associated with the regional code, etc.

2) *Data Plane*

- For the egress region, where the traffic is originated from, the data packet forwarding is based on the lookup of "Region/Country code" at all network devices. See the country CT1 in Figure 8.
- For the ingress region, where the traffic is destinated to, the data packet forwarding is based on the lookup of "the address of payload" at all network devices. See the country CT2 in Figure 8. In the example, the "address of payload" is IPv6 address.
- For the transit region, there are two approaches, one is Transparent Mode, another is Tunnel Mode.
  1. For Transparent Mode, the data packet forwarding is based on the lookup of "Region/Country code" at all network devices in a transit region. See the country CT3 in Figure 8.
  2. For Tunnel Mode, the data packet forwarding is based on the lookup of "Region/Country code" at edge network devices in a transit region. Proper packet encapsulation (at ingress router) or decapsulation (at egress router) are needed. See the country CT4 in Figure 8. In the example, the IPv4 tunnel is used and IPv4 address lookup for the tunnel is done on every network device within the region.
- For all scenarios, a very small table is needed to store all "Region/Country code" for the communication crossing regions. The table lookup will use "exact match". These two behaviors are different as the IP prefix lookup, which needs huge amount of table to store global IP prefix, and the lookup is Longest Prefix Match using TCAM (Ternary Content-Addressable Memory).

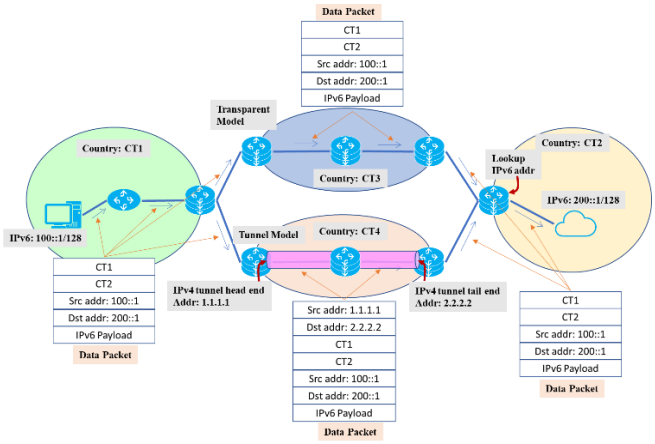


Figure 8. Homogeneous communication: Transparent Mode and Tunnel Mode (only the essential parts of packet header are shown)

3) *Heterogeneous Communication Between Region or Country*

The above discussions are about the homogeneous communication between regions or countries, or the address type are the same for all end users.

The new network protocol and architecture can support heterogeneous communication worldwide. Heterogeneous communications are communications with different types of address. This is very useful to many applications in security, privacy, IoT, etc., below are some supported address combinations for heterogeneous communication:

- Different length of IP for source and destination, e.g., IPv10 or other type of IP that the address length is not 32-bit and 128-bit.
- Different type of address for source and destination, e.g., between Ethernet and IP.
- No source address, the source address is hidden in the application data.
- Variable length public key as address.

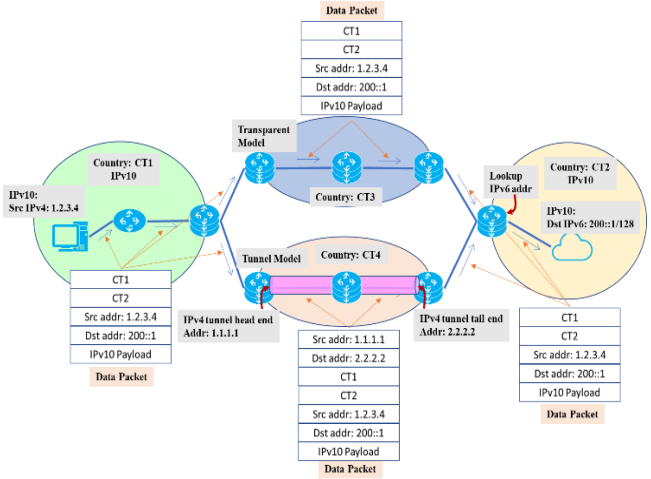


Figure 9. Heterogeneous communication: Transparent Mode and Tunnel Mode (only the essential parts of packet header are shown)

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Figure 9 illustrates the data plane for a case where IPv10 is supported in country CT1 and CT2, and how an IPv4 host in CT1 sends data to IPv6 host in CT2. For IPv10 case, both IPv4 and IPv6 address are supported, thus the lookup of IPv6 in CT2 is obviously supported. We can see that to support IPv10, only communication participants (CT1 and CT2) need to have an agreement to support it. This is much easier to have a global consensus to support IPv10.

VIII. Compatibility Issues

The major changes of the Internet based on the proposed new network protocol are the Internet resource management, the DNS architecture, and the use of new network protocol.

For the communication or IP service within the same region or country, the current IP based internet service can still be used, and there is no compatibility issue. The new Internet resource management and new DNS architecture have very little impact on the end-user application and network operation, i.e., some provisioning (to the DNS server and domain name management) may need to be changed.

For the communication or IP service crossing different regions or countries, the new network protocol needs to be used, and it is not compatible with the existing IP, but we can maximize the current Internet investment through the detailed design of new network protocol header.

It is easy to notice that the new network protocol packet header is very similar to the IPv4. This is intended to make the future design easier to be implemented in IPv4 capable hardware. We have two options in the final design of the packet header encoding: (1) re-use the IPv4 packet header for the new network protocol, or (2) only re-use the 32-bit IPv4 address space for the region code and redesign other fields in packet header. Since the current IPv4 header has design flaws in some areas, such as: (a) The protocol is not extensible due to the limited IPv4 option size, (b) The header checksum is not required, (c) Fragmentation is not a good design. So, we prefer the option (2): define the 32-bit source and destination region codes; redesign other fields in the packet header.

With the above design considerations and coupled with redesigned protocol running between regions, by the minimal re-programming, the existing hardware can be easily re-used for the future Internet.

IX. Advantages of New Network Protocol

A. Benefits

The proposed new network protocol is only for the interconnection between regions and countries. The Internet based on new protocol will have following benefits:

- Much less restriction at the protocol for interconnection:
  The new network protocol only defines the regional interconnection mechanism that is based on regional codes, but not limit the communication address and communication mechanism within a region or a country, thus reduces the restriction caused by globally uniformed IPv6 header for global network. Heterogeneous communication support will be easier to achieve between interested parties.
- Minimized changes on the current Internet architecture:
  The current IPv4 and IPv6 protocols and data forwarding can still work in a region or country. DNS changes very little. The architecture of IP based Internet is kept, and the investment is not wasted.
  The control protocol and data forwarding for interconnection between regions and countries can be realized based on extension of existing IP routing protocols and IP packet forwarding. It needs minimal investment.
  Existing and future IP based applications within a region can still run without any feeling that the underlayer networking is changed for the interconnection between regions. The application to reach outside of a region just needs minor modification for the address format to include the regional codes.
  The routing table size will be dramatically reduced due to the fact that routers in a region will only keep the prefix defined in the region. All addresses to outside of a region can be summarized as regional codes.
- Independent technology evolution:
  With the new network protocol, Internet technology can evolve in different regions or countries independently. It is expected to be much easier and faster than the current situation that the global consensus is needed, thus will mitigate the Internet ossification a lot.
- Distributed Internet resource management and DNS:
  The new Internet resource management and DNS are distributed and based on sovereignty and jurisdiction, thus has no legal obstacles to making the regional Internet technologies adaptive to local laws or regulations. It will make any security, privacy changes or enforcement much easier and faster.
  The new Internet resource management and DNS root servers are distributed and fully controlled by a region or country. The Internet service of any country will not be impacted by other countries. It makes the Internet more robust and resilient to any disasters and geopolitical interruption.
  The new distributed Internet resource management also makes each region or country able to use the whole IP address space and ASN space. This will not only eliminate the unfairness issues in IP address allocation, but also expand the IP address resource for all countries.
  The new architecture and network protocol gives each region or country full control and freedom of what type of address and communication are used for the internet service within the region. This will eliminate the IPv4 to IPv6 migration mandates if IPv4 is preferred in a region or country. Also, other new types of address can be invented and adopted locally.
- Internet integrity is maintained:
  Internet fragmentation [46] is always a concern for new technology proposals. From a technical perspective, the new proposal does not impede the ability of systems to fully interoperate and exchange data packets. The Internet functions are consistent as before at all end points. Internet interoperability, universal accessibility, the reusability of capabilities, and permissionless innovation are all not impacted. While the data protection and localization from

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

many regional regulations can be naturally satisfied by the architecture, more freedom in addressing can provide more possibilities for new technologies in security and privacy.

B. Advantages

Comparing with the existing proposals, RINA, SCION, New IP, IPv10 and EI, the new proposal has following advantages:

- Unlike RINA and SCION, the new proposal is not a clean slate solution, it can keep the current IP based internet service in a region or a country unchanged, it only impacts the interconnection between regions and countries. Considering most of internet traffic is local and international traffic crossing borders of countries are relatively small, the impact to current internet service is limited. Additionally, for the impacted interconnections between regions, proper migration strategy can be developed to upgrade inter-links individually to new protocol and minimize the service interruption.
- The new protocol is orthogonal to other variations of IP, like New IP, IPv10 and EI. It can make those technologies easier to be adopted locally without global consensus and impacts.

X. Disadvantages of New Network Protocol

The proposal will have disadvantages compared to the current Internet architecture; these include:

- The Internet is no longer a unified and flat network with the same type of addresses. While we can obtain the benefits of the new internet protocol such as diversified address, architecture and technologies, we also lose the simplicity of the current Internet.
- The traffic crossing the boundary of regions and countries are discouraged. This is not economical sometimes, i.e., the same application may have to deploy more servers in different regions to provide the local services. This is the same side effect as the requirement to provide the localized services based the regulations in some major countries and regions.
- The root DNS servers distributed in different region or country will require the information exchanging and database synchronization. This is not needed for the current DNS system.

XI. Conclusions And Future Work

The paper has proposed a new network protocol and architecture that can provide more flexibility and mitigate Internet ossification. The new architecture is distributed without any central control, thus making the Internet more robust and resilient to geopolitical interruption. It can also expand the usable Internet resources for each region and country. Meanwhile, the new proposal can keep the current IP based Internet in regions, thus it can minimize the impacts to Internet and maximize the old investments.

Further works are needed for detailed solutions in every area where the new technologies or protocol redesign are required, such as protocol for distributed DNS, the control protocols and forwarding engine for interconnection between regions, upgrading and migration approaches, etc.

It must be noted that the purpose of the paper is to analyze the internet ossification and possible solutions for future internet. It is expected that any solution including the proposal in the paper will face a lot of questioning, challenges, and objections. For example, the basic IPv4 and IPv6 packet formats have never been changed since the 1st version were proposed in IETF. But it is believed that doing something will be better than doing nothing. As the most important invention of human beings, the Internet can only be pushed forward after whole interested parties join the work and contribute the ideas.

References

[1] L. Han and R. Li, “On the Study of Internet Ossification and Solution,” Internet 2023, IARIA, https://www.thinkmind.org/articles/internet_2023_1_30_40006.pdf.

[2] “Internet Protocol,” RFC 791, Internet Engineering Task Force, Sept. 1981.

[3] A. Bridgwater, “What is a digital asset?,” Computerweekly.com. 2014. [Online]. Available: http://www.computerweekly.com/blogs/cwdn/2013/09/what-is-a-digital-asset.html. [Accessed on Aug. 7, 2023]

[4] J. Pohle and T. Thiel, "Digital sovereignty," Internet Policy Review, vol. 9, no. 4, 2020.

[5] S. Deering and R. Hinden, “Internet Protocol, Version 6 (IPv6) Specification,” RFC 8200, Internet Engineering Task Force, July, 2017.

[6] “Internet,” Merriam-Webster.com Dictionary, Merriam-Webster, [Online]. Available: https://www.merriam-webster.com/dictionary/Internet. [Accessed on Mar. 7, 2023].

[7] 3GPP, “5G System Overview,” [Online]. Available: https://www.3gpp.org/technologies/5g-system-overview. [Accessed on Mar. 7, 2023].

[8] “Internet of Things,” Encyclopedia Britannica, Encyclopedia Britannica, Inc., [Online]. Available: https://www.britannica.com/science/Internet-of-Things. [Accessed on Mar. 7, 2023].

[9] 3GPP, “Solutions for NR to support Non-Terrestrial Networks (NTN),” TS 38.821. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.821/38821-g10.zip.

[10] “Internet Assigned Numbers Authority (IANA),” [Online]. Available: https://www.iana.org/. [Accessed on Mar. 7, 2023].

[11] APNIC, “Understanding address management hierarchy,” [Online]. Available: https://www.apnic.net/manage-ip/manage-resources/address-management-objectives-2/address-management-objectives/ [Accessed on Mar. 7, 2023].

[12] “Autonomous System (AS) Numbers,” Internet Assigned Numbers Authority. [Online]. Available: https://www.iana.org/assignments/as-numbers/as-numbers.xhtml. [Accessed on Mar. 7, 2023].

[13] Y. Rekhter, T. Li, and S. Hares, “A Border Gateway Protocol 4 (BGP-4),” RFC 4271, Internet Engineering Task Force, Jan. 2006.

[14] “Domain Name Services,” Internet Assigned Numbers Authority. [Online]. Available: https://www.iana.org/domains. [Accessed on Mar. 7, 2023].

[15] “Root_Zone Management,” Internet Assigned Numbers Authority. [Online]. Available: https://www.iana.org/domains/root. [Accessed on Mar. 7, 2023].

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[16] "Root Servers," Internet Assigned Numbers Authority. [Online]. Available: https://www.iana.org/domains/root/servers [Accessed on Mar. 7, 2023].

[17] T. Narten, E. Nordmark, and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)," RFC 4861, Internet Engineering Task Force, Sept. 2007.

[18] T. Mrugalski et al., "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)," RFC 8415, Internet Engineering Task Force, Nov. 2018.

[19] C. Hopps, "Routing IPv6 with IS-IS," RFC 5308, Internet Engineering Task Force, Oct. 2008.

[20] R. Coltun, D. Ferguson, J. Moy, and A. Lindem, "OSPF for IPv6," RFC 5340, Internet Engineering Task Force, July 2008.

[21] E. Rosen, A. Viswanathan and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, Internet Engineering Task Force, Jan. 2001.

[22] Daniel O. Awduche et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels," RFC 3209, Internet Engineering Task Force, Dec. 2001

[23] L. Ginsberg, B. Decraene, S. Litkowski, and R. Shakir, "Segment Routing Architecture,". RFC 8402, Internet Engineering Task Force, July 2018.

[24] J. Postel, "Transmission Control Protocol," RFC 793, Internet Engineering Task Force, Sept. 1981.

[25] J. Postel, "User Datagram Protocol," RFC 768, Internet Engineering Task Force, Aug. 1980.

[26] J. Iyengar and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport," RFC 9000, Internet Engineering Task Force, May 2021.

[27] E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, Internet Engineering Task Force, August 2018.

[28] E. Rescorla and A. Schiffman, "The Secure HyperText Transfer Protocol", RFC 2660, Internet Engineering Task Force, August 1999.

[29] J. Mccauley, S. Shenker, and G. Varghese, "Extracting the Essential Simplicity of the Internet," Communications of the ACM, vol. 66, no. 2, pp. 64-74, February 2023.

[30] S. Deering, R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," RFC 2460, Internet Engineering Task Force, Dec. 1998.

[31] S.D. Schoen, J. Gilmore, and D. Täht, "IETF Will Continue Maintaining IPv4, " draft-schoen-intarea-ietf-maintaining-ipv4, Internet Engineering Task Force, Sept. 2022.

[32] J. Day, Patterns in Network Architecture: A Return to Fundamentals, Prentice Hall, 2008.

[33] X. Zhang et al., "SCION: Scalability, Control, and Isolation on Next-Generation Networks," *2011 IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 2011, pp. 212-227, doi: 10.1109/SP.2011.45.

[34] S. Jiang, S. Yan, L. Geng, C. Cao, and H. Xu, "New IP, Shaping Future Network: Propose to initiate the discussion of strategy transformation for ITU-T", TSAG C-83, Sept. 2019.

[35] K. Omar, "Internet Protocol version 10 (IPv10) Specification," draft-omar-ipv10, Internet Engineering Task Force, Sept. 2017.

[36] H. Balakrishnan et al., "Revitalizing the public internet by making it extensible," ACM SIGCOMM Computer Communication Review, vol. 51, no. 2, pp. 18–24, April 2021.

[37] International Computer Science Institute, University of California, Berkeley, "An Extensible Internet for Science Applications and Beyond," [Online]. Available: https://www.icsi.berkeley.edu/icsi/projects/extensible-internet/science-applications.

[38] T. Eckert, L. Han, C. Westphal, and R. Li, "An Overview of Technical Developments and Advancements for the Future of Networking," ITU Journal on Future and Evolving Technologies, vol. 3, no. 3, December 2022.

[39] "Host Identity Protocol Working Group," IETF, [Online] Available: https://datatracker.ietf.org/wg/hip/about/.

[40] R. Moskowitz and P. Nikander, "Host Identity Protocol (HIP) Architecture," RFC 4423, IETF, [Online] Available: https://www.rfc-editor.org/rfc/rfc4423.txt.

[41] S. Frankel and S. Krishnan, "IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap," RFC 6071, IETF, [Online] Available: https://www.rfc-editor.org/rfc/rfc6071.txt.

[42] "Understanding Denial-of-Service Attacks," Cyber Security and Infrastucture Security Agency, U.S.A, [Online]. Available: https://www.cisa.gov/news-events/news/understanding-denial-service-attacks.

[43] "Understanding and Responding to Distributed Denial-of-Service Attacks," Cybersecurity and Infrastructure Security Agency, U.S.A, [Online]. Available: https://www.cisa.gov/sites/default/files/publications/understanding-and-responding-to-ddos-attacks_508c.pdf.

[44] Nick Merrill, "Taiwan & the internet during world war," [Online]. Available: https://www.else.how/p/taiwan-and-the-internet-during-world.

[45] "E.164 : The international public telecommunication numbering plan," International Telecommunication Union. [Online]. Available: https://www.itu.int/rec/T-REC-E.164/. [Accessed on Mar. 7, 2023].

[46] W. J. Drake, V. G. Cerf, and W. Kleinwächter, "Internet Fragmentation: An Overview," [Online]. Available: https://www3.weforum.org/docs/WEF_FII_Internet_Fragmentation_An_Overview_2016.pdf. [Accessed on Mar. 7, 2023].

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Zero-Touch-Design Information-Centric Wireless Sensor Networking with Availablity Assurance

Shintaro Mori
Department of Electronics Engineering and Computer Science
Fukuoka University
8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan
e-mail: smori@fukuoka-u.ac.jp

***Abstract*—This paper presents a novel zero-touch-design information-centric wireless sensor network for smart-city applications. To promote self-growing in an autonomous-distributed environment, the proposed scheme adopts zero-touch technology implemented using micro-operators and micro-service providers with a focus on the lower layers where sensor nodes join the network. The scheme also aims to improve the overall availability by utilizing proxy caching and fragmented data management schemes. Computer simulations revealed that, for large-size sensing data (e.g., rich sensing data), the proposed scheme performs better when used with millimeter-wavelength band wireless networks. The results also showed that the scheme is effective in terms of availability and energy consumption. This study is a part of our ongoing research on the development of an ecosystem that enables a smart-city-as-a-service platform, where we are currently focused on the development and experimental trials through on-site testing.**

Keywords: Information-centric wireless sensor networks; Zero-touch-design; Availablity assuarance; Smart-city-as-a-service platform

I. INTRODUCTION

The Internet of things (IoT) has stimulated new trends and empowered innovative new developments in smart devices. The deployment of such devices at distributed locations is a typical scenario in smart-city applications. Wireless sensor networks (WSNs) are an elemental technology in this regard, and they require rapid deployment, initial configuration, and sensing-data provisioning, all of which remain challenging. Moreover, in next-generation wireless networks, such as beyond the fifth generation (5G), massive IoT devices will be deployed in a heterogeneous environment across multiple network domains and versatile service slices. Therefore, for scalability and sustainability, the IoT platform must be shifted from a centralized cloud-based framework to an autonomous-decentralized edge-based one that allows access from various end-users and applications ranging from individuals to enterprises or governments [2].

In light of this background, we focus on two key technologies: zero-touch and data-centric. The zero-touch design aims to completely automate the network management process to minimize the initial costs and set up individual execution environments. Ever since a zero-touch-design system was utilized in the first Linux operating system, there have been increased demands for service deployments that are versatile and flexible in cloud-native micro-services. As for the data-centric architecture, information-centric networking (ICN) (e.g., a content-centric network or named-data network) can be utilized to transform the current network world (e.g., the Internet). ICN natively supports functionalities, such as abstraction, naming, and in-network caching, which enables the data to be decoupled from its original location and the security of every piece of data to be adopted in the network layer. Combining ICN with WSNs is suitable for an autonomous-decentralized environment, which yields Information-Centric Wireless Sensor Networking (ICWSN) [3].

In our previous study [4], we investigated an ICWSN-based ecosystem with a blockchain for smart-city applications utilizing a scheme that achieves efficient and reliable caching. We also presented a blueprint for the system design of a zero-touch-design ICWSN on which the actual smart-city application services will be deployed [1]. In the current paper, utilizing these prior studies as a basis, we evolve and expand the ICWSN system with a focus on reliability and availability. We interpret reliability here as a benchmark indicating that the system operates correctly throughout a certain time interval, i.e., that a reliable system can tolerate any error that may occur (fault tolerant). We interpret availability as an indicator that a system is operating correctly and can continue running its functions at any time, i.e., a system with high availability should have undetectable periods of inaccessibility. In the proposed scheme, for reliability, we utilize a micro-operator (μO), where a network node can be joined only after the μO verifies its individual information when it turns on.

For system deployment in actual smart cities, to reduce the effort required for initialization, the first step is to enable automatic participation by the ICWSN. To this end, for zero-touch design, the proposed scheme also uses a micro-service provider (μSP). For the management of device information and application-service types in μOs and μSPs, the proposed scheme utilizes blockchain-based ledgers, as both the ICWSN and blockchain can work together under an autonomous-decentralized network without mutual trust. However, in ICWSNs, the network nodes are typically limited to resources and thus cannot feasibly support the blockchain network. The μOs provide a solution in that the proposed scheme guarantees the trustworthiness of the nodes during an initial process.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Therefore, the data generated by a reliable node can be trusted without needing to receive any additional verification. For this reason, blockchain-based storage for the data no longer requires traditional computation-intensive mining, and the blockchain can simply select alternative consensus schemes, (e.g., proof-of-authority or proof-of-elapsed-time algorithms) instead of the traditional ones [5]. For availability improvement, the proposed scheme utilizes proxy caching and cooperative data management schemes, where the proxy caching scheme transfers the role of responding to the sensing data from relatively low-reliability nodes to more reliable ones, and the cooperative data management is applied among the nodes that assigned the task of the proxy caching scheme.

In the proposed scheme, the protocol stacks are placed on the ICN layer of the wireless local area network (WLAN). Another option would be to directly place them on the datalink and physical layers, but this is not general purpose since it requires the construction of a special protocol suite. In addition, to support rich 3D sensing applications, virtual reality, augmented reality, and mixed reality, etc., in a future innovative society, the WLAN underpinning the proposed scheme should select the mesh network based on two radio-frequency bands, microwave and millimeter-wave (mmWave,) which are respectively specified as IEEE 802.11 ac/ax and IEEE 802.11 ad/ay. In our evaluation and feasibility demonstration of the proposed scheme, we focus on these two radio bands while performing system modeling, computer simulation, testbed development, and a fundamental experiment.

Section II of this paper presents an overview of the traditional technologies related to our work, and Section III describes the proposed scheme. In Section IV, we report the numerical results, and in Section V, we present the experimental results using a hardware-based testbed device. Related works are discussed in Section VI. We conclude in Section VII with a brief summary and mention of future work.

II. Wideband Information-Centric Wireless Networking Technologies

Various wireless communications and network systems have been investigated to meet different requirements regarding sensing-data collection, distribution, security, and privacy. This section presents an overview of ICWSNs and mmWave WLANs for wideband wireless communications.

A. Information-centric wireless sensor network

The ICWSN system deals with individual data as named data. For named data, one of two naming rules can be selected, hierarchical or flat, depending on the situation in which the ICWSN is deployed. When the data are wirelessly forwarded, the network nodes along the routing path cache (copy and store) them in the local storage for further retrieval of the same data. The caching method is generally categorized into on-path caching or off-path caching. On-path caching is an in-network caching scheme in typical ICNs, whereas in off-path caching, the nodes around the routing path also actively cache. Wireless communications are typically provided in a flooding (broadcasting) style, unlike wired networks, and this specific characteristic (i.e., overhearing phenomena) is what enables the off-path caching to be implemented. Thanks to these caching methods, the sensing data can be effectively expanded, making the retrieval accelerative.

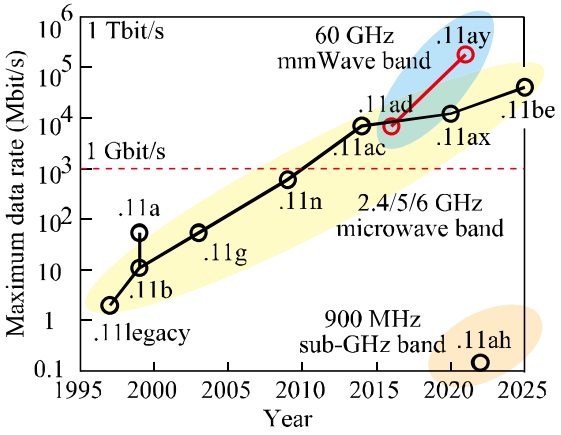


Figure 1. Evolution and history of wireless local area network in IEEE 802.11 specifications.

In the ICN layer, data packets are mainly transferred as interest and response packets. When data retrieval is performed, the requester sends the target packet to interest the network as an acquisition of the data. The node that receives the interest packet and matches the target data for this request plays the role of responder. The responder replies with the response packet encapsulated by the data. Note that, since the data is not distinguished from either the original or the cached data, the interest packet consists of data-requestor information and the properties of the required data. In contrast, the response packet includes the named data with a digital signature and lifetime (data freshness). For data exchange, the interest packets are forwarded to send back the data, and the trace information is recorded in the forwarding information base (FIB), which stores the outgoing interface(s) for each known naming prefix. The intermediate network nodes each have a pending interest table (PIT) that keeps a record of the incoming interfaces and the interest-packet information. The response packet follows the reverse path guided by the PIT entries, and these traversal records are removed while forwarding the data.

B. Wireless communications in mmWave band

The steadily increasing global demand for higher bandwidth has motivated the exploration of the underutilized mmWave spectrum. As shown in Figure 1, this spectrum has a higher potential than the microwave band (e.g., the sub-6-GHz and sub-GHz bands) and can be made available for much larger bandwidth allocation to enable the use of beamforming for greater spatial reuse. Historically, the mmWave has been utilized for fixed wireless access, but standards defined as IEEE 802.11 ad/ay have been commonly provided in the 60-GHz band [6] as well. Radio propagation in the mmWave band is characterized as free-space path loss and precipitation attenuation, which are typically a few dB per kilometer, unless heavy rain causes a significant attenuation (15 dB/km in

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

150 mm/hr). The 60-GHz band has further attenuation due to oxygen and the water concentration of objects, i.e., the 57–64 GHz bands are high-oxygen absorbing bands with 10–15 dB/km, and the moisture it contains (e.g., from leaves and humans) results in particularly strong attenuation across the radio path.

IEEE 802.11 ad is a member of the IEEE 802.11 family and the pioneer in standardizing the 60-GHz band (i.e., in the unlicensed 57–66 GHz bands). IEEE 802.11 ay is the latest standard for expanding the data-transmission rate up to 30 Gbit/s. The 60-GHz band has a greater amount of available bandwidth than all other unlicensed bands (e.g., 2.4, 5, and 6 GHz), but rain and atmosphere make the radio links useless; thus, it is currently utilized only for short-range indoor environments. This should be reconsidered because today's cellular-base-station cell size in urban areas is on the order of 200 m, which offers greater flexibility in the deployment of outdoor scenarios. Terragraph (TG) is an IEEE 802.11 ad/ay-compliant platform developed by Meta (Facebook) and has been successfully deployed by mobile operators and Internet service providers for the backhaul in wireless mesh networks [7].

The physical layer of mmWave WLAN uses a phased-array antenna to execute beamforming between the transmitter and receiver side nodes since the mmWave propagation results in severe path loss and signal attenuation. Note that beamforming technology can concentrate the transmission power and the receiver region over narrow beams. The medium access control (MAC) layer has similar functionalities to the traditional microwave band; however, TG's design only supports the single carrier PHY mode and 12 types of modulation and coding scheme (MCS) sets with data rates up to 4.6 Gbit/s.

III. Proposed Scheme

We introduce a receiver-side cooperation design into the ICWSN system to improve availability. The proposed scheme features two key technologies: proxy caching and fragmented data management. After modeling the ICWSN system, we describe the proposed scheme on the basis of the model.

A. Network model

The network structure of the proposed scheme consists of three main network sections: an ICWSN, an edge network, and a cloud network, as shown in Figure 2. The ICWSN includes sensor nodes (SNs) and relay nodes (RNs) as network nodes that are centrally orchestrated by a mobile base station (MBS). The SNs can perform a pull operation to answer the inquiries of other network nodes. Note that, in traditional WSNs, the sensing data is gathered for the cloud servers (push operation), and users retrieve it as needed. The SNs and RNs are both distributed across the smart-city area, and the SNs sense physical values. The sensing data are provided to users and then partially aggregated to the cloud network via the RNs and MBS. The RNs forward the data (as well as the legacy WSNs), and the data are cached in the local storage of the RNs (as well as the typical ICNs) during the data-forwarding process. The MBS is connected to both the ICWSN and the edge network, so it is relatively more resourceful than the SNs

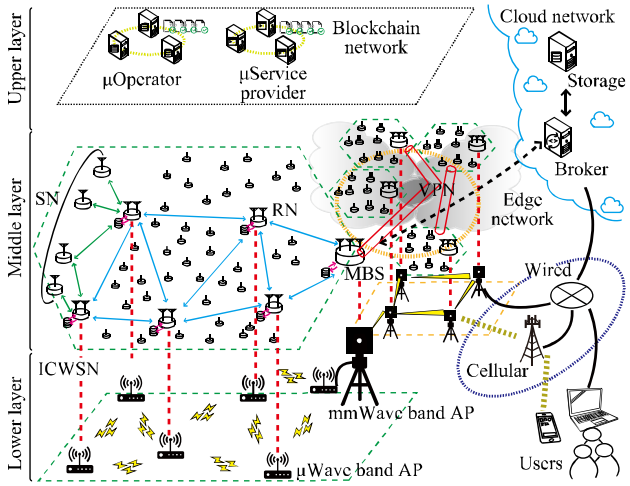


Figure 2. Overview of proposed scheme.

and RNs. The cloud network includes a storage server and a broker. The broker intermediates between the ICWSNs, edge networks, cloud networks, and users, i.e., it exchanges and translates the sensing data and the control messages between them as a gateway. This functionality enables interoperability between multiple regional ICWSNs and provides global scalability. The storage server stores and provides the (copied) partial caching data. Finally, the users are those who consume and obtain the data.

The proposed scheme utilizes two service providers to overlay the physical network for reliability and a zero-touch design: a micro-operator (µO) and a micro-service provider (µSP). The µO verifies whether the SNs can join the ICWSN and provides them with the required connectivity. The information provided here comprises either the authentication that an SN is a proper member of the ICWSN or the network construction settings for wireless transmissions. The former information is provided by the µO, and the latter is managed using the FIB in the RNs, MBS, and broker. The µSP provides the application service and its setting information for the registered SNs to perform as a specified actuator in the ICWSN. Namely, when an SN device is turned on, it sends a registration request to the µO and establishes a secure Virtual Private Network (VPN) link if approved. An ICWSN with a VPN implements the orchestration of distant ICWSNs, terminal fixation at the datalink layer, and secure data exchanges. After joining the network, the SN downloads and installs a configuration setting and application software from the µSP. (The detailed procedures regarding the initialization and registration are described in later sections.) In the proposed scheme, the databases to be referenced by µO and µSP should be used as distributed databases, such as a blockchain. This design principle ensures scalability across interregional networks with different governments, operators, and providers, which is effective in the smart-city scenario assumed in this paper.

As shown in Figure 2, as a network architecture, the proposed scheme uses a mesh network with IEEE 802.11 WLANs in the microwave and mmWave bands in the lower

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

layer. Since the ICWSN is constructed using WLANs, the ICWSN and edge network are in a closed local area network; meanwhile, we assume these networks can access the Internet via the wired or cellular network for global connection. The SNs, RNs, and MBSs can be virtually placed on the physical network in the middle layer, but they should also be placed on the WLAN access points for practical reasons related to location, power supply, and wireless features. The µOs and µSPs are deployed in the upper layer as a functionality of the overlay network technology.

B. Proxy caching scheme

One of the problems with ICWSNs is that SNs can suddenly disappear due to a lack of energy supply or the failure of a (cheap) device, and their turnover (including participating and withdrawing) is rapid because of node mobility. We overcome this problem by utilizing a proxy caching scheme, which ensures availability even in this situation.

An ICWSN consists of an MBS, several RNs, and several SNs, which can be expressed as a set of $\{\mathcal{R}, \mathcal{S}\}$. The set of RNs and SNs are respectively represented as $\mathcal{R} \triangleq \{r_0, r_1, r_2, \cdots, r_{|\mathcal{R}|}\}$ and $\mathcal{S} \triangleq \{s_1, s_2, \cdots, s_{|\mathcal{S}|}\}$. Note that r_0 is the MBS and the others are RNs, and the operator $|\mathcal{X}|$ is the number of elements in the set of $\mathcal{X}$. In the proxy caching scheme, the RNs around an SN are assigned as the SN's alternative nodes, named Proxy RNs (PRNs). The set of PRNs $\hat{\mathcal{R}}^{s_i}$ assigned to $\forall s_i$ $(\in \mathcal{S}, i = 1,2,\cdots,|\mathcal{S}|)$ is given by

$$\hat{\mathcal{R}}^{s_i} \triangleq \Omega(\mathcal{R}|s_i) = \{\hat{r}_1^{si}, \hat{r}_2^{si}, \cdots, \hat{r}_{|\hat{\mathcal{R}}^{s_i}|}^{si}\}, \quad (1)$$

where $\Omega(\mathcal{R}|s)$ is a function that lists up the set of RNs in which the members of $\mathcal{R}$ can be connected to s. In addition, we assume that the SN does not belong to multiple ICWSNs, i.e., there is no roaming across multiple µOs. In proxy-caching-enabled data retrieval, if s_i does not respond to the data-query request for any reason, one of the $\hat{\mathcal{R}}^{s_i}$ answers instead of s_i. The privilege of each $\hat{\mathcal{R}}^{s_i}$ is prioritized in accordance with its index, from small to large; specifically, $\hat{r}_1^{si}$ is called the primary PRN. The priority among $\hat{\mathcal{R}}^{x_i}$ is determined based on the wireless channel condition when the SN is initially registered to the ICWSN or should be re-registered when it moves.

C. Data fragmentation and its cooperative management

For the data fragmentation and management scheme, let $X^{s_i,t}$ denote the data generated by s_i at time t. The expired data, $X^{s_i,t+t_0}$, no longer has any value, where t_0 is the period of valid data. If the data size is large, it is divided into smaller chunks. Let $\delta^{s_i t}$ denote the number of fragmentations for $X^{s_i,t}$; then, this method can be expressed as

$$X^{s_i,t} = x_1^{s_i,t} \oplus x_2^{s_i,t} \oplus \cdots \oplus x_{\delta^{s_i,t}}^{s_i,t}, \quad (2)$$

where the operator $\oplus$ is a bit-by-bit combining. The PRNs cooperatively deal with the divided data and not only perform data backup but also reduce the cost of unnecessary stockholding when the data becomes worthless.

When not all data is complete during alternative data retrieval, the primary PRN gathers the partial loss of chunks from the other PRNs and compiles them. Among PRNs, the primary PRN is most likely to store the divided data, and the amount of available data will be limited in accordance with the size of its index. In some cases, however, only a certain chunk may be lost, which can be dealt with by using the scheme of erasure code and cooperative communications [8]. Using this technique, even if not all the data is complete (i.e., if some of them are lacking), the original data can be fully restored based on the incomplete divided data.

D. Procedure of proposed scheme

In this section, we present the signal processing procedure of the proposed scheme. When a new SN is appended to the ICWSN, it must be identified and its information recorded on the member nodes of the ICWSN. As shown in Figure 3(a), the SN sends a registration request to the neighbor RNs, and the message is then forwarded between RNs and delivered to the µO via the MBS. Note that the SN selects a primary PRN from among the available PRNs here. The µO inquires and verifies the SN's identification to the blockchain network. If the SN is officially approved and activated, the µO sends back the message of complete registration. In this process, the MBS registers the SN as a member of the ICWSN, and the PRNs know that it is the SN to which they should provide a proxy caching mechanism. In addition, it is properly updated in the FIB of the MBS and the RNs on the traced-back routing path. After the SN has been registered, it requests a pre-defined execution (actuation) from the µSP and downloads the application (and its configuration) to enable a zero-touch initialization. Note that a kind of image file is downloaded data for the virtual operating system (e.g., Docker).

When the SN moves, it takes over the task of the proxy caching scheme for the new PRNs, which includes selecting and registering them. As shown in Figure 3(b), the SN sends a withdrawal request to the current PRNs, but the primary PRN temporarily keeps this request on hold, and the SN simultaneously sends a reconstruction request to the neighboring RNs in a new location. A new primary PRN is selected and broadcasts the decision to the ICWSN. On the basis of this notification, the old primary PRN accepts the withdrawal request and removes its task from the old PRNs. In addition, the new and old PRNs should update their FIB at the same time. Through this process, the set of PRNs is changed from $\hat{\mathcal{R}}^{s_i}$ to $\hat{\mathcal{R}}^{s'i}$ corresponding to the movement from s_i to s_i', which is given by

$$\hat{\mathcal{R}}^{s_i'} \triangleq \Omega(\mathcal{R}|s_i'). \quad (3)$$

The PRNs of $\{\hat{\mathcal{R}}^{s_i} \cup \hat{\mathcal{R}}^{s_i'} - \hat{\mathcal{R}}^{s_i'}\}$ remove the registered s_i, and the RNs of $\{\hat{\mathcal{R}}^{s_i'} - \hat{\mathcal{R}}^{s_i} \cap \hat{\mathcal{R}}^{s_i'}\}$ newly register it. If the user moves to a different ICWSN, it is necessary to obtain the wireless-connection information from the µO and re-install and reconfigure the application service from the µSP.

If the radio channel is temporarily degraded (e.g., in the case of not receiving the data continuously), it is expected to recover over time, but since the reason typically originates from a failure of the SN device (e.g., a lack of battery supply,

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

continuous busy status, or permanent poor radio condition), the RNs should detect the cause and reassign the PRNs accordingly. As shown in Figure 3(c), the primary PRN initiates and sends the diagnosis message to the other PRNs. If the PRNs without a primary PRN have also not received the data, the result is reported to the μO, and then the appropriate action is conducted, such as notifying the manager to repair the SN device. When one of the PRNs has received the data, i.e., the primary PRN's radio channel is permanently worse, the secondary PRN takes over the primary PRN's task: specifically, the primary PRN sends the change of primary assignment message to the secondary PRN to switch their duties. If the secondary PRN suffers from a similar situation, it can be replaced by the third PRN, thus maintaining an optimal primary PRN selection.

If the data-requestor cannot directly access the ICWSN, (e.g., if a user cannot access the ICWSN via the Internet), the interest packet is alternatively sent from the broker. Note that users are not limited to human users, they also include the machine that periodically retrieves the collected data. As shown in Figure 3(d), if the required data have been cached in the cloud network, MBS, or RNs close to the MBS, the responder node replies with the cached data; otherwise, the primary PRN answers.

E. Formulation of proposed scheme's availability

In this section, we formulate the effect of the proposed scheme in terms of availability. When introducing system reliability engineering to the network research field, we can define availability as the probability of a data-retrieval request being successfully answered. The availability of the system can be calculated based on the summation of the network node's availability and the outage probability of wireless networks. The network node's availability A is calculated based on

$$A = T_{\mathrm{MTBF}} / (T_{\mathrm{MTBF}} + T_{\mathrm{MTTR}}), \quad (4)$$

where T_{MTBF} is the mean time between failures and T_{MTTR} is the mean time to repair. As for the outage probability of the wireless link, p_{o}, it is determined based on the radio-frequency band and radio-propagation environment, which will be modeled and illustrated in the next section.

Let A_{SN}, A_{RN}, and A_{MBS} denote the availability of SN, RN, and MBS, respectively. The RNs located close to the MBS have a greater effect on the availability of the overall system due to not only hardware failure but also network traffic (and its congestion). To simplify our analysis here, we ignore this effect, i.e., we assume the average availability, $\bar{A}_{\mathrm{RN}}$, for all RNs. For the same reason, we use the average outage probability, $\bar{p}_{\mathrm{o}}$, instead of p_{o} for time-varying channel conditions. Let A_{conv} and A_{prop} denote the availability of the overall conventional scheme and the proposed scheme, respectively, which can be calculated based on

$$A_{\mathrm{conv}} = (1-\bar{p}_{\mathrm{o}})A_{\mathrm{SN}} \cdot A_{\mathrm{MBS}} \cdot \prod_{n=1}^{N_{si}}(1-\bar{p}_{\mathrm{o}})\, \bar{A}_{\mathrm{RN}} \quad (5)$$

and

$$A_{\mathrm{prop}} = A_{\mathrm{PRN}}^{s_i} \cdot A_{\mathrm{MBS}} \cdot \prod_{n=1}^{N_{s_i}-1}(1-\bar{p}_{\mathrm{o}})\, \bar{A}_{\mathrm{RN}}, \quad (6)$$

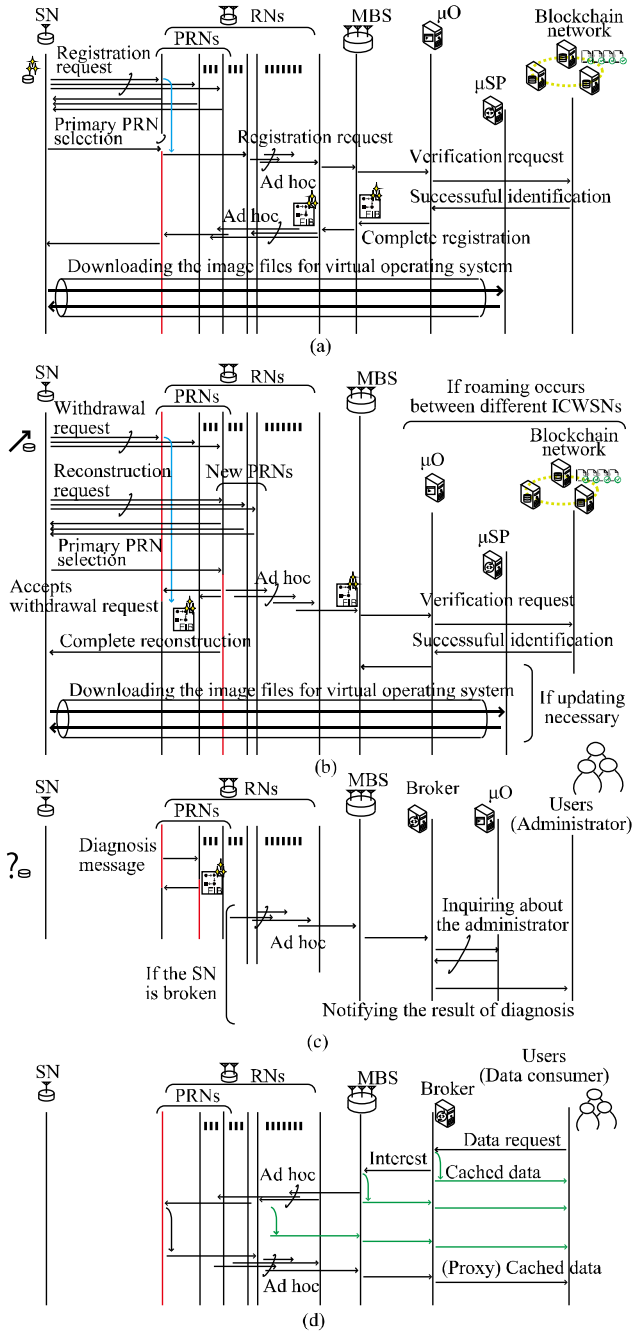


Figure 3. Procedure of the proposed scheme: (a) an SN is newly appended to the ICWSN, (b) the SN moves and changes its PRNs, (c) the SN's data does not reach the PRNs, and (d) the data is retrieved for the ICWSN from the users (data consumers).

where N_{s_i} is the number of hops from the primary PRN of s_i to the MBS, and $A_{\mathrm{PRN}}^{s_i}$ is the availability of the PRNs for s_i, which is given by

$$A_{\mathrm{PRN}}^{s_i} \triangleq 1 - [\, 1 - (1-\bar{p}_{\mathrm{o}})\bar{A}_{\mathrm{RN}} \,]^{|\hat{s}[s_i]|}. \quad (7)$$

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

By comparing (5) and (6), the availability section of the SN in the first term is replaced by a parallel RN, which leads to an availability improvement thanks to the proxy caching. This is because, the relationship between them is $A_{\mathrm{RN}} > A_{\mathrm{SN}}$, since the SNs are generally cheap and massively spread around in the observation area. At the same time, we need to ensure that A_{MBS} has a smaller failure rate, since the MBS is the single point of failure in a physical ICWSN and the proposed mechanism cannot change this. We adopt a high-reliability (industrial use) hardware device as a testbed demonstration in order to mitigate this issue, as discussed in the later section. It should therefore have the highest availability compared to all other nodes, i.e., $A_{\mathrm{MBS}} \gg A_{\mathrm{RN}} > A_{\mathrm{SN}}$.

F. Modeling of wireless communications

Radio propagation in the microwave and mmWave bands is typically characterized as free-space path loss. This can generally be represented as

$$L^{\propto d} = 20\log_{10}(\lambda/4\pi d_0) + 20\log_{10}(d/d_0) \ + \chi_{\mathrm{s}}\ (\mathrm{dB}), \quad (8)$$

where λ (= $300/f_{\mathrm{c}}$ MHz) is the wavelength of the carrier frequency f_{c}, d is the distance between the transmitter- and receiver-side nodes, and d_0 is the closed-in free-space reference distance (typically set to 100 m). χ_{s} is a shadowing variation, i.e., it is a random variable with a Gaussian distribution. Using the practical-experiment-based radio propagation model, (8) can be rewritten as

$$L^{\propto d} = \alpha + \beta \cdot 10\log(d) + \chi_{\mathrm{s}}\ (\mathrm{dB}), \quad (9)$$

where α and β are decided based on the individual radio propagation model. For example, we select the model of Erceg et al. [9] for the link between the nodes on the ground and the model of Amorim et al. [10] for the link between the ground- and air-nodes, as these models are formulated based on the experimental measurements for their respective practical scenarios. Specifically, in Erceg's model, in particular, β can be given by

$$\beta = (a - bh + c/h) + \varepsilon \cdot z, \quad (10)$$

where h denotes the antenna height, z is a random variable with the Gaussian distribution of $\mathcal{N}(0,1)$, and a, b, c, and ε are constant values depending on the surrounding environment [9]. In Amorim's model, α and β are constantly given depending on the UAV altitude [10].

In the microwave band, the link budget, i.e., the relationship between the transmission power P_{TX} and received signal strength $P_{\mathrm{RX}}^{\mathrm{microwave}}$, is represented by

$$P_{\mathrm{RX}}^{\mathrm{microwave}} = P_{\mathrm{TX}} - L^{\propto d} + G_{\mathrm{ANT}} - L_{\mathrm{DEV}}\ (\mathrm{dB}), \quad (11)$$

where G_{ANT} and L_{DEV} are the antenna gain and device loss, on both the transmitter- and receiver-side nodes. In contrast, the radio propagation in the mmWave band is characterized as not only free-space path loss but also precipitation attenuation, which is typically a few decibels per kilometer, unless heavy rain. The moisture is influenced by the presence of natural materials (e.g., leaves, humans), which results in particularly strong attenuation across the radio path. Therefore, the received signal strength in the mmWave is given by

$$P_{\mathrm{RX}}^{\mathrm{mmWave}} = P_{\mathrm{TX}} - L^{\propto d} + G_{\mathrm{ANT}} - L_{\mathrm{DEV}} - L_{\mathrm{RAIN}} - L_{\mathrm{O}^2} - L_{\mathrm{H_2O}}\ (\mathrm{dB}) \quad (12)$$

where L_{RAIN}, L_{O^2}, and $L_{\mathrm{H_2O}}$ are atmosphere attenuation due to rain, oxygen, and natural moisture materials, respectively.

The statistical distribution of the outage probability under the condition that P_{RX} exceeds the desired signal power $P_{\min}$ can thus be given by

$$p_{\mathrm{o}} = \Pr\left(P_{RX} > P_{\min}\right) = 1 - Q\left(\frac{P_{\mathrm{RX}} - L^{\propto d}}{S_{\sigma^2}}\right), \quad (13)$$

where $Q(x)$ is the Gaussian Q function defined as

$$Q(y) \triangleq \int_{\xi}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\xi^2}{2}} \mathrm{d}\xi\,. \quad (14)$$

When p_{o} is simply obtained from the received power, (13) suffices; otherwise, in practice, the signal-to-noise ratio (SNR) should be considered in accordance with the modulation method and additive noise. In this case, (13) is rewritten as

$$p_{\mathrm{o}} = \Pr(\gamma_s < \gamma_0) = \int_0^{\gamma_0} \frac{1}{\gamma_s} e^{-\gamma/\gamma_s} \mathrm{d}\gamma, \quad (15)$$

where γ_s is the received SNR and γ_0 is the required SNR.

IV. Computer Simulation

In this section, we conduct computer simulations to clarify the conditions under which the proposed system can perform, identify numerical examples of availability, and evaluate its power consumption for deployment in a smart city.

A. Simulation environment

The computer simulations were implemented in C++ language on a PC (Windows 10 OS, Core i5 2.9 GHz, 16 GB RAM). Assuming an experimental network composed of 1,000 SNs, ten RNs, and an MBS, we implemented a scenario in which all SNs send the sensing data to the MBS via the PRNs. In the simulation, these nodes are deployed in a 1-km^2 area. The communication range of the nodes is equal in distance and all of them have the same outage probability.

B. Conditions under which the proposed system performs

Let ℓ denote the distance between the SN and RN and ρ denote the number of randomly distributed RNs per km^2. Figure 4 shows the relationship between the number of SNs that do not have PRNs in the case where ρ = 50, 100, 150, 200, and 250. As ℓ increases, the percentage of SNs with no PRNs decreases because the RNs that come under the coverage of the SN increase. Here, ℓ should be kept as small as possible because increasing it would require additional energy consumption for wireless communications. Note that, according to Shannon's information theory, the communication capacity is linearly increased based on the radio bandwidth and logarithmically increased based on the

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

SNR, and thus more energy consumption is necessary to expand the communication distance of ℓ. In addition, for larger ρ, the area assigned to the RNs as the PRNs becomes smaller, and thus the percentage also decreases. Since a larger ρ requires the placement of more RNs with higher functionalities (i.e., more expensive hardware equipment) compared to SN, ρ should be reduced from the viewpoint of cost reduction. Consequently, we set ℓ = 200 m, 150 m, 120 m, 100 m, and 80 m for ρ = 50, 100, 150, 200, and 250, respectively, as the minimum communication distance under no SNs without having PRNs. The computer simulation results will be evaluated on the basis of these parameters.

For data retrievals, Figure 5 shows the probability of successful data acquisition for the average outage probability of the wireless link $\bar{p}_o$. Figure 5(a) shows the case where the ICWSN responds to the proposed method of PRN cooperation, while (b) shows the case where only PPRN responds. As we can see, the curves in the cases of different ρ overlap and the evaluation results do not differ, which indicates that the data retrievals are not affected if a sufficient ℓ is assured and determined for ρ. Therefore, the proposed scheme is available (scalable) if appropriate parameters are given for the network scale for the pair of ρ and ℓ that makes up the parameters determined depending on the observation-area environment. As shown in Figure 5(a), the proposed scheme can reduce the response failures to 5% or less under $\bar{p}_o < 0.4$: specifically, the probability of data-retrieval success is 0.986, 0.978, 0.974, and 0.966 in the case where $\bar{p}_o$ = 0.1, 0.2, 0.3, and 0.4, respectively. From this result and the comparison of Figures 5(a) and (b), it is clear that the proposed scheme can improve the probability of data-retrieval success by 9.87%, 23.79%, 39.22%, and 60.65% in the case where $\bar{p}_o$ = 0.1, 0.2, 0.3, and 0.4, respectively, thanks to a PRN cooperation. However, in the case where $\bar{p}_o > 0.5$, i.e., under the poor wireless channel condition, the curves experience rapid degradation, making it difficult to improve the data-retrieval success even if the proposed scheme is introduced.

Figure 6 shows the mean number of PRNs with which the PPRN requests cooperation to collect data for PRNs. Specifically, this result represents the number of PRNs required to cooperate with each other until all the data divisions are completed. In Figure 6(a), the number of fragmentations is set to 10, while in (b), thanks to the erasure code, we assume that the original data can be recovered if ten of the 15 divided data are complete. With increasing $\bar{p}_o$, more cooperative PRNs are necessary because fewer divided data are cached in the RNs. When we compare Figures 6(a) and (b), it is clear that the use of the erasure code reduces the number of PRNs for data retrieval. This is an advantage when it comes to data acquisition, even though the total amount of sensing data increases during data generation. Specifically, the method with the erasure code improves by 53.7%, 55.2%, 53.3%, and 51.9% when $\bar{p}_o$ = 0.1, 0.2, 0.3, and 0.4, respectively. In particular, as shown in Figure 6(b), in the case where $\bar{p}_o < 0.05$, the number of cooperative PRNs is reduced to 0, which means that the data request can be fully completed by PPRN.

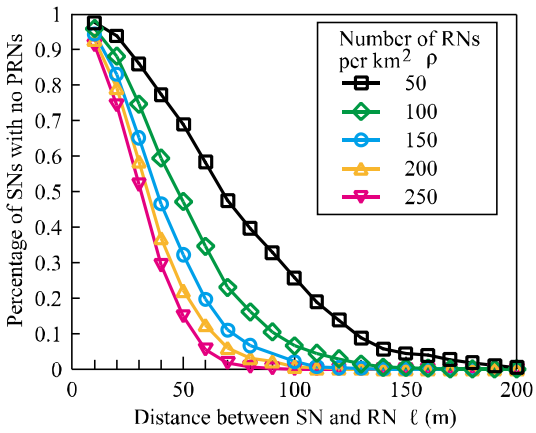


Figure 4. Percentage of SNs with no PRNs vs. distance between SN and RN.

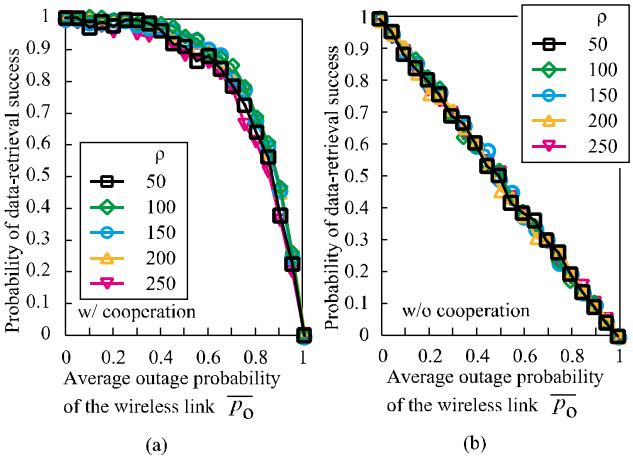


Figure 5. Probability of data-retrieval success vs. average outage probability of the wireless link under (a) proposed PRN cooperative environment and (b) conventional no cooperation PPRN-only environment.

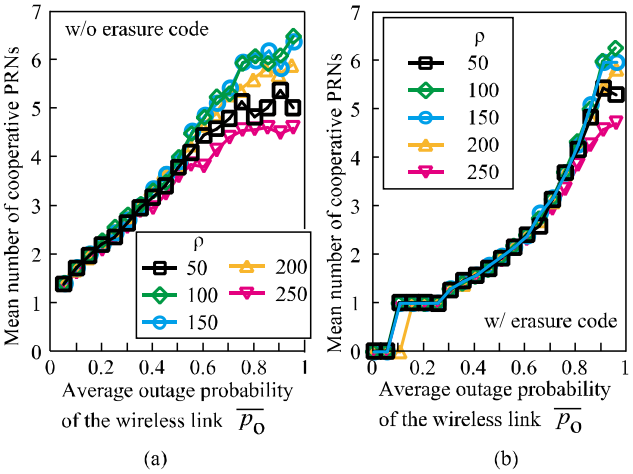


Figure 6. Mean number of cooperative PRNs vs. average outage probablity of the wireless link (a) without erasure code and (b) with erasure code.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

C. Evaluation results: Availability

In this section, we calculate the availability modeled in the previous section using computer simulations. Figure 7 shows a numerical example in the case where ρ = 50 (ℓ = 200) and ρ = 250 (ℓ = 80). These simulations were performed under three conditions where (a) SNs were more unreliable than RNs and MBS, (b) MBS and RNs were equally reliable, and (c) MBS was more reliable than SNs and RNs. Note that, as shown in (6), the overall availability can be improved in the proposed scheme thanks to the proxy caching scheme, i.e., the system uses reliable PRNs instead of unreliable SNs. As a comparable method, the conventional scheme was not used the proxy caching technique. In addition, the MBS is a single point of failure in the ICWSN system.

The results are shown in Figures 7(a–c), where we can see that the proposed scheme improves the overall availability by 5.93%, 14.7%, 22.1%, 31.8%, 36.6%, and 46.7% for ρ = 50 (ℓ = 200) and 5.91%, 14.5%, 23.3%, 31.2%, 38.8%, and 46.0% for ρ = 250 (ℓ = 80) in the case where $\bar{p}_o$ = 0, 0.05, 0.1, 0.15, 0.2, and 0.25, respectively. As shown in Figure 7(a), SN devices were inferior to those of RNs (i.e., A_{RN} = 0.99 and A_{MBS} = 0.999, which was the same condition, but A_{SN} was 0.95 versus 0.9). In terms of overall (ICWSN system) availability, the conventional scheme had a difference (with the cases under that A_{SN} was 0.95 versus 0.9) of 5.56%, while the proposed scheme experienced no degradation.

Figure 7(b) shows the results when the availability of the RN and MBS is in the same condition, with a difference of 0.91% for both the proposed and conventional schemes. This result, which can be explained by the relatively unreliable MBS driving the overall availability, indicates that the proposed scheme cannot ensure the overall availability of the MBS with low reliability. However, as shown in Figure 7(c), in the case where the MBS is more reliable (i.e., A_{MBS} = 0.9999 and 0.99999), the overall availability was only improved by 0.09% and 0.1% compared to the case of A_{MBS} = 0.9999. Therefore, the results in Figure 7(b) suggest that the MBS should be more reliable than RNs and SNs, but the results in (c) indicate that the overall availability does not significantly improve even if the MBS device is over-reliable due to paying much cost, such as high-reliable hardware-device development.

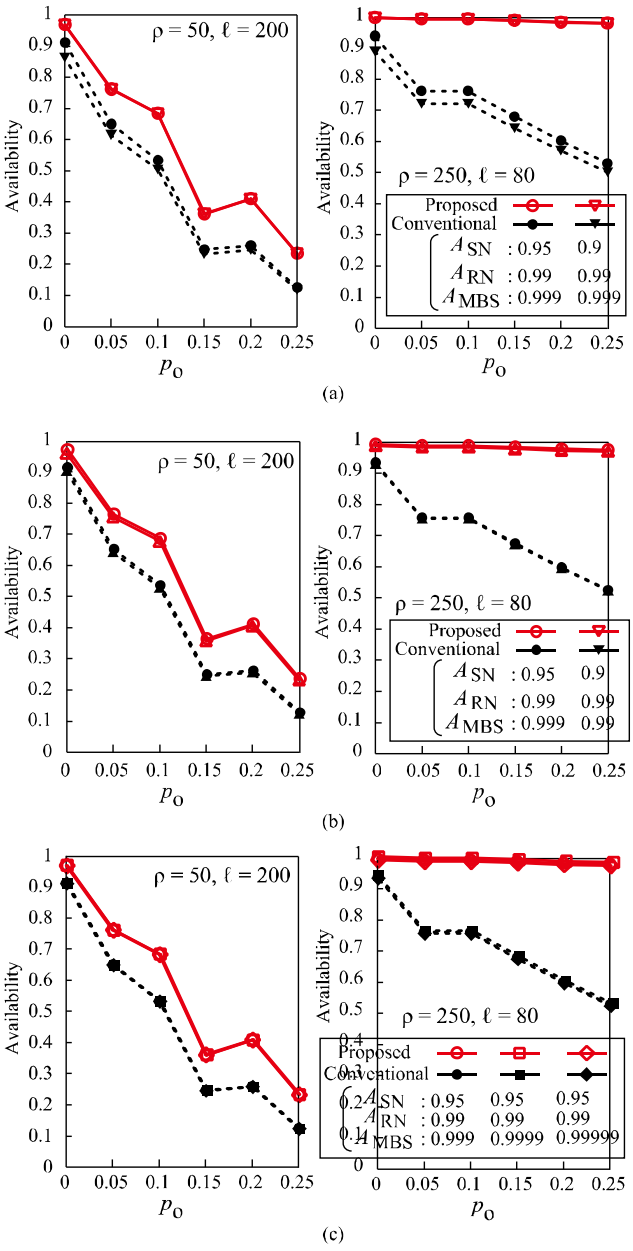


Figure 7. Computer simulation results for availability under three condintions: (a) SNs are unreliable, (b) MBS and RNs are equally reliable, and (c) MBS is more reliable than other nodes.

D. Evaluation results: Energy consumption

The challenge in deploying the proposed scheme is how to ensure a benefit in terms of energy consumption among the SN devices. This is because ICN has a pull-type network design and must always be on standby, and the blockchain also causes energy wastage. Figure 8 shows the computer simulation we ran to investigate the cumulative energy consumption in the conventional scheme (current application-programming-interface-based IoT platform) and the proposed scheme. Note that, when the SN does not execute any process, we assume the conventional scheme supports a sleep state with deep sleep and wake-up functionalities, whereas the proposed scheme waits in the idle state to be ready for data retrieval from any other node (because of the pull-type data acquisition). The energy consumption for each status is based on the actual measured values from our previous study [5].

As shown in Figure 8(a), the proposed scheme can reduce energy consumption by 1.91% if there are no additional requests for data retrieval in most cases of periodic data collection in ordinary situations. Moreover, even if 66 additional data retrievals per day are requested, the proposed scheme can outperform. Next, Figure 8(b) shows the total energy consumption in the ICWSN for 1,000 SNs, with the results converted into the power consumption per node. For these results, the number of data retrieval attempts for each

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

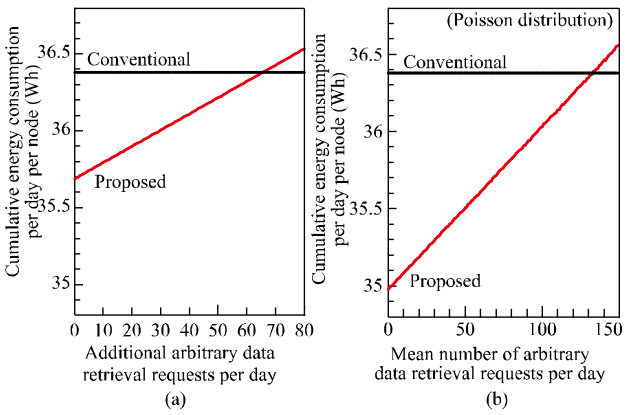


Figure 8. Simulation results for (a) additional data retrieval requests per day vs. cumulative energy consumption and (b) mean number of additional requests according to a Poisson distribution vs. cumulative energy consumption per unit for 1,000 SNs.

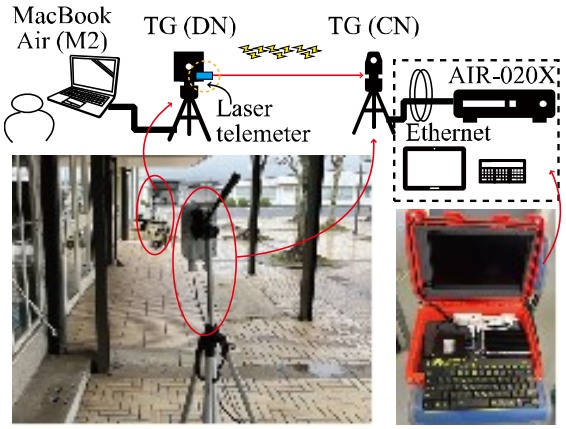


Figure 9. Experiment network and testbed device

TABLE I SPECIFICATIONS OF TG NODES

Spec.	MLTG-DN		MLTG-CN
Size	20×20×20 cm		18×11×4.3 cm
Weight	3.9 kg		1.1 kg
Frequency	f_c = 58.32 GHz (57.0–59.4 GHz)		
Tx power	43 dBm		38 dBm
Antenna	Gain: 28 dBi		Gain: 22 dBi
	Phased array antenna with 64 elements Azimuth range: -45° to +45° Elevation range: -25° to +25°		
LAN	Gigabit Ethernet (1x port)		
MCS no.	Modulation	Rate	Throughput
#9	π/2-QPSK	13/16	2,503 Mbit/s
#12	π/2-16QAM	3/4	4,620 Mbit/s

node was determined by a Poisson distribution, which is a more realistic calculation than the one in Figure 8(a). As we can see, the proposed scheme was able to reduce energy consumption by 3.85% and was advantageous until 138 retrieval attempts.

V. EXPERIMENTAL RESULTS

In this section, we implemented a testbed device and conducted a preliminary evaluation of the network performance, particularly for mmWave band WLANs. Note that we omitted the microwave band WLAN here because it is widely used and its features are well-known. The testbed here demonstrates a part of sensing-data processing and wireless communications for SN, RN, or PRN. Namely, the device will be able to perform the baseline tests for future test field construction and sensor node implementations. The testbed was implemented using an Advantech [11] AIR-020X (Six-core ARM v8.2 CPU, 8-GB RAM, Ubuntu 18.04 with Jet Pack OS), which is embedded in equipment for industrial use and adheres to the form factor of the NVIDIA Jetson; thus, the software and settings can be easily moved from any other prototype platform. The AIR-020X is high-performance for rich 3D sensing data, and we integrated it here with peripheral devices into a rectangular attaché case as a portable unit, as shown in Figure 9. The portable testbed device includes an IoT router, which is used to connect to the Internet via the cellular network for external time synchronization and emergency external control. The electrical power can be supplied from a wall outlet, through the devices can also alternatively be provided from the internal power-supply hub as well as the USB type A and C connections.

For the mmWave distribution networks, we utilized a pair of TGs consisting of distribution nodes (DNs) and client nodes (CNs), the specifications of which are listed in Table I. The TG was designed to construct a wireless mesh network. In particular, the DNs work and provide a backhaul wireless network, while the CNs can provide broadband wireless communications to the end users. In this paper, we evaluate the fundamental characteristics of mmWave communications, i.e., the link between the DN at the end of the backhaul network and the CN to the user terminal. Note that the hardware equipment for demonstrating the link between CN and DN and between DNs is different wireless communication equipment. However, we believe there is no significant difference in wireless communication characteristics.

According to the TG specification [9], a DN provides a mesh network among DNs within 15 hops and is composed of multiple sectors capable of communication in different directions. A CN is typically a one-sector station node that terminates Internet protocol (IP) connectivity for an end-user. The communication protocol between a DN and CN corresponds to the model of access point to station node in IEEE 802.11 ad/ay. To determine the best beam angle for the best SNR, TG has features of periodic beamforming and interference measurement. The MCS set is a combination of a modulation method (specifically BPSK, QPSK, or 16-QAM) and the low-density parity check coding scheme (with code rates of 1/2, 5/8, 3/4, 13/16, and 7/8). TG supports a total of 12 MCSs and typically operates at MCS nos. #12 and #9 (as

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

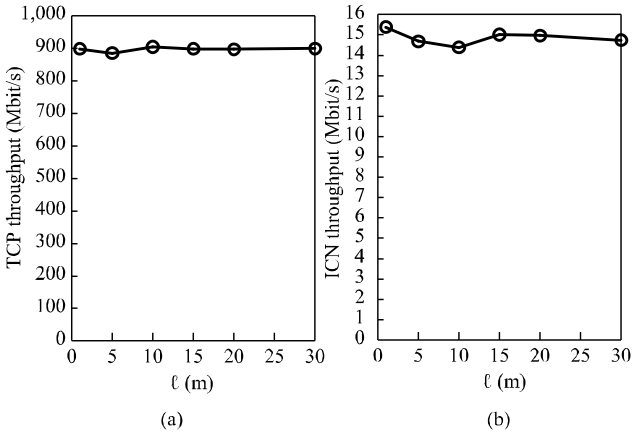


Figure 10. Experimental results for (a) TCP throughput and (b) ICN throughput vs. distance between SN and RN in mmWave WLAN.

shown in Table I) for 250-m-range coverage and achieves a throughput of 1 Gbit/s.

In the experiment, the network nodes are constructed using the implemented portable device and a 13-in MacBook Air. The TG equipment was mounted on a tripod, and both were respectively connected to the subscriber- and publisher-side devices. The horizontal plane of a DN and CN was maintained using a laser telemeter and the distance between the two nodes was measured at the same time.

This experiment was conducted in an outdoor environment, as shown in Figure 9, and there were no objects to interrupt the radio route, although some of the surrounding buildings caused reflections. The radio propagation was dominated by line-of-sight and building-reflected paths. Figure 10 shows the experimental results, including throughput at the transmission control protocol (TCP) layer using iPerf3, a common tool for measuring TCP throughput, and the ICN layer using Cefore [12], a CCNx-based ICN platform. Note that, when the proxy caching scheme is used, there might be a problem of cache inconsistency. In Cefore, the caching data can be managed using database in the daemon process of csmgrd and this concern is not present. As we can see, there was no significant throughput degradation depending on the distance between TGs, unlike that seen with IEEE 802.11 standards in Sub-6-GHz bands. In this experiment, we remove the reasons for the degradation of radio propagation, e.g., radio attenuation by trees and foliage, blocking materials (vehicle and human) between antennas, and multipath fading due to reflected waves. In addition, the mmWave communication must have the beamforming technique to bolster strong straightness with w signal weakness. In fact, the throughput was reduced by up to 50% when a person crosses between nodes, and throughput was reduced by up to 10% when the horizontal plane was not flat.

Since the mmWave band is highly directional (i.e., it has a radio-propagation characteristic similar to that of light), we investigated the degree to which it can be tolerated once a communication path between antennas was established by a beamforming technique. The DN and CN were placed three meters apart from each other, and we conducted the

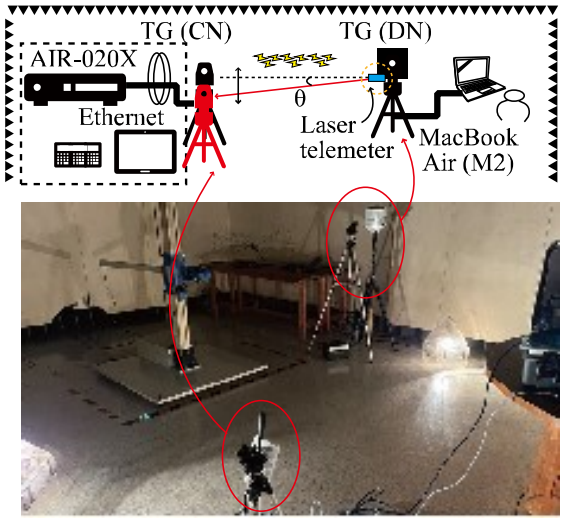


Figure 11. Overview of experimental site in anechoic chamber (shielded room).

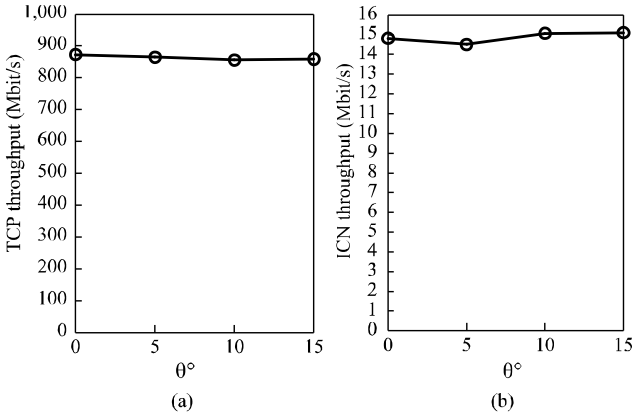


Figure 12. Experimental results for beamforming technique: (a) TCP throughput and (b) ICN throughput vs. degree from the horizontal plane in mmWave WLAN.

experiment in an electromagnetic anechoic chamber that is a shielded to eliminate the effect of other radio signals, including the reflected waves themselves. The DN and CN were maintained on a horizontal axis (determined using a laser telemeter), and the TCP and ICN throughputs were measured when the DN was fixed and the CN was moved to change its angle, as shown in Figure 12. The results showed that the wireless communication was stable and maintained even if the angle between CN and DN was changed several times. Note that, in our latest study [13], we develop and evaluate the testbed devices and test fields to evaluate the effectiveness of the proposed scheme for real WSN implementations.

VI. Related Work

Zero-touch management has become a hot topic in the process for standardization, e.g., by the European Telecommunications Standards Institute (ETSI) Zero-Touch Network and Service Management (ZSM) working group. Sanchez-Navarro et al. [14] provided a novel holographic immersive network management interface that extends the

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

standardized ETSI ZSM reference architecture to enable network administrators to understand real-time automated tasks in a 5G network without human intervention. Boškov et al. [15] proposed a zero-touch solution based on WLAN and Bluetooth technologies that can yield a sufficient performance for provisioning multiple devices without depending on the vendor's proprietary hardware and software. For resource management among nodes including an extended capability regarding 5G network slicing services, Theodorou et al. [16] formed marketplaces to facilitate the exchange service level agreements, where a blockchain was utilized for guaranteeing an untrusted and unreliable node.

For μO and μSP, Togou et al. [17] introduced a distributed blockchain-enabled network slicing framework that enables service and resource providers to dynamically lease resources, thereby ensuring high performances for their end-to-end services. The key component of the framework is its global service provisioning, which provides admission control for incoming service requests along with dynamic resource assignment by means of a blockchain-based bidding system. This is essentially a blockchain-based multi-operator service provisioning for 5G users with Intra and Inter spectrum management among multiple telecom operators. Gorla et al. [18] presented a blockchain-based implementation model for spectrum sharing between operators to minimize spectrum under-utilization in order to enable reliable quality of services. Another study [19] has suggested that the network slice provider will play the role of an intermediate entity between the vertical service provider and the resource provider, which makes a shift from a network-operator-oriented business to a more open system with multiple actors. Today, management and orchestration are considered prime components of the new network management layer [20], and multi-domain orchestration has helped in simplifying infrastructural operations while enabling better scaling and faster deployment of network services. Based on multi-constraint QoS, it fulfills the E2E slice request. Blockchain is also deployed to ensure trustworthiness between different telecom operators, introduce transparency, and automate the fulfillment of service-level agreements through smart contracts.

For mmWave-band communication systems, Rappaport, et al. [21] presented path loss models with directional and omnidirectional antennas based on over 15,000 measured power delay profiles (PDPs) at 28, 38, 60, and 73 GHz bands using wideband channel sounders. Tariq, et al. [22] measured received signal strength and delay spread values for each specified beam combination with massive antenna arrays in both indoor and outdoor scenarios using TG radios. Shkel, et al. [23] presented TG platform for promoting the research in mmWave propagation, systems, and networks. Aslam, et al. [24] measured the radio propagation characteristics for in-street backhaul environments and evaluated along with lay-based simulations. The results indicated that path loss in the urban canyon scenario was observed to be smaller when compared with the residential areas due to the rich number of multipath components. All these studies indicate that mmWave features a high path loss and a high material attenuation.

Sellami, et al. [25] proposed to implement an architecture for a distributed fog caching solution for ICN system, which consisted of sensor sub-networks connected to one or more fog super-nodes that maintained the internal caching policies and interactions with the fog in order to achieve efficient content caching and retrieval. Sukjaimuk, et al. [26] proposed an effective caching and forwarding algorithm for congestion control for ICWSN. The scheme utilized accumulative popularity-based delay transmission time for forwarding strategy and included the consecutive chunks-based segment caching scheme. Zhang, et al. [27] leveraged the machine learning technology to propose an intelligent caching scheme that could automatically adjust the caching nodes' caching parameters for the dynamic network environments. The simulation results showed that the scheme outperformed the existing approaches in terms of the total energy consumption.

VII. Conclusion

In this paper, we presented a zero-touch-design ICWSN to promote self-growing and ensure a reliable sensing-data distribution in which multiple players actively participate and exchange data. A computer simulation was conducted using a testbed of the proposed scheme and TG to investigate the conditions under which it can best perform, its overall availability, and its energy consumption, as the potential waste involved in the use of ICN and blockchain is significant. The results demonstrated the feasibility of our scheme and clarified the radio-propagation characteristics of mmWave band WLANs. In our ongoing research project, which we call the Decentralized Digital Twins' Ecosystem (D2EcoSys), we will further investigate the deployment of this scheme for real smart cities, which is our future work.

Acknowledgment

This work was partly supported by NICT Japan, Grant Number 05601. We are grateful to Dr. Kenji Kanai for his helpful discussions, and Advantech Japan, BeMap, Haft, Panasonic, and TEAD for their help with the on-site experiments.

References

[1] S. Mori, "A study on zero-touch-design information-centric wireless sensor networks," *Proc. IARIA the 22th Int. Conf. Networks (ICN 2023)*, Apr. 2023, pp. 7–9.

[2] H. Chergui, A. Ksentini, L. Blanco, and C. Verikoukis, "Toward zero-touch management and orchestration of massive deployment of network slices in 6G," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 86–93, Feb. 2022.

[3] B. -S. Kim, C. Zhang, S. Mastorakis, M. K. Afzal, and J. Tapolcai, "Guest editorial special issue on information-centric wireless sensor networking (ICWSN) for IoT," *IEEE Internet of Things J.*, vol. 9, no. 2, pp. 844–845, Jan. 2022.

[4] S. Mori, "Information-centric wireless sensor networks for smart-city-as-a service: Concept proposal, testbed development, and fundamental evaluation," *Proc. IEEE Consumer Commun. and Networking Conf. (CCNC 2023)*, Jan. 2023, pp. 945–946, doi: 10.1109/CCNC51644.2023.10060577.

[5] S. Mori, "Secure caching scheme using blockchain for unmanned aerial vehicle-assisted information-centric wireless sensor networks," *J. Signal Process.*, vol. 26, no. 1, pp. 21–31, Jan. 2022.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[6] Y. Ghasempour, C. R. C. M. Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 186–192, Oct. 2017.

[7] A. Nordrum, "Facebook pushes networking tech: The company's Terragraph technology will soon be available in commercial gear," *IEEE Spectrum*, vol. 56, no. 4, pp. 8–9, Mar. 2019.

[8] S. Mori, "Data collection scheme using erasure code and cooperative communication for deployment of smart cities in information-centric wireless sensor networks," *Int. J. Advances in Networks and Services*, vol. 14, no. 3&4, pp. 54–64, Dec. 2021.

[9] V. Erceg et al., "An empirically based path loss model for wireless channels in suburban environment," *IEEE J. Sel. Areas in Commun.*, vol. 17, no. 7, pp. 1205–1211, July 1999.

[10] R. Amorim et al., "Radio channel modeling for UAV communication over cellular networks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 514–517, Aug. 2017.

[11] Advantech, https://advantech.com/ (retrieved: Nov. 2023).

[12] Cefore, https://cefore.net/ (retrieved: Nov. 2023).

[13] S. Mori, "Test-field development for ICWSNs and preliminary evaluation for mmWave-band wireless communications," *Proc. IEEE Consumer Commun. and Networking Conf. (CCNC 2024)*, Jan. 2024. (in press)

[14] I. Sanchez-Navarro, P. Salva-Garcia, Q. Wang, and J. M. A. Calero, "New immersive interface for zero-touch management in 5G networks," *Proc. 2020 IEEE 3rd 5G World Forum (5GWF)*, Sept. 2020, pp. 145–150, doi: 10.1109/5GWF49715.2020.9221116.

[15] I. Boškov, H. Yetgin, M. Vučnik, C. Fortuna, and M. Mohorčič, "Time-to-provision evaluation of IoT devices using automated zero-touch provisioning," *Proc. IEEE GLOBECOM 2020*, Dec. 2020, pp. 1–7, doi: 10.1109/GLOBECOM42002.2020.9348119.

[16] V. Theodorou et al., "Blockchain-based zero touch service assurance in cross-domain network slicing," *Proc. 2021 Joint European Conf. Networks and Commun. & 6G Summit (EuCNC/6G Summit)*, June 2021, pp. 395–400, doi: 10.1109/EuCNC/6GSummit51104.2021.9482602.

[17] M. A. Togou et al., "DBNS: A distributed blockchain-enabled network slicing framework for 5G networks," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 90–96, Nov. 2020.

[18] P. Gorla, V. Chamola, V. Hassija, and N. Ansari, "Blockchain based framework for modeling and evaluating 5G spectrum sharing," *IEEE Network*, vol. 35, no. 2, pp. 229–235, Mar.–Apr. 2021.

[19] B. Nour, A. Ksentini, N. Herbaut, P. A. Frangoudis, and H. Moungla, "A blockchain-based network slice broker for 5G services," *IEEE Networking Lett.*, vol. 1, no. 3, pp. 99–102, Sept. 2019.

[20] V. K. Rathi et al., "A blockchain-enabled multi domain edge computing orchestrator," *IEEE Internet of Things Mag.*, vol. 3, no. 2, pp. 30–36, June 2020.

[21] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sept. 2015.

[22] M. H. Tariq, I. Chondroulis, P. Skartsilas, N. Babu, and C. B. Papadias, "mmWave massive MIMO channel measurements for fixed wireless and smart city applications," *Proc. IEEE 31st Annual Int. Sympo. Personal, Indoor, and Mobile Radio Commun. (PIMRC 2020)*, Sept. 2020, pp. 1–6, doi: 10.1109/PIMRC48278.2020.9217375.

[23] A. Shkel, A. Mehrabani, and J. Kusuma, "A configurable 60GHz phased array platform for multi-link mmWave channel characterization," *Proc. IEEE Int. Conf. Commun. (ICC 2021)*, June 2021, pp. 1–6, doi: 10.1109/ICCWorkshops50388.2021.9473724.

[24] M. Z. Aslam, Y. Corre, J. Belschner, G. S. Arockiaraj, and M. Jäger, "Analysis of 60-GHz in-street backhaul channel measurements and LiDAR ray-based simulations," *Proc. 14th European Conf. Antennas and Propagation (EuCAP)*, Mar. 2020, pp. 1–5, doi: 10.23919/EuCAP48036.2020.9135946.

[25] Y. Sellami, G. Jaber, and A. Lounis, "Distributed fog-based caching solution for content-centric networking in IoT," *Proc. IEEE Consumer Commun. and Networking Conf. (CCNC 2022)*, Jan. 2022, pp. 493–494, doi: 10.1109/CCNC49033.2022.9700561.

[26] R. Sukjaimuk, Q. N. Nguyen, and T. Sato, "An efficient congestion control model utilizing IoT wireless sensors in information-centric networks," *Proc. Joint Int. Conf. Digital Arts, Media, and Technol. with ECTI Northern Sec. Conf. Electrical, Electronics, Comp. and Telecommun. Eng.*, Mar. 2021, pp. 210–213, doi: 10.1109/ECTIDAMTNCON51128.2021.9425753.

[27] Z. Zhang, X. Wei, C. -H. Lung, and Y. Zhao, "iCache: An intelligent caching scheme for dynamic network environments in ICN-Based IoT networks," *IEEE Internet of Things J.*, vol. 10, no. 2, pp. 1787–1799, Jan 2023.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Energy Consumption Minimization in Data Centers for Cloud-RAN

Line M. P. Larsen*†, Simon Friis*, Henrik L. Christiansen† and Sarah Ruepp*
† Department of Mobile Innovation, TDC Net, Copenhagen, Denmark
* Department of Electrical and Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark
Corresponding author: lmph@dtu.dk

***Abstract*—Data centers are to become a vital part of mobile networks when exploring new architectures such as cloud radio access network, which is seen as a more energy efficient alternative to today's mobile network installations. However, the scale of a data center affects its energy efficiency, with larger data centers having more resources for energy-saving measures but at the same time different challenges than those faced by data centers of smaller scale. Hence, this work analyzes and compares energy efficiency of small, medium, and large data centers, and explores the energy minimization opportunities for cloud radio access network data centers. When moving processing from the radio access network to a data center to save energy in the radio access network, the mobile network operator does not want to move the energy consumption from the radio access network to the data center. Hence, energy consumption must be reduced in all segments of the network, and not moved to another segment. In the case study provided, cloud radio access network data centers are categorized as mid-scale and small-scale depending on the size of area they cover due to strict latency requirements of mobile network traffic flow. Thus, even though larger DCs prove to be more energy efficient, other factors, which in the case of mobile networks will be latency, will impact the size of the data center. Hence, this is a major driver to consider how energy consumption can be minimized in small data centers.**

Keywords-Data center; cloud; C-RAN; green RAN; energy efficiency.

I. Introduction

This work is an extended version of [1]. The Information and Communications Technology (ICT) sector, including Data Centers (DCs), communication networks and user devices, is accounted for an estimated up to 6% of global electricity use [2]. Thus, many initiatives aim at reducing the energy consumption of the various different sub-parts of the ICT sector. However, looking into solutions for reducing the energy consumption in one part of the sector should not just move the problem to another part of the sector. The solution must look at the ICT sector from a holistic perspective. A starting point is to trace the solutions and investigate what other implications they will bring. Mobile networks, are widely used communication infrastructure, and looking at their energy consumption, the Radio Access Network (RAN) plays a significant role [3]. One solution envisioned to reduce the energy consumption of the RAN, is to utilize the long time discussed Cloud RAN (C-RAN) architecture [4].

In C-RAN, the mobile network functions are divided into three functional units; the Radio Unit (RU), Distributed Unit (DU) and Centralized Unit (CU). The RU, is located in the antenna mast and contains part of the physical layer functions. Whereas the DU and CU, containing the upper layer baseband processing functions, can be virtualized and located in DCs. A C-RAN installation and a traditional RAN installation are compared in Fig. 1. The figure shows how all baseband processing is located on site in the traditional RAN in top of the figure, where baseband processing is divided into separate DU and CU and moved to DCs in the C-RAN installation. The concept is, to store processing from a number of sites in the same Data Center (DC), where they can share physical resources and exploit their different user movement patterns. This being users gathering in different areas at different times of day [4]. Hence, in order to make the RAN segment of mobile networks more energy efficient, much of the processing is moved into DCs.

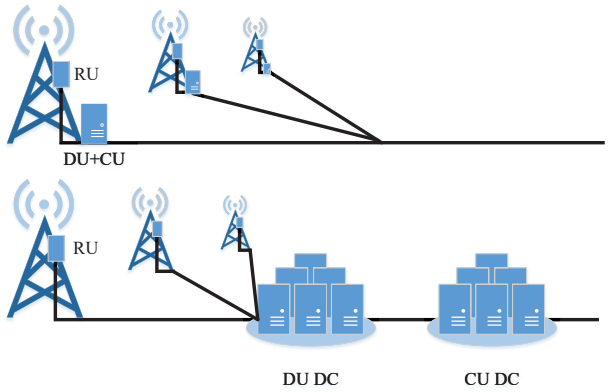


Figure 1. A traditional Radio Access Network (RAN) architecture [top] and a Cloud-RAN (C-RAN) architecture where data processing is moved to Data Centers (DCs) [bottom].

A DC refers to a number of servers located in the same building. Many different types of DCs exist and they are in this work categorized into three different sizes ranging from small, medium to large.

- *Small DCs* are categorized as having less than 1,000 servers, as well as less complex infrastructure, limited storage and thus; consume less power compared to larger DCs.
- *Mid-scale DCs* have a larger number of servers which ranges between 1,000 to 10,000 [5] with more complex infrastructure.
- *Large DCs* are defined as having more than 10,000

Support from Innovation Fund Denmark, through grant no. 1045-00047B and the Nordic University Hub on Industrial IoT, Nordforsk grant agreement no. 86220, is gratefully acknowledged.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

servers with an even more complex infrastructure [5]. Large DCs are typically used by large companies or governments.

Energy efficiency in DCs is a crucial topic of modern DC operations, as it can help to reduce energy costs and environmental impact of the ICT sector. DCs are energy-intensive facilities that consume a large amount of electricity to power servers, storage systems and cooling equipment. The energy consumption of DCs has become an increasing concern for the industry, as well as businesses and organizations that operate these facilities, as DCs are responsible for approximately 1% of global electricity demand [2].

This paper investigates methods for energy efficiency in DCs, with a focus on state-of-the-art technologies and techniques, as well as how and why these are beneficial for DCs of different scale. Furthermore, it is investigated how DCs, handling mobile network data processing in the C-RAN architecture, can become energy optimized. Section II presents other related papers, research projects and features our contribution to the topic. Section III explains the typical components in a DC, to be used in section IV, which goes in-depth with modern and commonly used strategies for minimizing DC energy consumption. Section V elaborates on the use of DCs in C-RAN mobile networks, while section VI discusses what and why some methods are most commonly used in DCs of certain sizes and their potential. Finally, the conclusion closes this paper. A list of acronyms is provided after the conclusion.

II. State of the Art

This study combines two directions of energy efficiency studies, namely DCs and mobile networks. There have been several studies and research conducted on energy efficiency in DCs in recent years. However, numerous surveys regarding energy efficiency in DCs tend to be older than 5 years, such as the work in [6], which presents an overview of energy-aware resource management approaches with focus on basic architecture of cloud DCs and virtualization technology. The survey in [7] investigates the green energy aware power management problem for Megawatt-scale DCs and classifies work that considers renewable energy and/or carbon emission. In [8], authors discuss several state-of-the-art resource management techniques, that claim significant improvement in the energy efficiency and performance of ICT equipment and large-scale computing systems, such as DCs. The work in [9] conduct an in-depth study of the existing literature on DC power modeling, covering more than 200 models. The concept of C-RAN was first mentioned by companies IBM [10] and China Mobile [11] and later explored in numerous surveys including [4], [12] and [13]. However, the following section of related work will focus on research conducted within the latest years.

A. Related work

Recent related work in the field of greening DCs includes [14], where the approaches moving towards green computing are investigated and categorized to help researchers and specialists within cloud computing expand green cloud computing and improve the environment quality. The work in [15], gives a brief overview of the state-of-the-art in green cloud computing. Existing research in the area is examined and categorized into different themes. Furthermore, challenges and opportunities in the field are discussed to provide insights into future directions for research. The paper provides valuable background information and a significant understanding of the current landscape of green cloud computing. In the survey [16], the authors discuss different mechanisms for lowering the power utilization in DCs. The survey provides in-depth details about the various mechanisms that can be employed at the hardware level so that the utilization of energy by component can be reduced. Techniques that can be applied at network, cluster of servers' level along with the various dynamic power management measures that can be employed at the hardware or firmware level and can lead to energy efficient or green DCs are also studied in detail [16]. Table I lists relevant research in the field of energy improved DCs categorized into relevance regarding the different DC sizes.

TABLE I. DATA CENTER REFERENCES BY SIZE

DC Size	References
Large scale	[6] [7] [8] [9] [14] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27]
Mid scale	[9] [14] [16] [18] [21] [22] [23] [24] [25] [27] [28]
Small scale	[14] [16] [18] [21] [22] [23] [24] [25] [27] [28] [29] [30] [31]

C-RAN has recently been surveyed in relation to energy consumption improvements in [3], [32], [33], [34]. Which all highlights the energy saving potential of shutting down equipment not in use. Hence, in DCs, some equipment can be shut down in low traffic periods such as during the night. Another benefit of C-RAN is the opportunity to assign extra capacity where it is needed. In the perspective of users being in residential areas in the morning and evening, and goes to work areas during the day [4]. Thus, current trends in C-RAN research point in the directions of: Load consolidation [35], [36], [37]; coordinated transmission [38], [39]; and with a focus on the transport network [40], [41].

The work in [42] acknowledges the problem of low server utilization when mobile network traffic is moved to DCs, and proposes a framework for a higher utilization of the physical machines. In [43], various methods for DC load balancing was surveyed and compared not only for resource utilization, but also in other parameters such as power usage and reaction time.

The authors of this work have previously addressed the topic of energy minimization of DCs in [1]. Which examined various methods for minimizing DC energy consumption in various sizes of DCs. To the best of our knowledge the combination of C-RAN and energy efficient DCs has not yet been explored.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

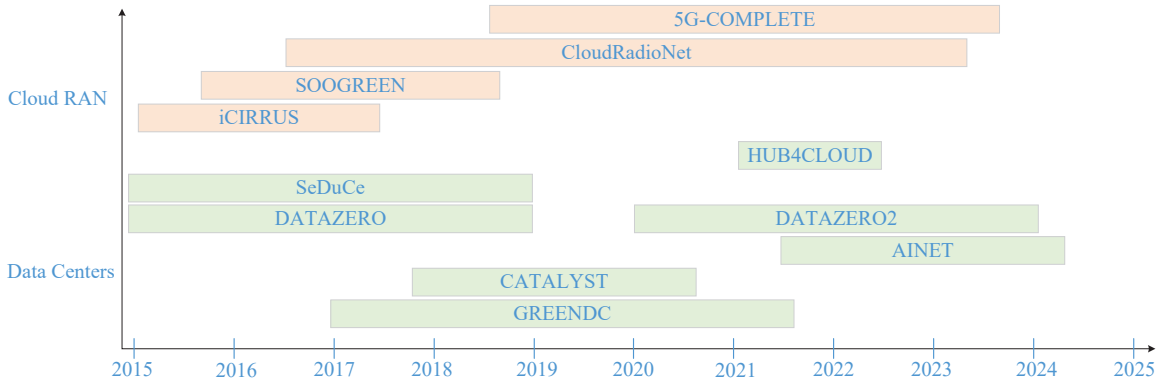


Figure 2. A timeline overview of various research projects within energy efficient Data Centers (DCs) and Cloud-Radio Access Networks (C-RAN).

B. Related research projects

DC energy consumption is a topic that receives broad attention. The European Commission has created the group "Data Centres Code of Conduct", to reduce DC energy consumption [44]. In the United States (US), Berkely Lab has a "Center of Expertise for Energy Efficiency in Data Centers" [45]. Some of the recent past and ongoing research projects in DC energy efficiency are listed below, and summarized in Fig. 2.

- The "Greening Datacenters", GREENDC project, is a recently finalized project where knowledge of DCs operations was transferred from industry to academic partners, where simulation based optimization for best practice of energy demand control was examined [46].
- The "Converting DCs in Energy Flexibility Ecosystems", CATALYST project, investigated how to combine existing and new DCs into flexible multi-energy hubs [47].
- The AINET project focuses on edge DCs to provide enablers and solutions for high-performance services by the use of Artificial Intelligence (AI), which also includes methods for minimizing energy consumption [48].
- The "DATAcenter with Zero Emission and RObust management using renewable energy", DATAZERO and DATAZERO2 projects, investigate how a DC only operated by renewable energies alone can work. Where the focus in DATAZERO2 is the operation and design of cooperating DCs [49].
- The "Sustainable Data Centers: Bring Sun, Wind and Cloud Back Together", SeDuCe project, aimed to design an experimental infrastructure dedicated to the study of DCs with low energy footprint. Resulting in a testbed for research on thermal and power management in DCs [50].
- The "The European Cloud Computing Hub to grow a sustainable and comprehensive ecosystem", HUB4CLOUD project, aimed to magnify the impact and relevance of cloud computing research, innovation, and policy-driven efforts in Europe with the ambition of environmentally sustainable cloud technologies and solutions [51].

For the sake of C-RAN, some of the recent past and ongoing research projects are listed below, and summarized in Fig. 2.

- The "Service-oriented optimization of Green mobile networks", SooGreen project, investigated how to reduce the energy consumption of services in light of the traffic evolution and exploit new network architectures including hybrid C-RAN [52].
- The "intelligent Converged network consolIdating Radio and optical access aRound USer equipment", iCIRRUS project, proposed an intelligent C-RAN solution bringing together optical fibre technology, low-cost but highly flexible Ethernet networking and wireless resource management [53].
- The "Cloud Wireless Networks: An Information Theoretic Framework", CloudRadioNet project, aimed at developing novel information theoretic concepts and techniques and their usage, to identify the ultimate communications limits and potential of different C-RAN structures [54].
- The "Computational and stOrage resource Management framework targeting end-to-end Performance optimization for secure 5G muLti-tEchnology and multi-Tenancy Environments", 5G-COMPLETE project, utilizes the C-RAN architecture to build an unified ultra-high capacity converged digital/analog fiber-wireless transport network for the RAN [55].

C. Our contribution

This current work explores numerous strategies for improving energy efficiency in DCs, while taking the different sizes of DCs into consideration. Furthermore, we examine the size of DCs to be used for C-RAN, to explore how these can become the most energy efficient, which is a major additional contribution of this paper compared to [1], as well as the extended overview of related work. Thus, we provide an in-depth overview of key challenges, opportunities and methods for improving energy efficiency in all types of DCs, but with a major focus area in DCs for C-RAN. Hence, we:

- Provide insights into effective ways to improve energy efficiency in DCs by synthesizing the state-of-the-art technologies and techniques for DCs of different sizes.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

- Analyze previous surveys and papers on energy efficiency in DCs, and provide an overview of current research projects.
- Offer engineering guidelines for DCs in C-RAN mobile networks.
- Examine recommendations towards achieving minimized energy consumption in DCs for C-RAN.

III. Data Center basics

This section introduces various types and components of a typical DC, before exploring the energy minimization opportunities in the next chapter. Two types of DCs include private enterprise DCs and public cloud DCs. As illustrated in Fig. 3, the end-users gain access to the DCs to store and process data through a network of computers, wireless Access Points (APs), switches/routers and the Internet. The computers are connected to both DCs through a switch or router, which directs the data traffic between them.

An enterprise DC is located inside the same local network as the users, while a cloud DC is located outside of the local network. Cloud DCs are typically managed and owned by third-party service providers and the services they provide are accessed through the Internet. Users can access both types of DCs by authenticating themselves and then proceed to transfer data through the nodes in the network. Fig. 3 gives an overview of a basic network architecture where users have access to both an enterprise DC and a cloud DC. The benefits of connecting to both a cloud DC and an enterprise DC for data access and exchange are:

- The cloud DC enables remote, on-demand access to data and application services from other providers through the Internet.
- The enterprise DC provides more secure, local access to other types of data and applications, which is beneficial for vulnerable data.

The architecture of a DC plays a crucial role in its overall energy efficiency. Several components make up a typical DC architecture, including [18]:

- *Server Racks:* The servers themselves, as well as the physical stations that house the servers in a DC and consume energy for processing and cooling. Server racks are designed to organize, store and manage numerous servers, while optimizing floor space at the same time.
- *Top of the Rack (ToR) Switches:* Switches connected to every server in a server rack and connects those to the network. A ToR switch can be located at the top of each server rack to provide the connection between the servers and the network. They are responsible for forwarding data packets between servers and the rest of the network. This is however; depending on the chosen DC architecture [56].
- *Aggregation Switches:* A centralized connection point for assigned ToR switches. Responsible for collecting data traffic from multiple servers and forward it.
- *Load Balancers:* Devices responsible for distributing network traffic evenly across several servers, reducing the probability of network failures by lowering workload of overwhelmed servers.
- *Access Routers:* A secure connection point for external network traffic.
- *Core Switches/Routers:* Devices responsible for forwarding traffic at a high speed within nodes of a DC network.
- *Edge Routers:* Handles incoming and outgoing network traffic by routing data from and to the DC.

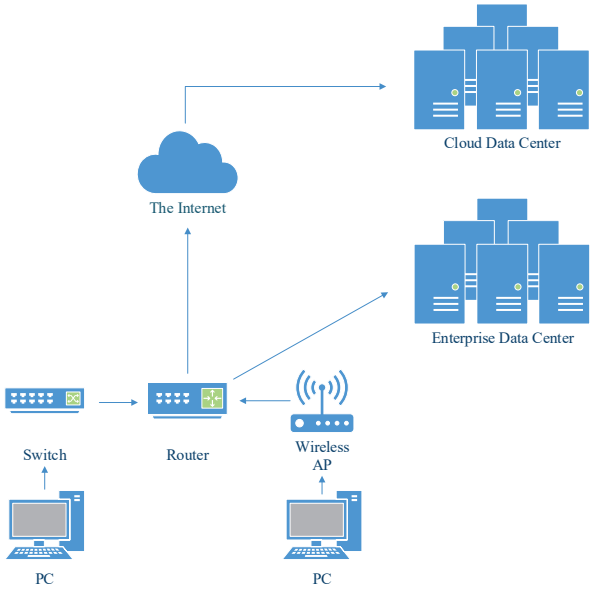


Figure 3. A basic network with computers connecting to both a cloud Data Center (DC) and an enterprise DC.

The various components cooperate to distribute, forward and transmit data traffic stored in a DC, and understanding the purpose and role of each device is key when optimizing energy efficiency in DCs. Fig. 4 illustrates the basic elements of a DC architecture, designed to efficiently process and manage data. It includes server racks connected to ToR switches which forwards traffic to the aggregation switch. The data is then directed to the load balancer which distributes the traffic between the servers. The access router controls access to the DC network and the core switch/router forwards to the edge routers which serve as the bridge between the internal DC network and the external network, being the Internet.

Table II provides an overview of the various DC components and their appearance in the different DC sizes. Furthermore, the table provides an overview of the energy consumption of the various elements using numbers from [56]. However, comparing these numbers to the work in [57], then the energy consumption of servers, storage and communications equipment only account for approximately 50% of the total DC energy consumption. The cooling systems require 40% and the power supply system require the remaining 10% [57]. Thus, in light of the total DC energy consumption the servers will consume 35%, aggregation switches 5%, access routers 7.5% and core switches 2.5%. The break down of energy consumption figures are illustrated in Fig. 5.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE II. DATA CENTER COMPONENTS BY SIZE

Component	Small scale	Mid scale	Large scale	Energy consumption
Servers	<1,000	1,000-10,000	10,000<	70% [56]
ToR switches	<50 [58]	50-500	500<	Expected to be part of the aggregation switch energy consumption
Aggregation switches	<25 [59]	25-250	250<	10% [56]
Access routers	<25 [60]	25-250	250<	15% [56]
Core switches	<12 [59]	12-120	120<	5% [56]

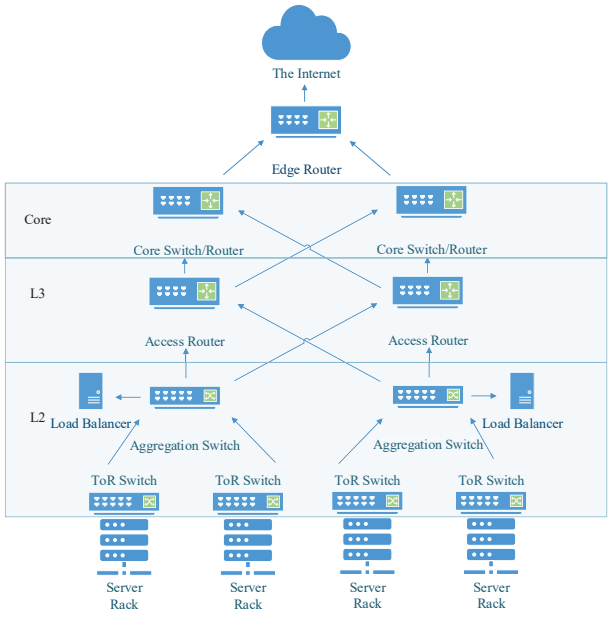


Figure 4. Key components of a basic data center architecture separated into different layers.

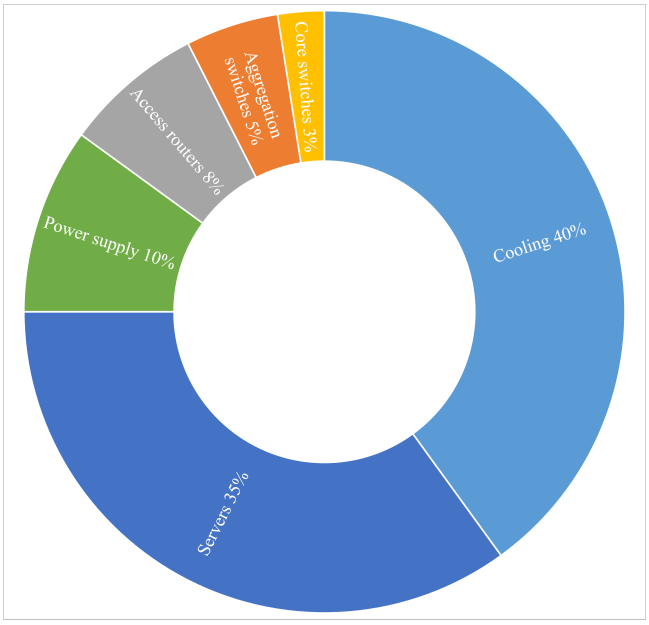


Figure 5. Data Center (DC) key components' energy consumption in relation to the total DC energy consumption.

IV. Energy Minimization Methods

Energy minimization methods refer to the numerous strategies used to reduce the energy consumption of DCs with the goal of minimizing energy consumption while maintaining high levels of performance and reliability. Server utilization in DCs are found to be under 20% most of the time and with the servers still running fully, this results in very low energy efficiency since servers still consume a significant amount of energy even when not fully utilized [21]. A common tool for measuring energy efficiency in DCs is the Power Usage Effiectiveness (PUE) metric. It is calculated by dividing the total amount of energy used by a DC, including all systems and components, by the energy used by the IT equipment within the DC [61].

This section will give a brief overview of multiple technologies and techniques as well as going in-depth with some subcategories of these strategies, being: sleep state methods and resource utilization in their own subsections. Fig. 6 illustrates where in a DC certain methods are utilized and what components are involved by highlighting the energy efficiency strategies with different colors:

- Green for load balancing and scheduling
- Blue for cooling systems optimization
- Yellow for sleep state methods
- Pink for Data Center Infrastructure Management (DCIM) tools

Figure 6 can be used as an overview of in which components the various energy consumption minimization strategies belong. The strategies will be further elaborated below.

A. Trending methodologies

Energy efficiency in DCs can be achieved through a variety of strategies. Examples of current research directions are:

- Advanced cooling systems
- Server virtualization
- DCIM tools
- Edge computing
- AI-driven DC Management
- Quantum computing

Advanced cooling systems are innovative technologies used for mainly cooling servers and can result in notable energy savings. Liquid cooling, free cooling and indirect cooling are some of the advanced types of cooling systems [30]. However, opportunities to place DCs underwater are also being investigated [62]. Another relevant method in this category is

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

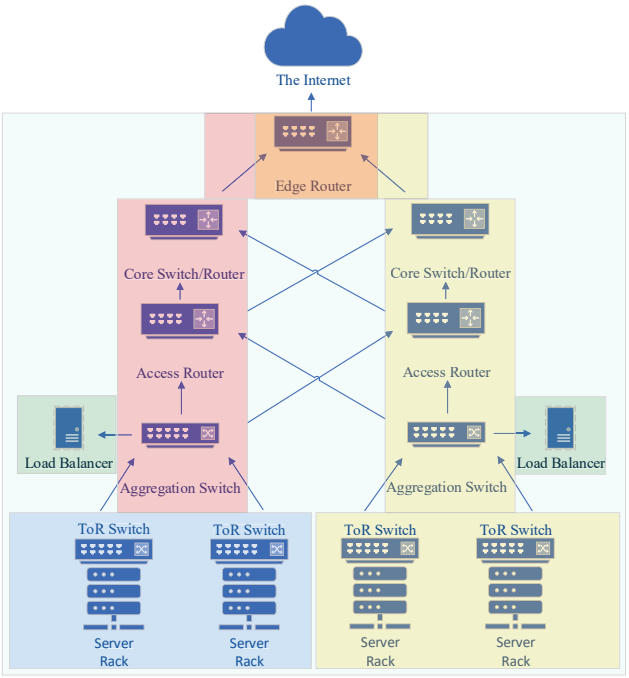


Figure 6. Energy minimization strategies highlighted in color for involved key components of a Data Center (DC). Hence, some of the components benefit from more than one strategy and thus; they are represented in different colors. The green colored components benefits from load balancing and scheduling. The dark blue components benefit from cooling systems optimization. The yellow components benefit from sleep state methods. The pink components benefit from DC Infrastructure Management (DCIM) tools.

heat re-use, which refers to the process of utilizing waste heat generated from one process or system and using it for another purpose, rather than letting it go to waste [3]. Such purposes could be to heat up greenhouses in cold regions [63] [64] [65], houses or whole cities [64] [66], also swimming pools and even laundries can make use of the heat generated in DCs [64].

Server virtualization can lower the number of needed servers in a DC by running multiple virtual servers on a single physical server, resulting in lower power consumption [22]. The technique is an enabler of methods to improve resource utilization, which will be examined later in this chapter [43]. Servers can be virtualized by realizing the functions in software by the use of either Virtual Machine (VM) or container systems. Thus, they need an integrator between the hardware and software, which can be a hypervisor for VMs or a container engine for containers.

DCIM tools are used to monitor, measure, manage and/or control DC utilization and energy consumption of DC equipment such as servers, storage systems and network switches/routers. This helps identify power-related issues and improve DC performance and energy efficiency [67].

Edge computing refers to a range of networks and devices at or near the users and enables processing data closer to where it is being generated. Edge computing can reduce the amount of data traffic that needs to be transmitted to a central DC, resulting in potential energy savings [23]. However, this represents a trade-off between centralization benefits for larger DCs and energy savings in the transport infrastructure. Hence, more centralized traffic in one DC will open up for more opportunities for sharing existing resources, but the cost is a comprehensive transport network, which is examined in [41], [68]. Edge computing however, might prove to be most beneficial for smaller DCs. Large DCs are much more centralized and have a much greater power density which counteracts the whole principle of edge computing. However, egde computing results in more and smaller DCs rather than one or a few large DCs and thus, if considering the amount of data processed, then the potential for higher resource utilization increases by the number of servers in the DC. Hence, more traffic will bring a larger potential for utilizing the servers at different times of the day. On the other hand, not all traffic can be transported to distant DCs, including time critical operations for mobile networks.

AI-driven DC Management is a method for automating control and monitoring of DC resources. By improving DC operations, energy efficiency improves as well [69].

Quantum computing defines super powerful computers that uses quantum technology to create a 3D reference model that significantly increases the computational performance compared to a normal computer. Quantum computing has the potential to increase energy efficiency in DCs by solving complex problems at an incredible speed compared to traditional computing methods. However, it is a new technology and still in its early stages [70]. Furthermore, the quantum computing requires extremely low temperatures and thus; the energy consumption of the cooling system is expected to be higher than the energy consumption of the computers [71].

B. Sleep states

Sleep states can be implemented to shut down several server components for a short period of time to reduce energy wasted on un-used server capacity. Fig. 6 illustrates the components in a DC that can be impacted by sleep state methods, being both switches, servers and routers at various levels. When utilizing sleep state methods, the components that are being powered down are the Central Processing Unit (CPU), cores of the CPU, memory and storage devices [21]. The devices and nodes that are involved when utilizing this method are highlighted in Fig. 6, marked by yellow. Modern processors support multiple types of sleep states, primarily:

- Core C-states
- Package C-states
- P-states
- Dynamic Random-Access Memory (DRAM) power mode

Core C-states work by stopping executions on the core. They range from C1-C6 and the differences between those being the varied amounts of power savings and exit latency costs, which will here be referred to as wake up time. C0 is the active state, with no CPU power savings. C1 is the state with the least power savings but with the shortest wake up time whereas C6 is having the longest wake up time at a 133μs transition time [21], however; this number is depending on

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

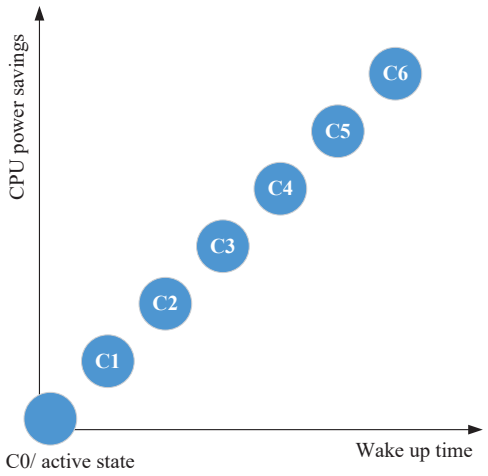


Figure 7. Illustration of the 6 Core C-states according to wake up time and opportunities for power savings.

the protocol used. The core c-states and their relation between wake up time and CPU power savings are illustrated in Fig. 7.

Package C-states are used when all cores are in state C1 to C6, hence; the entire CPU is idle. In this state a whole package of components turns off, such as shared caches, integrated Peripheral Component Interconnect Express (PCIe), memory controllers, and so on [24]. However, the concept is that additional power is saved compared to the power saved with the sub-components individually [24]. Package C-states can significantly reduce energy consumption but has the side effect of increasing the latency for cores going to or from low power states [21]. Furthermore, package C-states can be problematic because of high response times during re-activation when handling traffic spikes. Additionally, having the memory and/or storage of all servers to be available, even during times of light load, can be very beneficial. Lower latency can be achieved by AgileWatts (AW) [21] which is a deep idle core power-state architecture that reduces the transition latency to/from very low power states. AW has been proven to result in up to 71% power savings per core with a less than 1% end-to-end performance decrease [21].

P-states changes the frequency and voltage of a part of the system. This being the cores or other components such as a shared Layer 3, the network layer (L3) cache [24]. P-state is a module state affecting a collection of cores that share resources [24]. The concept of P-states is that a CPU running at lower frequencies requires lower performance and longer latency to complete a certain amount of work. Thus, under some circumstances, for example in low traffic periods, it is possible to complete a required amount of work with lower energy [24].

The DRAM power mode consists of two power-saving methods which are the Self-refresh function and the Clock Enable (CKE) mode. CKE sends a signal from the Memory-Controller (MC) to the DRAM device, and when this signal is no longer being sent, the DRAM is free to enter a low power state. In the Self-refresh function, the MC sends the refresh signal to the DRAM to ensure that the data is valid. DRAM has the ability to start the Self-refresh process itself, which can reduce the power consumption in the MC [72].

C. Resource utilization

DCs' load rises when more requests are received, and these requests can be received seasonally. Thus, the workload demands of the servers are changing dynamically and are determined by a real-time workload status. By balancing the load on the servers carefully and properly, it is possible to increase the energy efficiency of components in a DC. Fig. 6 illustrates resource utilization techniques within a DC, highlighted as green. This work examines two different ways of increasing the resource utilization:

- Load balancing
- Scheduling

Load balancing can be explored using various methods. Dynamic Time Scale based Server Provisioning (DTSP) is a method which takes the variability of workloads into consideration when providing servers for workload demands. For DTSP to load balance properly, key information is gathered constantly so that DTSP can accurately estimate workload requirements on servers and specify the appropriate number of servers for the dynamic workloads [19]. Irregular arrivals of requests impact the accuracy of the expected workload. To increase the estimation, the gathered information of incoming requests is standardized before it is used in later calculations. When it comes to workload, the algorithm looks at the three factors; arrival rate of previous requests, the arrival rate of current requests and the mean service time of current requests. With these factors, the algorithm is able to figure out the intensity of previous workloads and reflect the available remaining capacity for the unfinished waiting workloads, as well as measure the intensity and time needed for current workloads to complete. These factors are also used when calculating the workload demand of incoming requests and to determine how many servers are needed to finish current and remaining workloads while satisfying the Quality of Service (QoS) requirements [25]. DTSP has been proven to be able to estimate the workload demands of servers in a DC. By periodically adjusting service resources to match workload demands, DTSP significantly improves and maintains the system energy efficiency under an acceptable QoS level [19]. The work in [43] explores five different load balancing methods, the dynamic, predictive, energy-conscious application scaling, energy efficient and generic algorithm with population reduction. Results show that each of the investigated methods have individual pros and cons when evaluating them based on various parameters.

Scheduling can be used to prioritize which machines that run what jobs or are higher utilized. A cloud system uses virtualization technology to provide cloud resources such as CPU and memory to users in the form of virtual machines.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Tasks and job requests are assigned on these VMs for execution. The technique known as job scheduling is a method used to assign a job to a VM based on classification. By allocating jobs based on types and availability, it is possible to increase energy efficiency by making better use of available resources. Minimizing the number of hosts used when allocating resources reduces energy consumption. The Energy Aware VM Available Time (EAVMAT) scheduling algorithm does exactly this [26]. By categorizing jobs into three types and then assigning jobs based on a predefined policy with the earliest available resource. Energy consumption is then reduced since less hosts are in an active state and resource utilization is higher. This method has been tested and was able to achieve up to 46% energy savings [26].

V. Data Centers for Cloud-RAN

The C-RAN mobile network architecture centralizes the baseband processing of a number of sites in DCs, as illustrated in Fig. 1. Hence, for each RU the associated baseband processing is divided into DU and CU functions, which can remain on the cell site or be centralized in a DC. The potential maximum number of sites associated with one DC, will depend on several factors, being:

- Cell density in covered area
- Type of installations in covered area
- Latency limit in transport network
- DC efficiency

The *cell density* is an important parameter when estimating the amount of DUs and CUs in an area, and it requires knowledge about the Inter-Site Distance (ISD) in the current area. Hence, the ISD in an urban area is expected to be much shorter than the ISD in a rural area, due to the higher capacity requirement and more obstacles. More sites are equal to more installations and thus; more CUs and DUs.

The *type of installations* will vary based on the specific area. In an urban area more equipment is installed compared to a rural area due to the higher capacity requirement.

The *transport latency* limit is set by the requirements of the current network segment. The network segment connecting the RU to the DU is referred to as the fronthaul network, where the network segment connecting the DU to the CU is referred to as the midhaul network. The functions located respectively in the DU and CU, referred to as the high layer functional split or High Layer Split (HLS), is already standardized by 3rd Generation Partnership Project (3GPP) in [73]. However, the low layer functional split or Low Layer Split (LLS), separating the functions of the RU and DU is still a discussion topic amongst various industry alliances and standardization bodies [74].

The *efficiency of the DC* is also a parameter, since more efficient equipment can handle higher traffic loads, and on the contrary, if the capacity of the DC is not enough, it might not be able to handle traffic from all of its potential coverage area. Hence, if it is a private DC it will require more effort to upscale than using a public cloud solution where capacity is rented on demand. This is also a more difficult parameter to evaluate since it will vary based on vendor capabilities, and because this is an area in continuous development.

A. Engineering guidelines

In the light of the implementation factors mentioned, then in order to determine the size of a DC for C-RAN, the Mobile Network Operator (MNO) must be mindful about several conditions:

- On-site installation
- Number of basebands or DU and CUs in the current area
- Server efficiency
- Size of area to deploy C-RAN

The on-site installation defines what equipment in terms of RU, DU and CU are installed at the cell site. The type of RU is defined by the LLS used. Following 3GPP recommendations, for the HLS and a variety of LLSs, including the one used in today's installations; then the fronthaul transport delay must be < 250 μs [73] and the midhaul transport delay must be < 10 ms [73]. Thus, assuming a fiber propagation delay of 10 μs/km [3], then the maximum distance from the farthest site position and RU is 25 km to the associated DU DC and up to 1000 km to the associated CU DC [3]. These distances provides an approximation of how large an area a DU DC and a CU DC can cover, by assuming the DC covers a circular area with the maximum distance as the radius. Hence, this is corresponding to 1900 km^2 for the DU DC, which is the size of the Hawaiian island of Maui. On the other hand, the CU DC can cover more than 3 million km^2, which is larger than the country of Argentina, or approximately 1/3 of Europe. A MNO has three potential placement scenarios if they want to centralize their processing in DCs:

- Scenario A: To leave the DU on the cell site and move CU functions to a DC.
- Scenario B: To centralize the DU functions in one DC closer to cell sites and centralize CU functions from multiple DU DCs in one CU DC.
- Scenario C: To centralize both DU and CU in the same DC.

The three scenarios are illustrated in Fig. 8, where the various latency requirements are stated too. Hence, if the MNO wants to install CU functions in the same DC as the DU, then the DU transport latency requirements must be met.

The number of basebands will determine the size of the DC. This number is depending on the type of area, since an area with higher population or frequent visits by many people, like a huge train station or a concert hall, require more equipment and capacity. Thus, since more capacity can be added to an area by deploying more sites, then areas requiring high capacity will have a shorter ISD compared to areas with lower capacity requirements.

The server efficiency is difficult to measure and is an area in continuous development. In order to be able to compare the efficiency of a COTS server and a proprietary baseband installation, the performance must be measured. The performance can be quantified by examining the number of jobs executed in

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

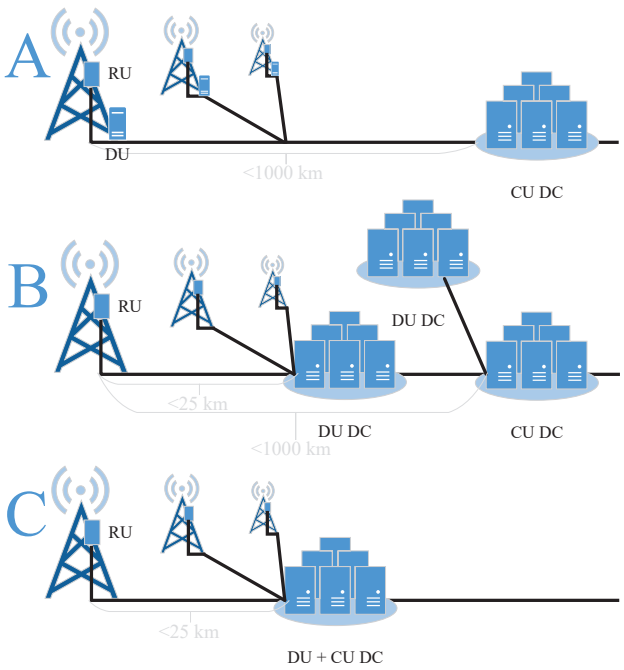


Figure 8. Three Cloud-Radio Access Network (C-RAN) installation options: Distributed Unit (DU) on the cell site and Centralized Unit (CU) centralized in a Data Center (DC) [A], DU and CU in different DCs [B] and in the bottom, DU and CU in the same DC [C].

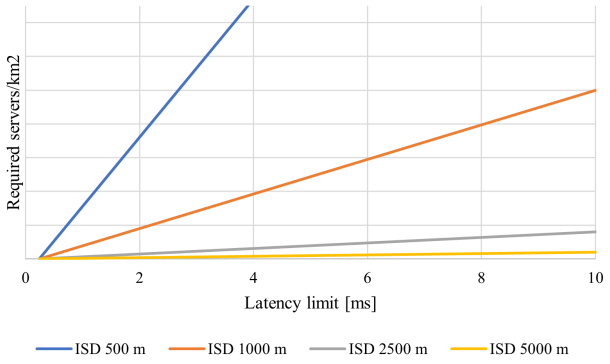


Figure 9. The figure shows the capacity in terms of required servers as a function of the latency required by the fronthaul network.

a certain time interval and the maximum load of one unit. In order to make an estimate for this work it is expected that one onsite DU+CU will correspond to one server, with the traffic distribution 2/3 to DU and 1/3 to CU. This is a topic that leaves room for further investigations, because how efficient is a Commercial off the Shelf (COTS) server actually compared to a proprietary baseband? However, if the efficiency of the proprietary baseband and the server(s) running the DU and CU functions are not 1:1 efficient then the number of COTS servers required might be less or (more likely) more than the proprietary installations in the traditional RAN. Hence, the server efficiency will affect the number of servers in the DC. The relationship between latency and capacity in terms of required servers in the DC is explored in Fig. 9. The figure shows the required number of servers for various ISDs when complying with different requirements to fronthaul latency.

The size of the area determines how many DCs are required to cover the current area in order to comply with the RAN latency requirements. Furthermore, the placement scenario selected will also determine the need for multiple smaller or one larger DC.

B. Case study

This case study, investigates the number and sizes of DCs required for the Danish MNO TDC Net [75] to convert to a C-RAN installation in their mobile network. Thus, real life numbers and approximations are used to evaluate the sizes of required DCs. TDC Net provides 99% geographical 5G coverage in the country of Denmark and utilizes approximately 4000 sites. The country of Denmark can be seen in Fig. 10. The country of Denmark is approximately 43,000 km^2 and thus; only one DC can according to the transport latency requirements carry all CU data of the whole country. According to transport latency requirements, then at least 23 DCs are necessary to handle the DU traffic. In this case we assume 23+ DU DCs since more might be required due to practical implementation specifics. In order to explore the C-RAN DC opportunities for TDC Net two areas with a radius of approximately 25 km are selected. One urban area with high traffic loads and many users present at all times of the day, and one rural area with the complete opposite capabilities. Table III summarizes the parameters of the urban area, rural area and the whole country, utilizing the parameters stated in the engineering guidelines. As stated in the table, the capacity illustrated by utilized spectrum in the current area, differs a lot in the different areas. The capacity here includes both Long Term Evolution (LTE) and New Radio (NR) cells. The rural area chosen has a lower average capacity compared to the average capacity in the whole country, where the urban area used here have a much higher average capacity compared to the average capacity for the whole country. In the following subsections, scenarios A, B and C presented under engineering guidelines, will be explored in the light of the case study.

1) Scenario A: Only one CU DC, is necessary for covering all of Denmark's CU traffic. This DC shall be able to handle upper layer traffic from all 4100 sites covering the whole country, with a total of 7100 basebands. Thus, expecting 1:1 performance of current installations and DC servers, with 1/3 traffic handled in the CU as described under DC efficiency in chapter V. Then 2400 servers will be required to handle CU traffic, corresponding to a mid-scale DC. Data is summarized in Table IV.

2) Scenario B: 23+ DCs are nescessary to handle all DU traffic. The DU DCs' sizes will depend on the cell density and installation types of the current area. Thus, two areas will be considered, a rural and an urban area in Denmark. The areas are compared in Table V considering an urban area with a total of 900 sites (including macro, pico and indoor systems) and a rural area covered by 50 sites, both areas covering approximately 25 km from the area center. When expecting

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE III. AREA SPECIFICS

Parameter	Urban area	Rural area	Denmark
Type of RU	LLS8	LLS8	LLS8
Fronthaul latency limit	< 250 μs	< 250 μs	< 250 μs
Total macro sites in area	900	50	4100
Total basebands in area	1400	100	7100
Average capacity per site	109 MHz	77 MHz	83 MHz
Estimated server efficiency	1:1	1:1	1:1
Approximated size of area	1900 km^2	1900 km^2	43,000 km^2

Figure 10. The country of Denmark. The image is a creative common under license CC BY-SA.

TABLE IV. CU DC

Parameter	Denmark
Assumed traffic distribution	1/3
Servers in DC	2400

1:1 performance of current installations and DC servers, and DU traffic corresponding to 2/3, then 940 servers will be required to handle DU traffic in the urban area, corresponding to a small-scale DC. For the rural case, only 60 servers will be necessary to handle DU traffic, corresponding to a minor small-scale DC. Data is summarized in Table IV.

3) Scenario C: If the DU and CU is both located in the same DC, then the covered area will be limited by the latency boundaries of the DU, and thus; 23+ sites are necessary to cover the whole country of Denmark. Table VI summarizes the required size of DC in a rural and an urban area. As the table shows, then the rural area is still covered by a small-scale DC. However, the urban area DC becomes a mid-scale DC with 1400 servers required to handle DU and CU traffic.

TABLE V. DU DC EXAMPLES

Parameter	Urban	Rural
Assumed traffic distribution	2/3	2/3
Servers in DC	940	60

TABLE VI. DU AND CU DC EXAMPLES

Parameter	Urban	Rural
Assumed traffic distribution	1/1	1/1
Servers in DC	1400	100

VI. Discussion

DCs being responsible for approximately 1.5% of global carbon emission with an annual growth rate of 4.3% have become an area of focus within the last decade [28]. However, a lot of attention has been directed towards the larger DCs, which are only responsible for a small portion of the overall energy consumption of DCs in general, since small-/mid-scale sized DCs are responsible for approximately 50% of the energy consumption [28]. Due to the increased attention, large-scale DCs have therefore advanced more than small-scale DCs and have numerous energy efficient methods implemented already. It is shown in [28], that energy efficient strategies such as virtualization are adopted less in smaller DCs compared to large DCs. Small DCs are in general behind on the energy efficiency front with around 43% of them not having energy efficiency objectives in place at all [28]. When energy optimizing the mobile networks, the baseband processing of the RAN is moved into DCs categorized as small-or mid-scale. Thus, if the energy usage should not just be passed on to the next segment of the network, ie. the DC, it is important to consider methods for minimizing the energy consumption in DCs for C-RAN as well.

The benefits of different strategies used for energy efficiency in DCs varies depending on the size of the DC. Below is recommended a set of guidelines for optimizing energy efficiency in DCs and evaluated based on the three different sizes/categories; small-, mid- and large-scale. However, it is important to stress that techniques and technologies recommended for small-scale DCs are also excellent methods for larger DCs, whereas methods recommended for large-scale DCs are not always realistic/beneficial options for smaller

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

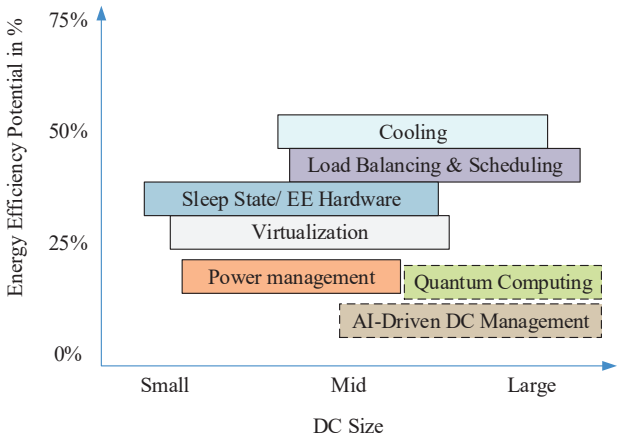


Figure 11. The graph compares the energy efficiency improvement percentages achieved through different energy efficiency strategies in DCs of different scale.

DCs because of price and other circumstances such as the scale. On the other hand, small-scale DCs might see greater improvements when utilizing some of these strategies, since they are size-wise easier to manage, which can result in energy-efficient technologies and practices being adopted more easily. Large DCs managing thousands of servers and hundreds of server racks will likely achieve greater power savings by investing in advanced cooling systems than small-scale DCs managing less than hundred servers. Here, small-scale setups might see greater benefits investing in other technologies and techniques as described in the next subsection.

A. DC Strategies at Different Scale

Many different factors are decisive for how effective certain strategies are when it comes to the energy efficiency for DCs of various sizes. This makes it difficult to generalize the different methods as all DCs differ in relation to infrastructure, scale and utilization, environmental factors and what energy efficient technologies are already in place. Some energy efficiency strategies can provide the best results for smaller DCs compared to larger DCs, since small-scale DCs have fewer resources available as well as generally not even having implemented any energy efficiency strategies at all [28]. Fig. 11 shows potential power savings of different strategies for varying DC sizes. Furthermore, below various opportunities for minimizing energy consumption in DCs of different scale are examined:

Small-scale and mid-scale DCs benefit from energy efficiency strategies such as sleep state methods and power management tools, as well as virtualization, load balancing and energy efficient hardware. Additionally, small-scale DCs can also benefit from design optimization including efficient cooling systems and energy efficient infrastructure.

Large-scale DCs have access to more resources and can allocate those towards many different energy-saving measures, including advanced cooling systems, server virtualization, load balancing as well as renewable energy sources. Having access to additional resources opens up for other strategies such as AI-driven DC management and quantum computing as large-scale DCs also have more data traffic to handle. Modern energy efficiency strategies such as advanced cooling systems have proven to potentially achieve energy saving of up to 50% [30], virtualization has proven possibilities of 30% [22], sleep state methods can provide up to 34% energy savings [27], and resource utilization methods can reduce energy consumption by up to 46% [26]. All these strategies are beneficial for DCs of all sizes but can vary in potential energy savings depending on multiple different factors. DCIM and PUE are also excellent methods for working towards more energy efficient DCs and can provide beneficial tools for analysing DCs of all sizes. That being said, as well as being able to utilize and implement the technologies and techniques mentioned for smaller DCs, large-scale DCs does also have other possible methods for achieving greater energy efficiency. AI driven DC management and quantum computing are both methods which will most commonly be seen in large-scale DCs since the owners are able to provide sufficient resources for these technologies to be implemented and these methods are therefore recommended for large-scale DCs, along the methods mentioned for smaller DCs.

Numbers provided in section III illustrate the energy consumption of the various components of the DC and table VII shows an overview of the various methods for energy savings examined throughout this paper and the savings they provide. Furthermore, the table shows how large a reduction in the overall DC energy consumption each of the proposed methods will bring, as well as which of the components will save energy by the current method. Finally, the table also shows whether the various DC sizes can benefit from the different solutions. The table leads to the clarification that small and mid scale DCs can potentially minimize their energy consumption by up to 23% by utilizing methods mentioned earlier in this section and in the table, where large DCs can save up to 34% energy consumption by utilizing the methods proposed.

B. Energy Minimization potential for C-RAN DCs

When moving baseband processing from mobile networks into DCs, the network functions are already virtualized, otherwise they could not operate on COTS hardware. Thus, multiple DUs or CUs can run on the same physical hardware. This opens up for opportunities in resource utilization including load balancing and scheduling. By utilizing these methods it is possible to shut down un-used hardware resources in low traffic periods. Mobile traffic does vary by time and is especially lower during the night, thus; this is a great potential energy saver. Hence, during the day the users of mobile networks will move around between different areas, for instance residential and work areas, leaving one area under-utilized. The take away points from the case study is that even in urban areas, the C-RAN DC is still a minor mid-scale DC. Hence, the strategies for small-scale DCs energy minimization can be applied. By examining the strategies for small-scale DCs energy minimization, it is possible to utilize

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE VII. DATA CENTER ENERGY EFFICIENCY BY COMPONENT

Method	Saving	Reduction	Component(s)	Small scale DCs	Mid scale DCs	Large scale DCs
Cooling	50% [30]	20%	Cooling	Yes	Yes	Yes
Virtualization	30% [22]	10.5%	Servers	Yes	Yes	Yes
Resource utilization	46% [26]	23%	Servers, aggregation switches, access routers, core switches	No	No	Yes
Sleep states	34% [27]	17%	Servers, aggregation switches, access routers, core switches	Yes	Yes	Yes
Power management	20% [76]	2%	Aggregation switches, access routers, core switches	Yes	Yes	Yes

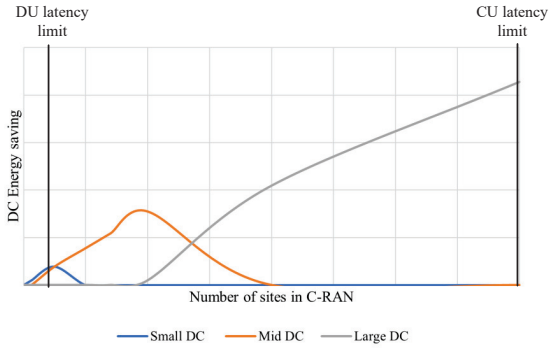


Figure 12. The graph illustrates the energy usage in small-, mid- and large scale Data Centers (DCs). The latency boundaries from Fig. 9 for a DU corresponding to ISD 0.5 km and a CU corresponding to ISD 5 km, are outlined in the graph.

sleep modes, where core C-states can be used in different ranges depending on the traffic pattern. Hence, the MNO must be aware of the exit latency, which increases with the CPU power savings. Furthermore, package C-states can be used for longer idle periods. However, since much mobile traffic is latency sensitive, P-states are not recommended.

Exploring how many DCs are actually beneficial to cover a certain area will depend on the size of the area as well as the chosen size of DC(s). From an energy efficiency perspective, both many small and one large DC have different pros and cons. Hence, looking at the various elements in the DC, as presented in Fig. 5, then some components will remain the same number when increasing the amount of DCs while others will increase, this is stated in table VII. Figure 12 shows how the various DC sizes are beneficial for different numbers of sites in C-RAN. In the figure, the maximum DC sizes stated in table II, are utilized, and when exceeding this number, more DCs of the current size are added increasing the number of cooling systems and power supplies, which are expected to be only one per DC. Thus, the figure outlines how latency requirements, as are a vital part of mobile network data flow, will be the limiting factor for re-routing DU traffic to larger DCs.

VII. Conclusion

This work investigated how the scale of a DC can impact its energy efficiency where large DCs in particular face opportunities in terms of energy minimization. On the other hand, the C-RAN trend in the RAN segment of mobile networks requires smaller and local DCs. Small and mid-sized DCs, can achieve notable energy-savings by improving design and infrastructure, as well as improving resource utilization. On the other hand, large-scale DCs can make use of the greater amount of available resources to increase energy efficiency in the same and other ways such as with AI driven resource management, new cooling methods and quantum computing. This work provided a set of engineering guidelines to be used for determining the size of a DC for C-RAN. These guidelines implicate the on-site installation, which particularly restricts the latency between the current site and the DC handling the mobile traffic. Furthermore, the size of the area and the number of installations in the current area affect the size of the C-RAN DC, and the server efficiency which is a yet greenfield area of exploration. Thus, when exploring the case study two candidate areas within the latency limit of one DU brought an insight in the size of DC required to support the C-RAN architecture if adopted in the mobile network, but highlighted the limitations within the mobile traffic latency requirements.

Acronyms

3GPP 3rd Generation Partnership Project.

AI Artificial Intelligence.
APs Access Points.
AW AgileWatts.

C-RAN Cloud RAN.
CKE Clock Enable.
COTS Commercial off the Shelf.
CPU Central Processing Unit.
CU Centralized Unit.

DC Data Center.
DCIM Data Center Infrastructure Management.
DCs Data Centers.
DRAM Dynamic Random-Access Memory.
DTSP Dynamic Time Scale based Server Provisioning.
DU Distributed Unit.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

EAVMAT The Energy Aware VM Available Time.

HLS High Layer Split.

ICT Information and Communications Technology.
ISD Inter-Site Distance.

L3 Layer 3, the network layer.
LLS Low Layer Split.
LTE Long Term Evolution.

MC Memory-Controller.
MNO Mobile Network Operator.

NR New Radio.

PCIe Peripheral Component Interconnect Express.
PUE Power Usage Effiectiveness.

QoS Quality of Service.

RAN Radio Access Network.
RU Radio Unit.

ToR Top of the Rack.

US United States.

VM Virtual Machine.

REFERENCES

[1] S. Friis, L. M. P. Larsen, and S. Ruepp, "Strategies for minimization of energy consumption in data centers," *Proceedings of Twenty-second International Conference on Networks, ICN*, pp. 17–22, 2023.

[2] The Parliamentary Office of Science and Technology. (2022) Postnote 677 september 2022 energy consumption of ict. Accessed: 15/06/2023. [Online]. Available: https://post.parliament.uk/research-briefings/post-pn-0677/

[3] L. M. P. Larsen, H. L. Christiansen, S. Ruepp, and M. S. Berger, "Toward Greener 5G and Beyond Radio Access Networks-A Survey," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 768–797, 2023.

[4] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 146–172, 2018.

[5] S. Moss. (2022) In search of the world's largest data center. Accessed: 15/03/2023. [Online]. Available: https://www.datacenterdynamics.com/en/analysis/in-search-of-the-worlds-largest-data-center/

[6] X. Wang, X. Liu, L. Fan, and J. Huang, "Energy-aware resource management and green energy use for large-scale datacenters: A survey," *Advances in Intelligent Systems and Computing*, vol. 255, pp. 555–563, 2014.

[7] F. Kong and X. Liu, "A survey on green-energy-aware power management for datacenters," *Acm Computing Surveys*, vol. 47, no. 2, p. 2642708, 2014.

[8] M. Zakarya, "Energy, performance and cost efficient datacenters: A survey," *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 363–385, 2018.

[9] M. Dayarathna, Y. Wen, and R. Fan, "Data center energy consumption modeling: A survey," *Ieee Communications Surveys and Tutorials*, vol. 18, no. 1, p. 7279063, 2016.

[10] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi, "Wireless network cloud: Architecture and system requirements," International Business Machines, Tech. Rep., 2010.

[11] China Mobile Research Institute, "C-RAN The Road Towards Green RAN," China Mobile, Tech. Rep., 2011.

[12] D. A. Temesgene, J. Núñez-Martínez, and P. Dini, "Softwarization and optimization for sustainable future mobile networks: A survey," *Ieee Access*, vol. 5, pp. 25 421–25 436, 2017.

[13] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networks - a technology overview," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 1, pp. 405–426, 2014.

[14] L. R. Jahangard and A. Shirmarz, "Taxonomy of green cloud computing techniques with environment quality improvement considering: a survey," *International Journal of Energy and Environmental Engineering*, vol. 13, no. 4, pp. 1247–1269, 2022.

[15] M. H. M. Gavali, M. S. S. Patil, M. P. U. Patil, S. P. Mane, and M. K. N. Rode, "Green cloud computing," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 4, pp. 581–583, 2022.

[16] A. Katal, S. Dahiya, and T. Choudhury, "Energy efficiency in cloud computing data center: a survey on hardware technologies," *Cluster Computing*, vol. 25, no. 1, pp. 675–705, 2022.

[17] S. Mustafa *et al.*, "Performance evaluation of energy-aware best fit decreasing algorithms for cloud environments," *Proceedings - 2015 Ieee International Conference on Data Science and Data Intensive Systems; 8th Ieee International Conference Cyber, Physical and Social Computing; 11th Ieee International Conference on Green Computing and Communications and 8th Ieee International Conference on Internet of Things, Dsdis/cpscom/greencom/ithings 2015*, pp. 464–469, 2015.

[18] L. A. Barroso, U. Hölzle, and P. Ranganathan, "Data center basics: Building, power, and cooling," *Datacenter As a Computer*, pp. 75–98, 2019.

[19] C. Hu, Y. Guo, Y. Deng, and L. Lang, "Improve the energy efficiency of datacenters with the awareness of workload variability," *Ieee Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1260–1273, 2022.

[20] M. Zakarya, "Energy, performance and cost efficient datacenters: A survey," *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 363–385, 2018.

[21] G. Antoniou *et al.*, "Agilepkgc: An agile system idle state architecture for energy proportional datacenter servers," *Proceedings of the Annual International Symposium on Microarchitecture, Micro*, vol. 2022-, pp. 851–867, 2022.

[22] M. S. B. M. Desa *et al.*, "Energy efficient approach using server virtualization in cloud data center," *2018 Ieee 4th International Symposium in Robotics and Manufacturing Automation, Roma 2018*, p. 8986732, 2018.

[23] L. C. Yan, Y. Li, H. Song, H. D. Zou, and L. J. Wang, "Edge computing based data center monitoring," *Proceedings - Ieee International Conference on Edge Computing*, vol. 2021-, pp. 17–24, 2021.

[24] C. Gough, I. Steiner, and W. Saunders, *Energy efficient servers: Blueprints for data center optimization*. Apress Media LLC, 2015.

[25] C. Hu, Y. Deng, G. Min, P. Huang, and X. Qin, "Qos promotion in energy-efficient datacenters through peak load scheduling," *Ieee Transactions on Cloud Computing*, vol. 9, no. 2, pp. 777–792, 2021.

[26] S. Loganathan, R. D. Saravanan, and S. Mukherjee, "Energy aware resource management and job scheduling in cloud datacenter," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 4, pp. 175–184, 2017.

[27] V. Anagnostopoulou, S. Biswas, H. Saadeldeen, A. Savage, R. Bianchini, T. Yang, D. Franklin, and F. T. Chong, "Barely alive servers: Greener datacenters through memory-accessible, low-power states," *Design Technologies for Green and Sustainable Computing Systems*, pp. 149–178, 2013.

[28] T. L. Vasques, P. Moura, and A. de Almeida, "A review on energy efficiency and demand response with focus on small and medium data centers," *Energy Efficiency*, vol. 12, no. 5, pp. 1399–1428, 2019.

[29] B. Speitkamp and M. Bichler, "A mathematical programming approach for server consolidation problems in virtualized data centers," *Ieee Transactions on Services Computing*, vol. 3, no. 4, pp. 266–278, 2010.

[30] Y. Gong, F. Zhou, G. Ma, and S. Liu, "Advancements on mechanically driven two-phase cooling loop systems for data center free cooling," *International Journal of Refrigeration*, vol. 138, pp. 84–96, 2022.

[31] Geetanjali and S. J. Quraishi, "Energy savings using green cloud computing," *Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, Icicict 2022*, pp. 1496–1500, 2022.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[32] F. Marzouk, J. P. Barraca, and A. Radwan, "On energy efficient resource allocation in shared rans: Survey and qualitative analysis," *Ieee Communications Surveys and Tutorials*, vol. 22, no. 3, pp. 1515–1538, 2020.

[33] A. Israr, Q. Yang, and A. Israr, "Power consumption analysis of access network in 5g mobile communication infrastructures — an analytical quantification model," *Pervasive and Mobile Computing*, vol. 80, p. 101544, 2022.

[34] M. Masoudi, M. G. Khafagy, A. Conte, A. El-Amine, B. Francoise, C. Nadjahi, F. E. Salem, W. Labidi, A. Sural, A. Gati, D. Bodere, E. Arikan, F. Aklamanu, H. Louahlia-Gualous, J. Lallet, K. Pareek, L. Nuaymi, L. Meunier, P. Silva, N. T. Almeida, T. Chahed, T. Sjolund, and C. Cavdar, "Green mobile networks for 5g and beyond," *Ieee Access*, vol. 7, pp. 107 270–107 299, 2019.

[35] M. Zhu, J. Gu, X. Zeng, C. Yan, and P. Gu, "Delay-aware energy-saving strategies for bbu pool in c-ran: Modeling and optimization," *Ieee Access*, vol. 9, pp. 63 257–63 266, 2021.

[36] M. R. Aktar and M. S. Anower, "Improvement of energy efficiency by dynamic load consolidation in c-ran," *International Journal of Communication Systems*, vol. 35, no. 6, p. e5087, 2022.

[37] M. Alemam, A. A. El-Sherif, and T. Elbatt, "Energy efficiency optimization through rrhs on/off switching technique in c-ran," *Ieee Wireless Communications and Networking Conference, Wcnc*, vol. 2019-, p. 8885821, 2019.

[38] M. R. Aktar and M. S. Anower, "Improvement of spectral and energy efficiency by coordinated transmission in c-ran," *6th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, Ic4me2 2021*, p. 4 pp., 2021.

[39] F. Zanferrari Morais, C. André da Costa, A. M. Alberti, C. Bonato Both, and R. da Rosa Righi, "When sdn meets c-ran: A survey exploring multi-point coordination, interference, and performance," *Journal of Network and Computer Applications*, vol. 162, p. 102655, 2020.

[40] J. Francis and G. Fettweis, "Energy efficiency maximization in massive mimo-aided, fronthaul-constrained c-ran," *Ieee International Symposium on Personal, Indoor and Mobile Radio Communications, Pimrc*, vol. 2019-, p. 8904159, 2019.

[41] D. Lopez-Perez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, "A survey on 5g radio access network energy efficiency: Massive mimo, lean carrier design, sleep modes, and machine learning," *Ieee Communications Surveys and Tutorials*, vol. 24, no. 1, pp. 653–697, 2022.

[42] S. Zhang, Y. Zhang, X. Gong, and R. Wang, "Freevm: A server release algorithm in datacenter network," *Ieee International Conference on Communications*, p. 6 pp., 2021.

[43] A. Kaushik, G. Khan, and P. Singhal, "Cloud energy-efficient load balancing: A green cloud survey," *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, Smart 2022*, pp. 581–585, 2022.

[44] Team E3P, European Commission. (2016) Data centres code of conduct. Accessed: 19/06/2023. [Online]. Available: https://e3p.jrc.ec.europa.eu/communities/data-centres-code-conduct

[45] Berkely Lab. (2023) Center of expertise for energy efficiency in data centers. Accessed: 19/06/2023. [Online]. Available: https://datacenters.lbl.gov/

[46] GREENDC. (2023) Green dc greening datacenters. Accessed: 19/06/2023. [Online]. Available: https://www.greendc.eu/

[47] CORDIS EU research. (2020) Converting dcs in energy flexibility ecosystems (catalyst). Accessed: 19/06/2023. [Online]. Available: https://cordis.europa.eu/project/id/768739

[48] AINET project. (2022) Ai-net-ainara. Accessed: 26/06/2023. [Online]. Available: https://aniara.ai-net.tech/home/

[49] DATAZERO. (2022) Datazero as green as possible. Accessed: 26/06/2023. [Online]. Available: https://www.irit.fr/datazero/datazero2/

[50] SeDuCe project. (2022) A testbed for research on thermal and power management in datacenters. Accessed: 26/06/2023. [Online]. Available: https://www.irit.fr/datazero/datazero2/

[51] HUB4CLOUD project. (2022) Improving cloud computing performance and sustainability in europe. Accessed: 26/06/2023. [Online]. Available: https://cordis.europa.eu/project/id/101016673

[52] SOOGREEN project, "Service-oriented optimization of Green mobile networks," 2015, accessed: 19/06/2023. [Online]. Available: https://soogreen.eurestools.eu/

[53] UNIVERSITY OF KENT, "Intelligent Converged network consol-Idating Radio and optical access aRound USer equipment," 2015, accessed: 19/06/2023. [Online]. Available: https://cordis.europa.eu/project/id/644526

[54] CloudRadioNet project. (2023) Cloud wireless networks: An information theoretic framework. Accessed: 26/06/2023. [Online]. Available: https://cordis.europa.eu/project/id/694630

[55] 5G-COMPLETE project. (2023) 5g-complete concept. Accessed: 26/06/2023. [Online]. Available: https://5gcomplete.eu/concept/

[56] D. Kliazovich, P. Bouvry, and S. U. Khan, "Greencloud: A packet-level simulator of energy-aware cloud computing data centers," *Journal of Supercomputing*, vol. 62, no. 3, pp. 1263–1283, 2012.

[57] H. Rong, H. Zhang, S. Xiao, C. Li, and C. Hu, "Optimizing energy consumption for data centers," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 674–691, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115016664

[58] L. Huff. (2015) Network architecture in the data center. Accessed: 04/08/2023. [Online]. Available: https://connectorsupplier.com/network-architecture-in-the-data-center/

[59] S. M. Nabavinejad and M. Goudarzi, "Chapter five - communication-awareness for energy-efficiency in datacenters," in *Energy Efficiency in Data Centers and Clouds*, ser. Advances in Computers. Elsevier, 2016, vol. 100, pp. 201–254. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0065245815000698

[60] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "Vl2: A scalable and flexible data center network," *Communications of the Acm*, vol. 54, no. 3, pp. 95–104, 2011.

[61] N. Horner and I. Azevedo, "Power usage effectiveness in data centers: Overloaded and underachieving," *Electricity Journal*, vol. 29, no. 4, pp. 61–69, 2016.

[62] J. Roach. (2020) Microsoft finds underwater datacenters are reliable, practical and use energy sustainably. Accessed: 26/06/2023. [Online]. Available: https://news.microsoft.com/source/features/sustainability/project-natick-underwater-datacenter/

[63] H. M. Ljungqvist, L. Mattsson, M. Risberg, and M. Vesterlund, "Data center heated greenhouses, a matter for enhanced food self-sufficiency in sub-arctic regions," *Energy*, vol. 215, p. 119169, 2021.

[64] NeRZ and Eco, "Utilization of Waste Heat in the Data Center," whitepaper. [Online]. Available: https://international.eco.de/topics/datacenter/white-paper-utilization-of-waste-heat-in-the-data-center/

[65] "RI.SE". (2023) Data center greenhouse - data collection plattform. Accessed: 26/06/2023. [Online]. Available: http://seduce.menaud.fr/

[66] W. E. District project. (2023) Heating & cooling solutions. Accessed: 26/06/2023. [Online]. Available: https://www.wedistrict.eu/

[67] D. Huang, "Data center infrastructure management," *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*, pp. 627–644, 2021.

[68] L. M. P. Larsen, M. S. Berger, and H. L. Christiansen, "Energy-Aware Design Considerations for Ethernet-Based 5G Mobile Fronthaul Networks," *Proceedings of the Fourth International Conference on Green Communications, Computing and Technologies*, 2019.

[69] A. Garg and D. Shenkar, "Drive data center management and build better ai with it devices as sensors," *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*, pp. 669–673, 2021.

[70] J. Liu, C. T. Hann, and L. Jiang, "Quantum data center: Theories and applications," p. 24, 2022, accessed: 15/03/2023. [Online]. Available: https://arxiv.org/abs/2207.14336

[71] M. J. Martin, C. Hughes, G. Moreno, E. B. Jones, D. Sickinger, S. Narumanchi, and R. Grout, "Energy use in quantum data centers: Scaling the impact of computer architecture, qubit performance, size, and thermal parameters," *Ieee Transactions on Sustainable Computing*, vol. 7, no. 4, pp. 864–874, 2022.

[72] J. Haj-Yahya, Y. Sazeides, M. Alser, E. Rotem, and O. Mutlu, "Techniques for reducing the connected-standby energy consumption of mobile devices," *Proceedings - 2020 Ieee International Symposium on High Performance Computer Architecture, Hpca 2020*, pp. 623–636, 2020.

[73] 3GPP, "TR 38.801 V14.0.0: Study on new radio access technology: Radio access architecture and interfaces," 3GPP, Specification, 2017.

[74] L. M. P. Larsen, H. L. Christiansen, S. Ruepp, and M. S. Berger, "Deployment Guidelines for Cloud-RAN in Future Mobile Networks," *Proceedings of 11th International Conference on Cloud Networking*, pp. 141–149, 2022.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[75] TDC NET. (2023) We connect Denmark. For everyone. Accessed: 02/08/2023. [Online]. Available: https://tdcnet.com/

[76] Panduit, “Real-Time Data Center Intelligence Pays Off,” accessed: 07/08/2023. [Online]. Available: panduit-capacity-management-via-dcim.pdf

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Intelligent Cipher Transfer Object for IoT Data Security

Bishal Sharma
Ingram School of Engineering
Texas State University
San Marcos, TX, USA
email: dxa6@txstate.edu

Bishal Thapa
Ingram School of Engineering
Texas State University
San Marcos, TX, USA
email: b_t220@txstate.edu

Stan McClellan
Ingram School of Engineering
Texas State University
San Marcos, TX, USA
email: stan.mcclellan@txstate.edu

***Abstract*—The availability of robust data security technologies to provide end-to-end, verifiable provenance of information is increasingly important. This study explores the new Intelligent Cipher Transfer Object (ICTO) technology as a novel approach to comprehensively securing digital data. The technology is assessed in terms of its performance and robustness relative to current security and data transport paradigms. Machine Learning algorithms are used to identify residual artifacts that may be exploited, and potential security threats associated with ICTO are described.**

Keywords-ICTO; IoT; Cybersecurity; Information Security; Blockchain; TLS; Encryption; Machine Learning; Cryptanalysis.

I. INTRODUCTION

This paper compares and contrasts key performance characteristics of technologies which are commonly used for securing information, both at rest and in transit, with emerging Intelligent Cipher Transfer Object (ICTO) technology. In previous work, we compared the performance of ICTO with well-known network transport technologies such as MQTT and conventional TCP [1]. The present work is an extension of the previous evaluation to a broader set of technologies and performance metrics, including the use of Machine Learning to evaluate the properties and potential weaknesses of a commercially available ICTO implementation. We conclude that ICTO may offer comprehensive data security independent of data location or transport even though aspects of data leakage may prove complex to resolve.

The current state-of-the-art in data security focuses on securing data when it is traveling (in transit) between network endpoints. Several complex operations, including payload encryption, may be performed on the data so that it cannot be accessed or modified during transit. However, when data is not in transit (at rest), it may be in possession of users, applications or systems where it may not be protected. Generally, multiple techniques or mechanisms are required in modern commercial or public data exchange settings (e.g., any client-server based interchange or web service) to securely transfer a piece of information from one point to another. These mechanisms have to be tightly integrated with one another to prevent any data-related leakage or other accidental disclosure of private information. The novel ICTO technology addresses the issue of robust integration of complex security techniques by creating a secure "intelligent", "self-aware", and "self-governing" object that can allow, deny, track, lock, or destroy itself based on the entity that is trying to access it, regardless of the security posture of the public (transitory) communication channel, or private (resting) environment. However, such a holistic approach to secure information may be costly in terms of computational or network performance.

After brief introductions to the specifics of blockchain and ICTO in Section II, comparative performance analyses are provided in Sections III-VI which contrast ICTO with several conventional approaches to information security. Section III describes the general experimental setup and important parameters. Section IV presents system-level performance measures which were collected and analyzed. Section V compares blockchain and ICTO in terms of computational performance and storage requirements, considering identical payloads. Section VI analyzes an ICTO implementation using cryptanalysis and unsupervised machine learning (ML). Section VII summarizes the experimental results and provides useful conclusions for the various technologies and implementations. Section VIII summarizes potential future work with these technologies.

II. BACKGROUND

Data or information security is the science of using and developing tools and techniques to prevent unwanted access to information. The fundamental elements of data security include:

- Confidentiality, which protects information from unauthorized access,
- Integrity, which guarantees accuracy and completeness of data providing assurance that it has not be tampered with,
- Availability, which makes information available to authorized parties whenever necessary,
- Authentication which provides a means of verifying the identity of users,
- Authorization, which grants access to specific resources to a user's identity, and
- Non-repudiation, which takes away false denial of possession or origination of information.

These key elements are necessary to provide individuals and organizations with assurance that their privacy is maintained and information they are using is trustworthy

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

and invulnerable to threats, regardless of how such data may have been generated.

Cybersecurity systems are designed and implemented using some combination of hardware and software so that these fundamental characteristics are in place. In an interconnected computing framework, data resides either inside memory within server systems or an end-client computer – considered to be "at rest", or in a network channel traveling from one endpoint to another – considered to be "in transit." Organizations may use public cloud infrastructures or private computing infrastructures to manage and store data [2].

Cloud computing technology is an internet based, decentralized, distributed, and virtualized computing paradigm in which operations on data as well as data transportation is carried out, often through web-based services [3]. This technology offers scalability, ease of deployment, and cost- effectiveness among other benefits [4]. However, from the perspective of cybersecurity, there are some issues and challenges that require extra preventive measures by both the cloud service provider as well as the cloud service user [5-6]. Although a typical cloud security stack includes fundamental services like authentication, access control, and encryption, data loss may still be encountered due to a plethora of issues, including server or application misconfiguration, malicious attacks, insecure data-flow pipelines, vulnerable Application Programming Interfaces (API), and mislocated data [7].

Legacy systems, which typically use privately owned/leased, on-premises systems to process digital data may seem to provide better security as data resides within relatively secure boundaries. However, complex logistical factors provide motivation to migrate toward cloud platforms, including high setup and operating cost, reduced flexibility, complicated integration, deployment, and maintenance procedures [8].

Considering the various modalities for data generation, transfer, and storage, employing the convenient perspectives of "application security", and "network security" is warranted [9]. For instance, encryption is a mechanism that is designed to prevent access to data travelling in a network. So, encryption can be regarded as a "network security" technique. However, authentication mechanisms – more broadly – those similar to Role Based Access Control (RBAC) are used to allow data access to known users, and so is an "application security" mechanism. Conventional Authentication, Authorization, and Accounting (AAA) security frameworks that enforce access controls to data and network resources and maintain accountability of network resources are prime examples of network security that is widely accepted and implemented. A robust cybersecurity framework should provide both "application security" and "network security" components to ensure protection. The Internet of Things (IoT) is an example of an important and complex domain where data security concepts may need to be viewed using multiple perspectives.

A. Security Frameworks

To gain a better idea of security of data in transit (network security) and at rest (application-level security) in modern frameworks, we use the well-known "Alice and Bob" scenario where users are communicating over the internet using a messaging website. Alice wants to say "hello" to Bob, so she uses her browser to access the messaging website. To establish a secure link, Alice's computer and the website server use public-key cryptographic schemes to first authenticate each other, and then set up a symmetric session key to encrypt outgoing messages from Alice's browser. The use of encryption along with the established session key secure the channel between the website's server and Alice. Several complex algorithms including Rivest Shamir Adleman (RSA), Diffie-Hellman, Secure Hash Algorithm (SHA), digital signatures, digital certificates, and Advanced Encryption Standard (AES) are used just to get to this point [10]. As soon as Alice's data reaches the server, it gets decrypted and thus is no longer obscured. Since the messaging website would most likely be hosted by an enterprise that handles user data, it uses IAM mechanisms to make sure that such data cannot be accessed by unauthorized or unintended entities (applications or individuals) within the enterprise. This can be seen as protecting data at rest [11]. A similar procedure is followed between the server and Bob as he receives Alice's message.

This example provides a simplified but useful perspective on modern security frameworks which employ channel security and site security for data in transit and at rest. It is important to note how the several complex cryptographic operations and access-control systems are necessary to ensure end-to-end data protection. Successful attacks on such systems are commonplace. An example of an in-transit data breach includes the 2009 attack on the A5/1 encryption algorithm used in 2nd generation Global System for Mobile Communication (GSM) network [12]. An example of an at-rest data breach includes a leak of the personal data of 100 million Capital One customers due to misconfiguration of a web application firewall in the company's cloud infrastructure [13]. Such attacks (among many, many others) show that complex integrations are prone to misconfigurations or mismanagement that can lead to catastrophic results.

Security frameworks in IoT networks use similar integrations. However, authentication and encryption of IoT sensor data is usually performed by a network gateway or other intermediate system instead of by an end-device. This architecture allows for exploitation of potentially vulnerable or unprotected links between the sensors, end-devices, and gateway or intermediate system.

Many alternative concepts have been developed to address issues in current security frameworks, including secured data self-moderating access and authorization, inserting encryption/decryption algorithm as metadata into data files, and specifying access and authorization criteria into the data objects [14-15].

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

B. Blockchain

A notable technology gaining traction for IoT-related applications is Blockchain. Blockchain is an eponymous, immutable ledger that stores transactions in blocks of data which are linked together like a chain. Blockchain can also be defined as a decentralized, shared, immutable database that makes it easier to track assets and record transactions in a network [16], enabling transparent information sharing.

The blocks in a blockchain are linked together by a hash function. Every block has a timestamp, transaction, and the hash of the previous block. The first block in the blockchain is referred to as the "Genesis block". The embedded hash validates the integrity and non-repudiation property of the data that is stored inside the block [8]. Every participating node in the blockchain network is updated regularly with the copy of the original. On a blockchain network used as a distributed ledger, practically anything of value (e.g. cryptocurrency) may be recorded and traded, lowering the risks involved for all parties [16].

Unfortunately, sophisticated security techniques such as blockchain require greater processing resources and power. As a result, blockchains have been modified to address the resource constraints in various domains. IoT devices typically do not have powerful processors and large memory. Instead, they possess just enough computational resources to periodically perform a limited number of tasks. Since these devices are battery operated, relatively inexpensive, and designed for highly specific purposes, they are not usually equipped with substantial data protection mechanisms. Techniques designed for general purpose, commercial, or enterprise systems therefore may not be well suited for application in the IoT space.

Often, the term "blockchain" is understood in casual usage to mean "cryptocurrency" due to the recent popularity of digital currency exchanges. Blockchain is the core technology behind cryptocurrency, but the application of blockchain is not limited to cryptocurrency. The distributed ledger is the major technology that introduced the concept of a Peer-to-Peer Electronic Cash System called Bitcoin in 2009 [17]. Distributed ledgers do not rely on centralized governance and exist virtually or digitally. This legitimate use of distributed ledgers, and the cryptography for securing transactions has made it a reliable alternative to traditional banking systems. The crypto architecture leverages computational power to solve complex mathematical puzzles, a process commonly known as "mining" [18]. Mining yields cryptocurrency units such as Bitcoin or Ethereum. While the influence of blockchain on IoT may seem obvious, a debate of "where to host the blockchain" remains [19]. Several implementations of blockchain technology exist to suit different applications.

Distributed ledgers address the "single point of failure" issue. Consensus mechanisms such as Proof of Work (PoW), Proof of Authority (PoA), Proof of Stake (PoS), Delegated Proof of Stake (DPoS), and Practical Byzantine Fault Tolerance (PBFT) provide a trustworthy distributed system [20]. The most common consensus mechanism is PoW which relies on computational work to validate new blocks and add them to the distributed chain. In contrast, systems implementing a centralized architecture [21] may be less reliable as sensitive data is consolidated, making it a prime target for cyberattacks [21]. The risk of tampering and counterfeiting is addressed in blockchain by the distributed nature of the ledger. Multiple sources confirm the transactions before recording. This provides high levels of reliability and security. Some applications manipulate a tremendous volume of data, and it isn't practical to store all the data in the blockchain itself. In such cases, decentralized storage or off-chain storage can be used.

Regardless of architecture or application, distributed ledgers and Blockchain can be effective in certain application domains and can provide a practical framework for distributed data storage as well as protection [22].

C. Intelligent Cipher Transport Object (ICTO)

Protecting information is difficult when all design parameters are considered, especially for IoT. Even if the channel is assumed to be secure, securing endpoints is still a challenge. A new technology capable of securely encapsulating data and embedding it with thorough access control policies may be a promising approach to address security issues in modern systems, irrespective of user, device, network, or operating system. This research aims to explore the usefulness of one such specific technology in terms of security and efficiency - Intelligent Cipher Transfer Object (ICTO) [23].

ICTO is a security technology that includes mechanisms for participant authentication and authorization for access of data which is protected by cloaking patterns. A portable dynamic rule set, which includes executable code for managing access to the protected set of participants and the protected data, is included within the ICTO. For a given user, the ICTO may provide access to some participants while preventing access to other participants based on this set of constraints [23]. The ICTO concept extends the idea of conventional AAA and RBAC concepts by cloaking data at the point of generation with specific user-defined rule sets. The owner of the data has substantial control over how or when protected data can be accessed by another party, regardless of the storage, transport, or operational environment.

ICTO is a form of encapsulation which achieves data security by embedding security techniques into the data itself. This provides a form of self-defense, user authentication, and governance and tracking ability which is independent of network or system security [23]. The key feature of this concept is that it allows data to be protected at the point of origin, eliminating a dependence on the security of the communications channel. In addition, access control policies and authentication parameters can be set by the data owner during object creation and therefore eliminate the need for third-party digital certificate providers. Fig. 1 shows the components/modules that are embedded with user data to create an ICTO object, also called "digital mixture" or "self-governing data" [23]. Implicit in the use of ICTO is a common execution platform or trusted set of libraries implemented on systems which manipulate ICTO objects.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

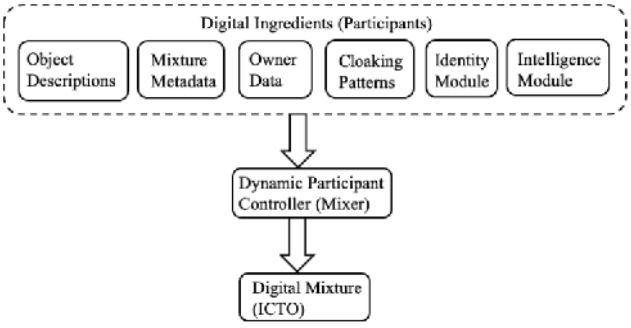


Figure 1. ICTO Creation Process.

The ICTO comprises a set of participants including a portable dynamic rule set (PDRS) which is responsible for enforcing/evaluating the rules and policies in order to allow/deny access to an entity attempting to access data. Fig. 2 and Fig. 3 illustrate simplified use cases where users can protect their data and can define specific conditions to allow or deny access depending on policies/rules set during ICTO object creation.

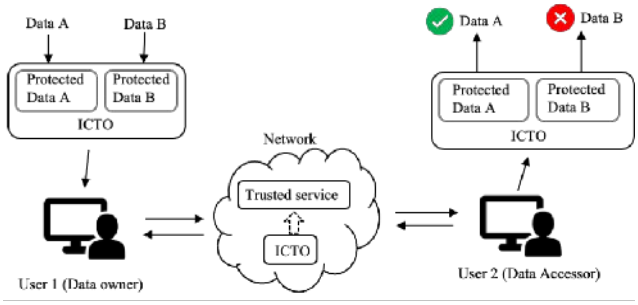


Figure 2. ICTO use case. ICTO object created by User 1 for partial data access by User 2 can be transported over an insecure network [23].

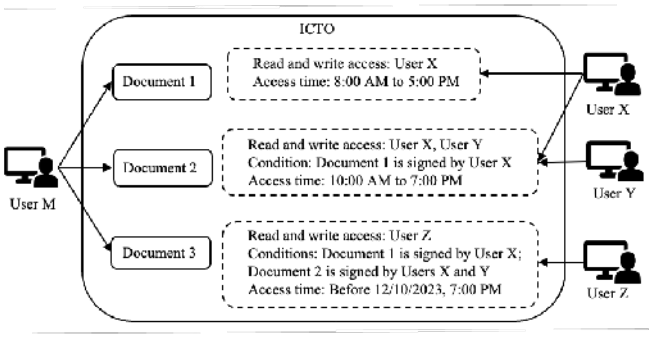


Figure 3. Access control policies in an ICTO object. Example of multiple access control policies configured within an ICTO object created by user M for multiple users X, Y, and Z [23].

Through software and tools, cipher objects containing cloaked data along with other modules are created that can only be utilized/deciphered by using compatible software.

The necessary software and tools used for this study was provided by Sertainty Corporation. Sertainty's UXP is an implementation of the ICTO concept that focuses on protection at the data layer, targeting any kind of unstructured/structured data format. UXP objects are essentially a secure, portable filesystem that hides data and access policies within a single file. For simplicity, the terms UXP objects, UXP files, ICTO, or ICTO objects are used interchangeably in this paper.

To create a UXP object, an XML file specifying user identification and definitions, access control and authorization policies, and other information is required. This XML file is used to generate a protected ID file that is then used during creation of the UXP object. The ID file is a digital object containing access control and authorization policy information, user definitions, authentications parameters etc. in a secure format. The ID object and user data are combined to create the UXP object (".UXP" filename extension). Fig. 4 summarizes the process of combining the XML file containing data policies, the resulting ID object (".iic"), and the provided user data to create a UXP object, which is an instance of an ICTO.

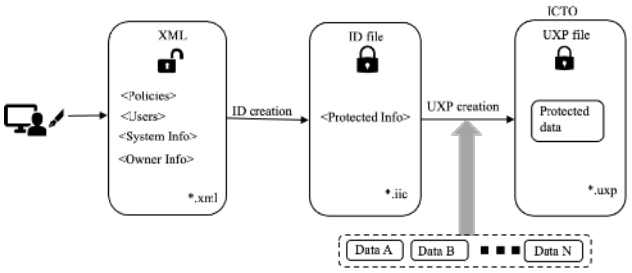


Figure 4. ICTO object creation process.

III. Experimental Setup

The ICTO implementation considered in this research is proprietary, and the property of Sertainty Corporation. The Sertainty UXP technology is chosen for evaluation as a compelling instantiation of the ICTO concept.The purpose of the outcomes in this paper summarize our understanding of the security features and performance of this promising technology in comparison with other similar competing or enabling technologies. Most experimental results are phrased in the context of an IoT-related application, and the technologies are compared or contrasted from this perspective.

A. System Performance Measurement

The computational cost of using an ICTO to secure user data is a concern, particularly for resource-limited systems. Thus, this work investigated parameters associated with ICTO creation by gathering system resource usage/overhead and network overhead data related to creation of UXP objects.

To gather performance data, a computer running a Linux operating system was used. The computer system was equipped with an Intel i5 processor having an average clock speed of 2.4 GHz, 5 MB cache, and 8 GB RAM. Linux shell and Python programming languages were used for running various experiments and for data processing.

To gather data related to network performance, secure and unsecure user data in plain text format was transported via both secure and unsecure channels to another machine

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

with similar specifications in a private network (LAN) setting.

Statistical evaluation of conventional system burden including clock cycles, memory usage, and elapsed time provided a baseline for system performance analysis whereas metrics including total transmit time, transmission overhead, and related network parameters provided a baseline for network performance analysis.

All experiments were repeated at least 200 times for each independent variable (user payload size) to obtain statistically meaningful interpretation of the resulting data. Mean and 95% confidence intervals were calculated for the statistical study, and maximum/minimum values were noted.

For experimental data related to memory utilization, the following memory metrics were monitored:

- Data resident set size (DRSS) – The amount of main memory occupied by the data segment of a running process, excluding code/instruction memory.
- Proportional share size (PSS) – The portion of main memory occupied by a process, composed by the private memory of that process plus the proportion of shared memory with one or more other processes.
- Resident set size (RSS) – Represents the total amount physical memory (RAM) currently occupied by a process, including memory for both executable instructions and data.
- Virtual set size (VSZ) – A measure of all memory that a process can access, including memory that is swapped out, unused allocated memory, and memory used by shared libraries.
- Number of bytes which a task causes to be read from storage.
- Text resident set size (TRS) – The amount of memory devoted to executable code.

These memory metrics cannot be used as a reliable estimate for system memory utilization when used individually. However, coherent results can be gathered if all of these metrics are considered.

B. Network Performance Measurement

Network performance is typically measured by looking at parameters including total transmit time, control plane overhead, round trip time of packets, and similar. For analyzing the network performance of the ICTO technology, two important metrics of total transmit time and control plane overhead were recorded and statistically analyzed.

The experimental setup involved two client and server computers with similar specifications. The client created user data that is protected via the UXP implementation of ICTO [23], and performance of the client system during UXP creation and transport was recorded. For comparison, system performance was measured in three configurations:

1. Unprotected user data sent via "plain" TCP.
2. User data protected using UXP and sent via TCP.
3. User data sent via TLS over TCP.

Since TLS v1.2 and TLS v1.3 are the most common protocol for secure communication over the Internet, TLS v1.3 was selected for the experiments [24]. In addition to comparative performance measurements, experiments were also performed to observe how compressible and incompressible data types are handled during UXP encapsulation. Plain text containing ASCII characters and digits, as well as image files of the same fixed sizes comprised the user payloads for UXP objects.

Linux utility 'forkstat' [25] was used to capture process IDs (PIDs) of processes for Python-based programs on the client system and supplied to the 'perf' utility [26] for CPU monitoring as well as to the 'ps' utility for memory monitoring. The remaining test configurations were identical to those used for system performance assessment discussed in Section III.A.

C. Blockchain Comparison

Blockchain implementations are often customized based on use cases where the objective is to solve some critical issue. This research uses a version of blockchain with a centralized cloud architecture optimized for lightweight endpoints. This implementation addresses complexity and scalability issues related to the traditional distributed ledger. However, this model doesn't address the "single point of failure" that is solved by distributed ledger.

For experimentation purposes, an alternative version of blockchain was also designed as an implementation of a distributed ledger. This implementation incorporated the PoW consensus mechanism with four difficulty levels. PoW is a popular blockchain consensus mechanism widely used in distributed ledger technology. Thus, the blockchain architecture incorporating the PoW consensus mechanism is referred to here as the "ledger." The experimental setup involved running the blockchain and ledger on a Linux system with 8 GB RAM and 4 CPUs operating at 2.4GHz.

The payloads for blockchain and ledger experiments were compressed (for losslessly compressible payloads) using the LZ77 algorithm with Huffman coding. Compressible and incompressible payloads were also encrypted using Elliptical Curve Cryptography (ECC) hybrid encryption with Advanced Encryption Standard (AES). Python's 'tinyec' library was used to generate an ECC key pair [27].

IV. System Level Performance Measurements

This section presents experimental results pertaining to system related as well as network related parameters such as CPU time, main memory, elapsed user time, and total transport time. The results are compared with other settings for a more comprehensive assessment.

A. System Performance

System performance measurement includes two primary phases of experiments. In the first phase, system statistics are recorded during UXP object creation. The second phase compares to the size of the payload data with the size of the resultant UXP objects.

To measure first-phase system performance, UXP objects were created with varying payload (user data) sizes of 1 byte, 1 kilobyte (kB), 20 kB, 100 kB, 500 kB, 1

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

megabyte (MB), 2 MB, and 3 MB. Performance metrics including number of CPU cycles, memory allocated, and total elapsed time for creating UXP objects are presented in Fig. 5 and Fig. 6 along with 95% confidence intervals, computed using the t-distribution for data with unknown variance.

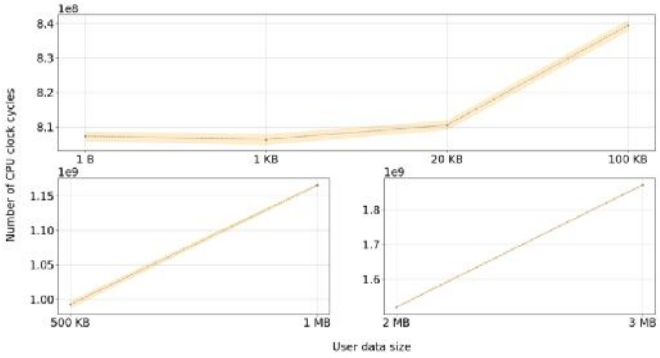


Figure 5. Plots showing number of clock cycles necessary to create UXP objects. User data size (X-axis) represents the size of user data (payload) that is protected by the UXP object. The lower set of plots accentuate the 95% confidence intervals, and indicate consistent, limited variability in the UXP creation process.

Fig. 5 shows that as the size of the UXP payload increases beyond 100 kB, the number of CPU cycles also increases, which is logical. For the smallest payload (1 byte), around 0.8 billion cycles are required for the minimum number of clock cycles to create a UXP object. As payload size increases, the number of CPU cycles increases steadily. The number of cycles required for the maximum payload size is twice the number of cycles required for the minimum payload size.

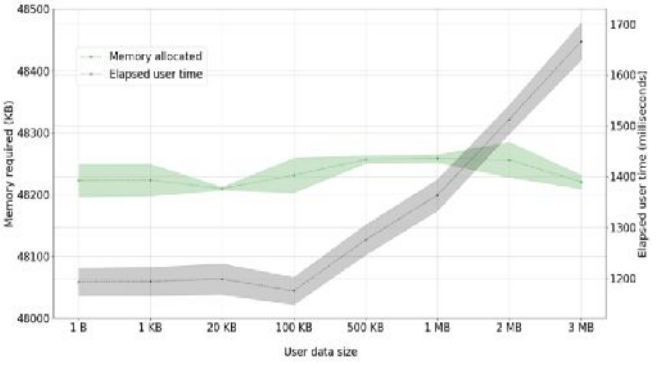


Figure 6. 95% confidence interval plot of memory utilized/time spent during UXP creation vs. size of user data protected. System memory (RAM) used is shown in the left side Y-axis and time spent is shown in the right-side Y-axis.

Fig. 6 contains two sets of data: allocated memory and elapsed time. The first set of data plotted in Fig. 6 shows that for payloads of 1 B to 100 kB, elapsed time remains relatively constant. Thus, a minimum time-to-create Fig. 6 of roughly 1 second may be observed which will vary based on system characteristics.

The second set of data plotted in Fig. 6 suggests that the total memory allocated/utilized during UXP creation remains fairly constant at under 48.3 MB regardless of payload size.

Thus, on average a minimum of 1 second and 800k cycles are necessary to create a useable, secure UXP object, based on an allocated memory of just under 48.3MB. These performance figures increase essentially linearly with payload size beyond 100kB (elapsed time) and beyond 20kB (cycles). As a result, a payload of 4MB would be expected to require approximately 1.9 billion cycles, would consume roughly 48.3MB, and would complete in under 1.8 seconds on a system comparable to those used for testing.

UXP object creation produces different results for different types of user payloads. Fig. 7 shows the size of UXP files after protecting image and text files of varying sizes. For text data, the resulting UXP object is smaller than the data for payloads larger than about 2500 kB.

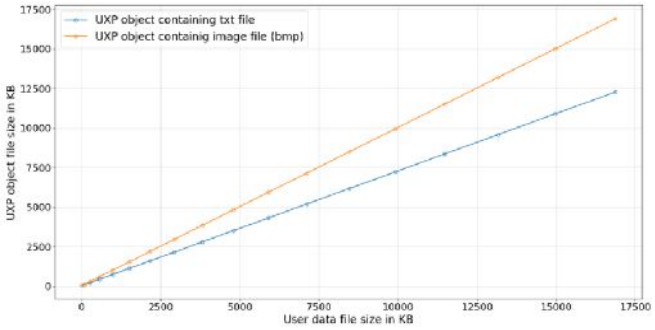


Figure 7. UXP object size for varying sizes of text and image payloads.

When the user payload is comprised of text data, the difference in resultant UXP file and original data file is negative for payloads around 264 kB. This difference increases with payload size suggesting that a lossless compression mechanism is employed during UXP object creation. However, with incompressible payloads, an almost constant positive difference is present regardless of payload.

B. Network Performance

After measuring system performance for UXP object creation, the objects were transported via a controlled network and network performance was observed. Network related performance was analyzed using system metrics such as RAM and CPU usage as well as transmission metrics such as control plane overhead and transmit time.

1) System Performance During Transport: To evaluate system performance, the number of CPU cycles and total memory allocated when unprotected user data is transported using “plain” TCP and using a TLS protected TCP channel is presented. These results are contrasted with similar results from the transport of UXP objects using “plain” TCP. Processing and memory requirements during transport of payloads of multiple sizes, across multiple settings are summarized in Fig. 8 and Fig. 9 using mean values with 95% confidence intervals.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

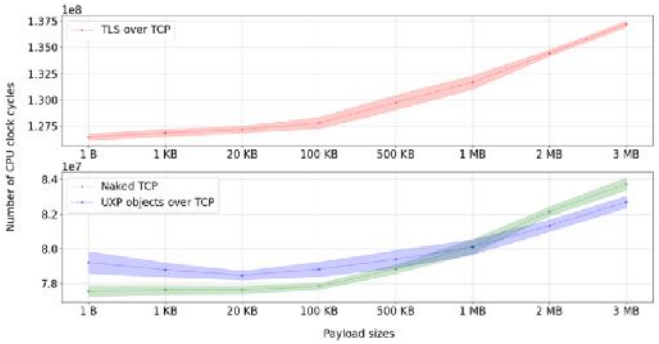


Figure 8. 95% confidence interval plot of number of CPU cycles for various payload sizes during transport.

The upper portion of Fig. 8 shows the number of CPU cycles vs. user payload with a 95% confidence interval when user data is transported via TCP protected with TLS. The lower portion of Fig. 8 shows the same data for plain TCP (user data unprotected), and for plain TCP with user data protected via UXP. From the plots, it is clear that using TLS for security is about 1.6 times more computationally expensive than using UXP objects as the security medium over a plain TCP channel.

As expected, CPU cycles increase steadily as the user data size increases. Surprisingly, the number of CPU cycles required to transport UXP objects via TCP gradually decreases as payload increases as compared to transport of unprotected data over TCP. This may be explained using Fig. 8, which suggests that UXP object creation includes lossless compression of user data.

Regardless, this observation suggests that protecting user data using ICTO technology provides improved performance for data larger than 500 kB.

Thus, UXP is substantially more efficient than TLS in channel and CPU utilization as well as transmit time, particularly for large payloads, and provides benefits for data at-rest as well as in-transit.

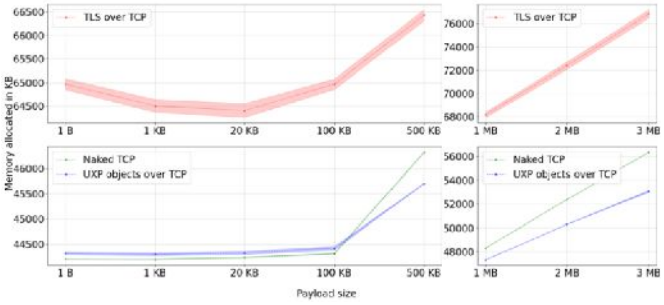


Figure 9. 95% confidence interval plot of amount of memory allocated during transport vs. user data size.

Fig. 9 shows allocated memory vs. user data sizes during transport of user data over plain TCP, over TLS secured TCP, and UXP-secured user data over TCP. All plots employ 95% CI.

As shown in Fig. 9, transporting user data over TLS is the most expensive in terms of allocated/required memory as well as compared to the other two modes of transport. TLS secured data transport is observed to generally require an additional 20 MB of memory compared to unprotected data transport or ICTO protected data transport.

Further, the memory required seems to be lower when UXP objects protecting user data larger than 100 kB are transported over TCP. Contrary to intuitive expectation, UXP protected data transport over TCP is observed to be more memory efficient than transport of raw unprotected data over TCP.

2) Network Performance During Transport: For analyzing the network performance of the UXP technology, two important metrics – total transmit time, and control plane overhead were recorded and statistically analyzed.

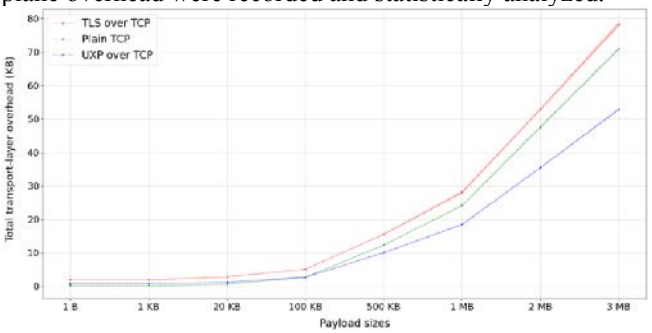


Figure 10. 95% C.I. plots showing control plane overhead vs. user data size. Overhead measurements are from the transport layer and above.

Mean and 95% confidence interval figures of total control plane overhead size for transmission as well as total transmit time for various payload sizes in multiple settings are presented in Fig. 10 and Fig. 11.Fig. 10 indicates that using UXP objects as a means of transporting user data seems to be most efficient in terms of overhead because it has the least amount of control plane overhead during transmission. Using TLS over TCP for secure data transport has at least 50% larger overhead as compared to secure transport of user data using UXP objects. On average, about 1950 bytes of overhead is introduced by TLS over TCP for transporting a single byte of data whereas UXP object and raw data transport over plain TCP introduce only 862 and 184 bytes of overhead respectively, which increases with payload size.

Hence, for transporting a single byte of data, using UXP objects as a means of securing the payload introduces only 50% of the overhead of TLS. For the maximum payload size (3 MB), TLS over TCP requires overhead of about 78 kB whereas using UXP objects over TCP requires 52 kB. Thus, for transporting 3 MB of user data, TLS introduces 1.5 times more network overhead. Further, compressible payloads reduce the size of UXP objects, so that required network overhead is less than 50% as compared with plain TCP. Thus, UXP is substantially more efficient than TLS in channel utilization, particularly for large payloads, and provides benefits for data at-rest as well as in-transit.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

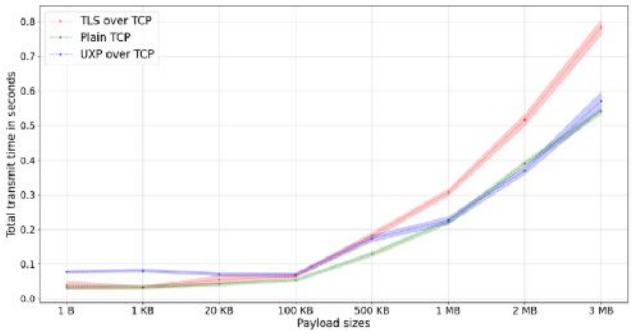


Figure 11. 95% C.I. plots of transmit time vs. payload size. Transmit times are calculated for transporting user data over TCP, user data protected with UXP objects over TCP, and via secure channel using TLS over TCP.

Fig. 11 shows the average transmit times for varying payload sizes for transport of user data in different configurations. As shown in the Fig. 11, the total time for transmission remains below 100 milliseconds for payload sizes up to 100 kB. Time required to transmit 1 byte to 100 kB payloads are highest when transporting UXP objects over TCP and the lowest when transporting raw user data over TCP. However, for payloads larger than 500 kB, UXP protected user data requires less time to transport than other configurations. The plot shows that for higher user payload sizes, transmission time is significantly lower for UXP object transmission over TCP than for that for TLS over TCP. Thus, UXP is substantially more efficient than TLS in transmit time, particularly for large payloads, and provides benefits for data at-rest as well as in-transit.

Although the transmit-time data presented in Fig. 11 were recorded in a controlled network environment, some incoherency is still evident. This may be explained by irregular handling of traffic by the network access point that connects the client and server machines. In addition, the interfaces at the communicating machines (client and server) may also have irregular scheduling/process priority for certain network related processes, affected by services/applications running in the background. Nonetheless, somewhat distinct trends are still observable and are enough to draw meaningful conclusions.

V. BLOCKCHAIN VS ICTO

The results of experiments and data gathering consists of experiments performed on a general-purpose computer running the Linux operating system. The experiments were further categorized into two parts: compressible data, which encompasses experiments conducted with lossless compression techniques on payloads, and incompressible data, which comprises of payloads that could not be compressed using lossless compression techniques.

A. Storage Comparisons

Memory storage is important irrespective of the context of data security applications, and in UXP as well as blockchain implementations. In this section, a general comparison between blockchain technology and UXP is presented, considering that both versions of the blockchain i.e., core blockchain and ledger have the same memory requirements for the experiments.

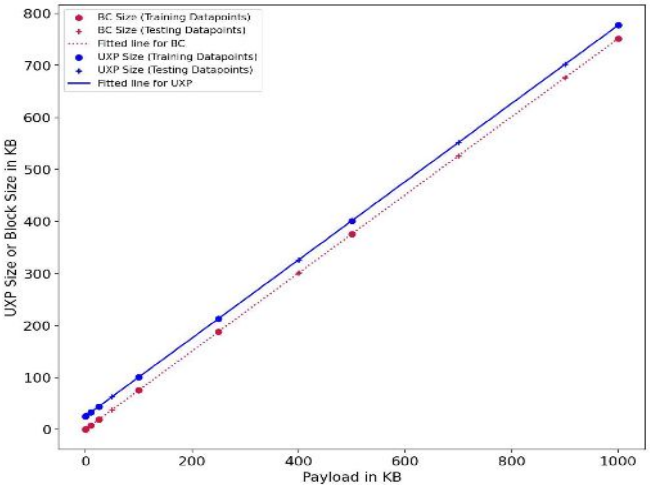


Figure 12. Fitted line for blockchain and UXP storing compressible payload with training and testing datasets.

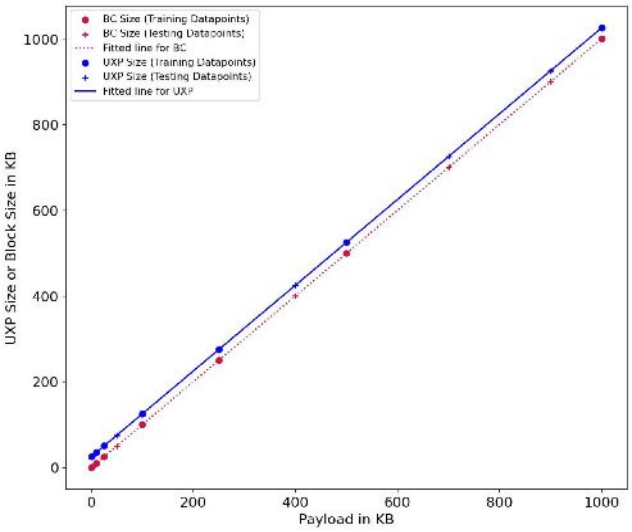


Figure 13. Fitted line for blockchain and UXP storing incompressible payload with training and testing datasets.

Fig. 12 and Fig. 13 summarize experimental analysis regarding the memory storage used by blockchain and UXP for compressible and incompressible payloads. The experimental dataset was used to perform linear curve fitting and generate equations that serve as an approximation for determining the block size and UXP size for various payloads. The relationship between the size of the payload and size of the protected payload was found to be linear.

Payload sizes within specific range of 1Byte to 1MB was used as the training dataset for fitting the linear equations. To assess the accuracy and performance of the equations, a separate testing dataset was prepared. The testing dataset included payload sizes of 25kB, 400kB, 700kB and 900kB for both experiments. It is evident from Fig. 12 and Fig. 13 that the testing datapoints, represented

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

by red dots for blockchain and blue dots for UXP, align closely with the straight line approximated by the training dataset. The confidence interval had a significantly low range, which made it infeasible to illustrate in Fig. 12.

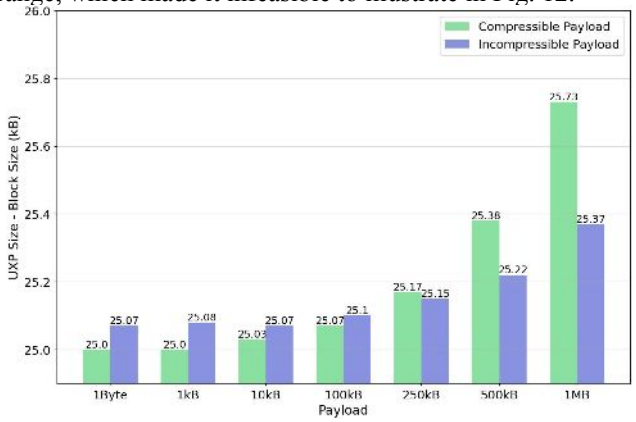


Figure 14. Difference in UXP Size vs Block Size.

Fig. 14 displays the disparity in size between a block and UXP encapsulating compressible and incompressible payloads of the same size. Storing the same size payload in a UXP object vs. a blockchain block requires a mean of 25.20 kB for a compressible payload and 25.15 kB for an incompressible payload.

B. Execution Time

Measuring the time required to store or encapsulate data is crucial for understanding performance of the system especially in applications consisting of endpoints with limited resources. Regarding this context, an experiment was conducted to compare the time required to store payload in blockchain, time required to store data in the ledger, and the time required to encapsulate payload using UXP technology. Fig. 15 shows the time required to store incompressible and compressible payloads in blockchain, encapsulated via UXP, and stored in the ledger respectively.

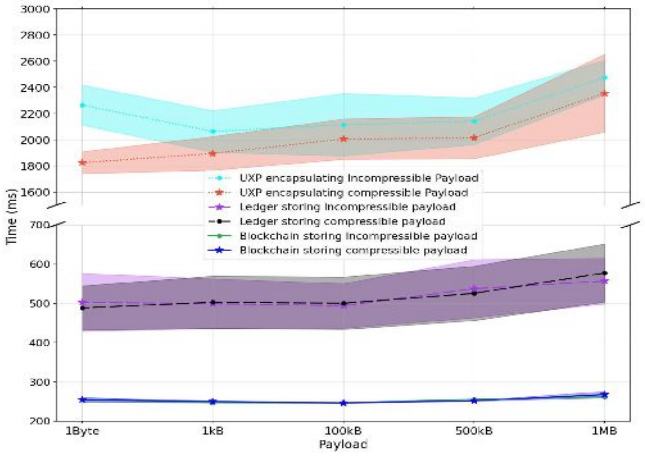


Figure 15. Blockchain vs UXP vs ledger in terms of time required to store payload.

As noted, the time to encapsulate data using UXP is more than 7 times greater than the time it takes to store data in the blockchain. Similarly, the time taken by UXP technology is more than 3 times greater than the time it takes to store data in the ledger. This disparity in execution time between the blockchain and ledger can be attributed to the computational requirements of PoW, which requires calculating a hash with specific difficulty level. This task poses a challenge for devices with limited resources and computing power.

Interestingly, the time difference between storing 1 Byte of payload and 1 MB of payload in the blockchain, on average, was just 7.15ms and 13.25ms. This indicates that the time required to store data in the blockchain is relatively independent of payload size.

Fig. 15 also highlights the narrowness of confidence intervals for time measurements of blockchain, indicating extremely low standard deviation and consistent time requirements regardless of payload size. Conversely, time measurements for UXP experiments have larger standard deviation, reflecting greater variability in the dataset. This could possibly stem from the different encryption and cloaking mechanisms used in UXP encapsulation. To depict this variability and randomness in the experiments, confidence intervals are included in the figures.

C. CPU Clock Cycles

CPU clock cycles directly influence the power consumption, heat dissipation, and resource allocation, in an embedded computing system, making it a crucial parameter to understand.

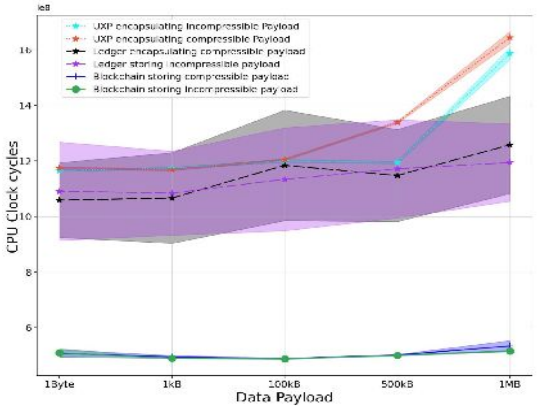


Figure 16. Comparison of CPU clock cycles for blockchain, ICTO, and ledger.

Fig. 16 compares CPU clock cycles for blockchain, UXP, and ledger technologies storing compressible and incompressible payload of various sizes.

As indicated in the Fig. 16, the CPU clock cycles required for UXP encapsulation were 2 times greater than storing data in blockchain. Interestingly, the CPU clock cycles required for UXP, and ledger were nearly equal for compressible and incompressible payloads.

D. Random Access Memory (RAM)

When choosing data protection technology, the balance of security, resource efficiency, and performance is important. The Resident Set Size (RSS), which measures the

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

amount of physical RAM consumed by a process [25, 26] is a useful indicator of memory utilization for blockchain, ledger, and UXP creation on an individual computer.

Fig. 17 shows the average RSS values for compressible and incompressible payloads encapsulated via UXP or stored via blockchain or ledger. This data clearly illustrates that, for the same payloads, RSS required for blockchain, and ledger is 50% greater than RSS required for UXP encapsulation, regardless of payload type.

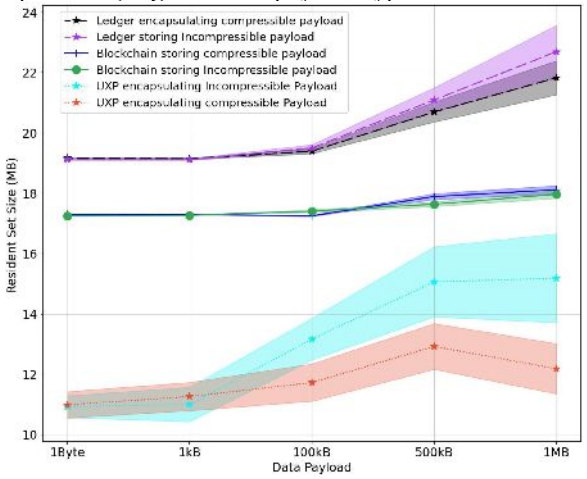


Figure 17. RSS memory comparison for blockchain, ledger, and ICTO.

E. Network Analysis

Network performance for blockchain and ICTO transferred via TCP was evaluated using identical client and server systems via data gathered on the client side. Blockchain payloads were compressed using the LZ77 algorithm and encrypted with ECC encryption prior to transmission. UXP objects were created using compressible payloads.

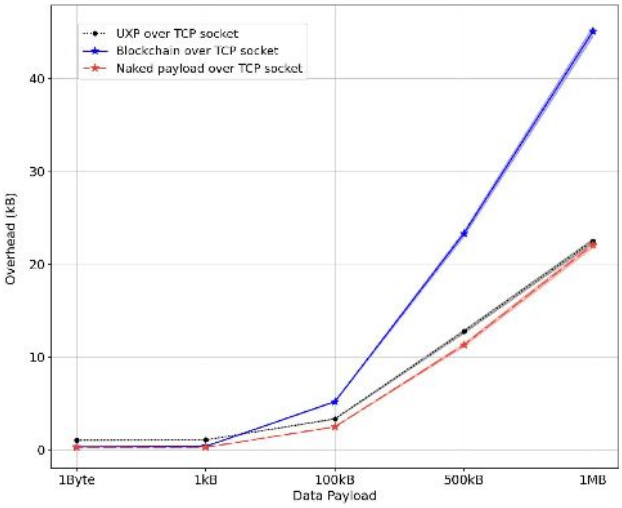


Figure 18. Cumulative Header Size vs Payload Size.

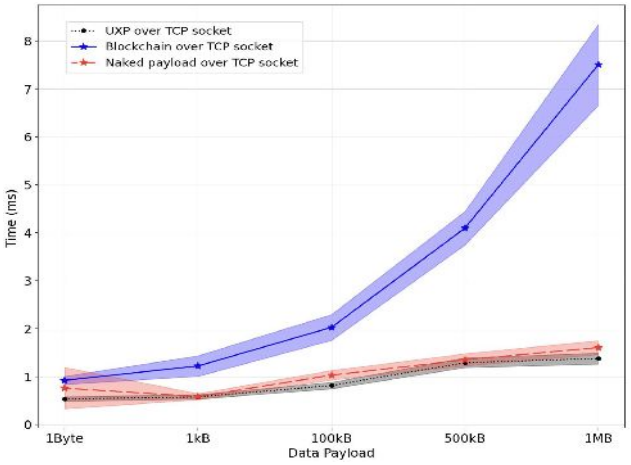


Figure 19. Network Latency vs Payload Size.

The overhead for plain TCP connections (denoted "naked payload" in the figures) and UXP/TCP was found to be similar for larger payloads as seen in Fig. 18. The overhead for blockchain was found to be significantly higher than for plain TCP and UXP/TCP for large payloads. This can be attributed to the relatively larger blockchain blocks vs. plain TCP and UXP/TCP. The encryption overhead of the blockchain blocks also increases with payload. However, the overhead of UXP/TCP is constant with mean value of 25.20 kB.

Fig. 19 illustrates that the total time required to transport UXP/TCP was surprisingly efficient as compared with blockchain, as block transport required as much as 500% of the time required for UXP/TCP.

F. Memory Footprint

The memory footprint of UXP encapsulation compared with encrypting blocks of blockchain with Ascon is also important for IoT applications. The Ascon encryption algorithm is designed for lightweight usage and easy implementation with minimal overhead [29]. "Ascon-128" with key size of 16 bytes was used in this research to encrypt individual blocks of the blockchain. The memory footprint of the resulting block was compared with UXP objects for the same payload.

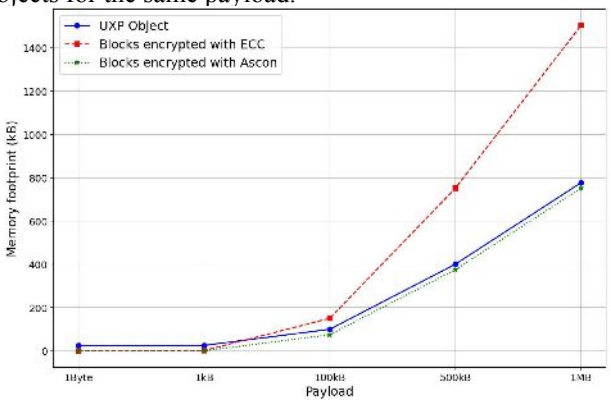


Figure 20. Memory footprint comparison for compressible payloads.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

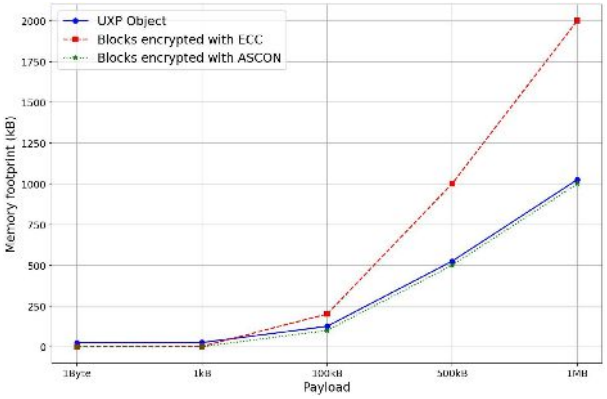


Figure 21. Memory footprint comparison for incompressible payloads.

Fig. 20 and Fig. 21 comparison memory requirements of UXP encapsulation and blocks of blockchain or ledger encrypted by ECC and Ascon algorithms for compressible and incompressible payloads respectively.

From the figures, it is clear that the memory requirements of UXP and Ascon encrypted blocks were almost identical for larger payloads. However, this requirement for blocks encrypted by ECC increased with payload size. Ascon is specifically designed for constrained implementation and low memory footprint. The fact that UXP memory requirements are similar to Ascon is surprising, particularly considering that UXP encapsulation includes multiple layers of encryption.

VI. Security Analysis

A major part of this research is concerned with the investigation of strengths and weaknesses of ICTO technology as a data security measure. Assessment of the security provided by the technology through conventional cryptanalytic techniques as well as modern approaches such as machine learning is a particularly important aspect of the investigation.

Cryptanalysis is the study and practice of analyzing data and cryptosystems for weaknesses and vulnerabilities that may be used to extract useful information [30, 31]. Two main categories of cryptanalysis are symmetric cryptanalysis for symmetric ciphers, and asymmetric/public key cryptanalysis for asymmetric ciphers. Some common symmetric cryptanalytic techniques include brute force attacks, differential cryptanalysis, and algebraic attacks [31-34]. Common techniques for public-key cryptanalysis include factoring attacks and discrete logarithm problem solving [35-37]. Cryptanalysis also relies on the availability of related information, such as the cryptographic algorithm applied, the plaintext used to generate ciphertext, and the ciphertext itself.

Unlike cryptanalysis of symmetric or asymmetric encryption algorithms, where a string of plaintext of a given length results a ciphertext of comparable length, ICTO objects typically have a minimum size of 25 kilobytes regardless of payload size. This makes it difficult to determine where the payload is located within the object. UXP objects were therefore analyzed with the assumption that the plaintext is not available. This approach mimics the role of an attacker who can observe UXP objects in flight or at rest.

As a preliminary measure, a large number of UXP objects created using the same XML policy file and user data were hashed and the resulting hashes were compared with the aim of finding a collision/repetition. None of the resulting hashes matched, which suggests that every UXP object is unique regardless of embedded user data or policy.

A. Frequency Analysis

In classical cryptanalysis, frequency analysis (letter counting) is the study of frequency of occurrences of letters or group of letters in ciphertexts [30-32]. It is one of the most basic and common methods to analyze ciphertexts and has been used for breaking many classical ciphers.

Frequency analysis was performed for a large number of UXP objects (n=500) protecting user payload. Characters were read at each index/offset/position in the UXP file, and the number of occurrences were tabulated. This analysis was performed individually for each object file to obtain a scatter plot, and then repeated for the entire set of files to obtain an average character-frequency plot.

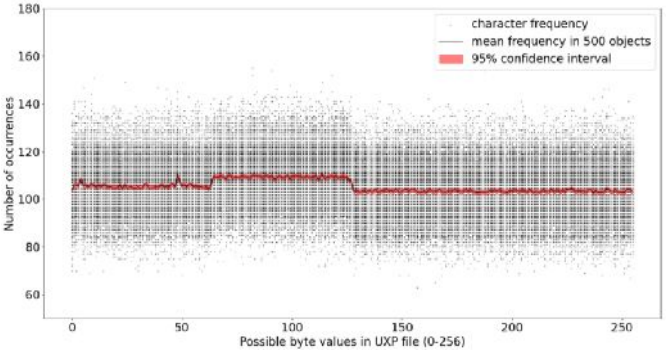


Figure 22. Plots showing byte value of characters vs. frequency of occurrence. The points plotted in black represent character frequency for individual files. The line surrounded by red band represents 95% C.I. plot of average frequency of characters in all 500 object files.

Fig. 22 shows the result of frequency analysis through a scatter plot and a line plot. Each black mark in the scatter plot represents the frequency of corresponding character (expressed in base 10) for a single UXP object. Each character is represented with an 8-bits byte and thus the character pool has 256 possible values. The composite scatter plot contains data for 500 UXP objects.

The red line in Fig. 22 displays the 95% confidence interval for each character value and location. Notably, the scatterplot indicates that regardless of character position, the distribution of values is distinctly non-uniform. This is an unexpected and potentially problematic outcome.

Characters with base-10 values greater than 63 and less than 127 are observed to have higher frequency of occurrence. Also, slightly higher frequency of occurrence near 105 is observed for characters 0 to 63 compared to characters 127 and above which have an average frequency of about 103. This observation counters the intuitive notion

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

that character values in the UXP objects would be uniformly distributed regardless of position.

B. Positional Analysis

As a measure of finding a structure or similarity that may be common in UXP files, coincidence counting – a technique of putting two or more texts side-by-side and recording the number of times and position where identical characters repeat was performed for several UXP objects. It was observed that sets of characters at positions 492 to 494 (3 characters) and at positions 503 to 512 (10 characters) repeat for any UXP object. This observation along with the evidence of non-uniformly distributed character sets led to the analysis of characters based on their position/index in UXP objects.

A scatter plot of character position vs. value using 150 UXP object files all protecting 1 byte of user data is presented in Fig. 23. For simplicity, only the first 700 positions of each UXP file are considered in the plot. It can be observed from the plot that the characters within a range of positions always have a value within a fixed range. This outcome is concerning and non-intuitive for a collection of encrypted/cloaked segments of data.

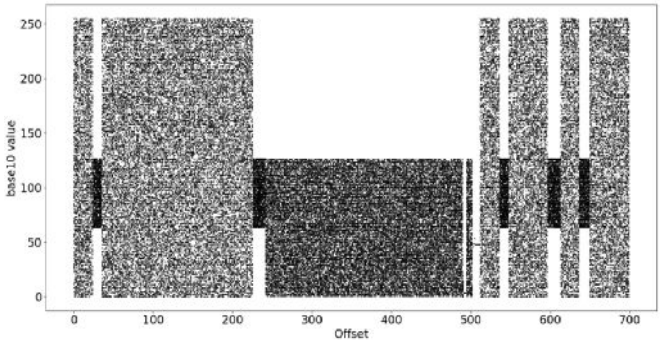


Figure 23. Scatter plot of character position vs. character value for 150 UXP files. The first 700 positions are considered for each object file.

Characters having base-10 values in range of 0-225, 64-127, and 1-126 are observed to be occurring in fixed ranges of positions. This pattern was observed throughout the entirety of UXP files. However, such patterns occur more frequently within the first 650 positions. Although it is not ideal for ciphertexts to contain a fixed set of characters occurring at a given range of positions, UXP objects are observed to have a clearly defined pattern/structure. These patterns were found to be distributed throughout the entirety of every UXP files regardless of variations in policy parameters set in the XML file.

Through careful analysis, it was found that some of the positions in the UXP files always contain a fixed set of characters. This is seen specifically for position 491 which contain only 14 possible characters. Positions 492 to 494 always contain a character with base-10 value of 4. Similarly, positions 503 to 512 always contain the same character. This data validates observations and inferences from frequency analysis because these characters evidently have a higher frequency of occurrence and are always present in several fixed positions of UXP objects. Such a composition of characters observed in UXP objects were found to remain unaffected even when the original XML file (containing policy specification) or payload was modified. This outcome is concerning and counter-intuitive for a collection of data which is encrypted or cloaked.

C. Entropy Analysis

Different levels of "surprise," "uncertainty," or "information" can be expressed in the information theory metric of entropy [31]. Entropy is the expected value or mean of the "information function" of the probability distribution for a set of characters. It indicates the "uncertainty" of a subsequent realization from a random source with a certain probability distribution. For UXP objects, the entropy for each position in the object yields an estimate of the number of bits required to store or transmit the information contained. Alternately, the positional entropy can measure the uncertainty related to a particular character in each position in the UXP object.

To calculate the positional entropy, character occurrences for each position in 500 UXP objects protecting 1 byte of user data were recorded. Kernel density estimation (KDE) [37-40] was employed to approximate the probability density of sample data recorded for each position and used to calculate the positional entropy for the first 700 bytes of each UXP object.

Fig. 24 presents entropy vs. unit index plot for the first 700 bytes of 500 UXP files. The X-axis contains 700 units and 350 units in the bottom and top axes to represent 1-byte units and 2-byte units respectively. Note that each 1-byte unit contains 8-bit characters (256 possible values), and each 2-byte unit contains 16-bit characters (65,536 possible values).

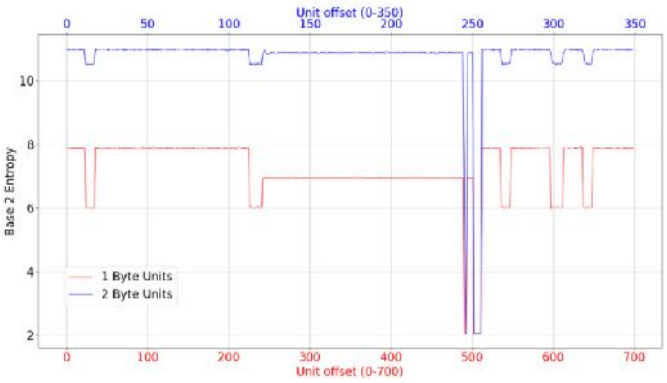


Figure 24. Line plot of entropy vs. unit index. Entropy of 1-byte units (red) and 2-byte units (blue) are shown. The vertical axis is presented using the base-2 entropy values for each unit offset.

The plots in Fig. 24 show large variations in entropies with respect to unit offsets/indices. Dips in entropy values are observed in indices where a fixed subset of characters were found to occur in Fig. 23. Note that the maximum value of base-2 entropy for an 8-bit index is 8 and for a 16-bit index is 16. This indicates the number of bits required for lossless transmission of information carried by a single unit. Most positional entropy values are maximized, but the regularity of "dips" in entropy are concerning.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Dips in entropy are observed in unit indexes where character groups appeared to "cluster" in Fig. 23. For instance, in the 1-bye unit vs. entropy plot, dips in entropy values are seen in the positions/offsets 25 to 36, 227 to 242, and so on. The trends observed in Fig. 24 correlated specifically with the constrained character occurrences observed in Fig. 23.

Lower entropy values indicate that the possible outcomes of a random variable or data source (characters occurring at given positions in this case) have a non-uniform probability distribution. In other words, the probability distribution may be skewed/warped heavily towards a single outcome. This result indicates that UXP objects are "leaking" information at well-defined locations.

In contrast, the positional entropy of ciphertexts generated from random data samples using AES and 32-bit keys is presented in Fig. 25.

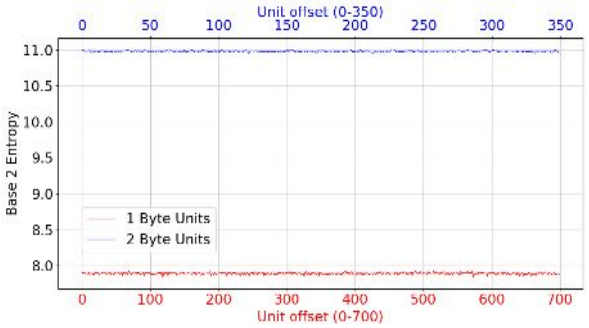


Figure 25. Line plot of entropy vs. unit index. 1-byte and 2-byte units were taken to obtain line plots in red and blue colors respectively.

Unlike the "dips" in positional entropy for UXP objects, AES-encrypted data produces consistent positional entropy of around 7.9 and 11 for 1 byte and 2-byte units respectively, regardless of character position. Thus, characters at each index of the ciphertexts are uniformly distributed and hence do not exhibit any kind of structure, or potential data leakage.

D. Cryptanalysis Using Unsupervised Machine Learning

To analyze patterns in UXP objects using different unsupervised ML techniques, a dataset containing relevant features was created using a large number of objects. The base-10 character values in each position from each object was recorded, and features of each array were extracted. Five characteristics including largest value, smallest value, mean, difference of largest and smallest values, and entropy were recorded or calculated. For simplicity, only the first 700 positions of the objects were considered. As a result, 700 records each with 5 features are present in the final dataset to be submitted to ML algorithms. The accuracy of the features depends upon the number of objects used for calculating each feature value.

Two popular unsupervised ML algorithms, k-means and agglomerative clustering were used to discover clusters of similar orientations in the set of UXP objects.

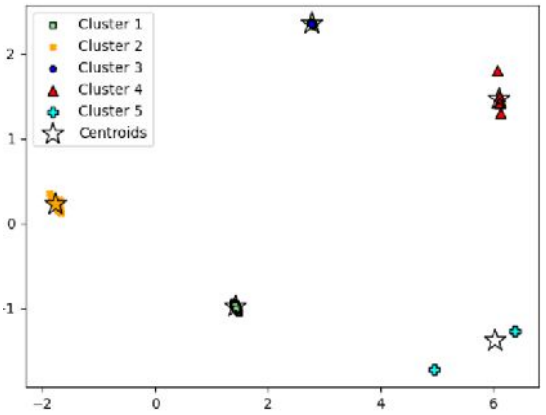


Figure 26. Clusters visible with k-means clustering.

Fig. 26 indicates that 5 distinct clusters with high degree of separation are present in UXP objects. These clusters indicate that distinct patterns or structures are present in every UXP object. The frequency of such patterns/structures in the objects correspond to the number of data points in each cluster. Information of this nature could be relevant for attackers because patterns in data can potentially expose or leak information and may act as the weakest points of attack.

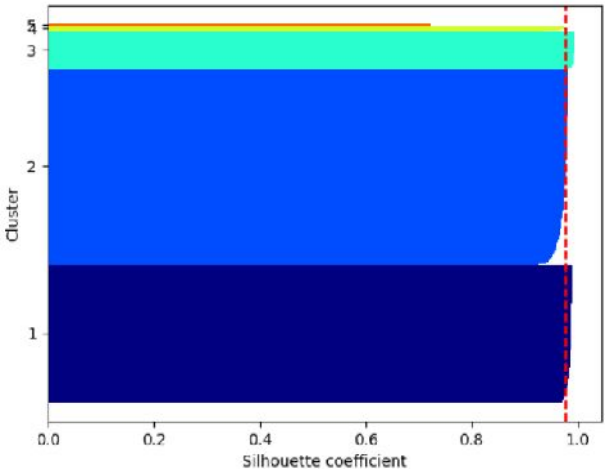


Figure 27. Silhouette plot of resulting k-means clusters.

Fig. 27 shows the silhouette plot obtained using data points from each cluster, which graphically depicts how well data points fit into the clusters to which they have been assigned, as well as the quality of separation. Silhouette values signify good or bad clustering with a range of [-1, 1]. The mean silhouette value in Fig. 27 is roughly 0.9 which is near the maximum value of 1.

Combining observations from Fig. 26 and Fig. 27, it is clear that UXP object have at least 4 distinct patterns of data within them. These inferences also align well with prior observations. As a result, it seems clear that more sophisticated or in-depth machine learning processes may reveal additional information about the UXP objects, or the data payloads contained within them.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

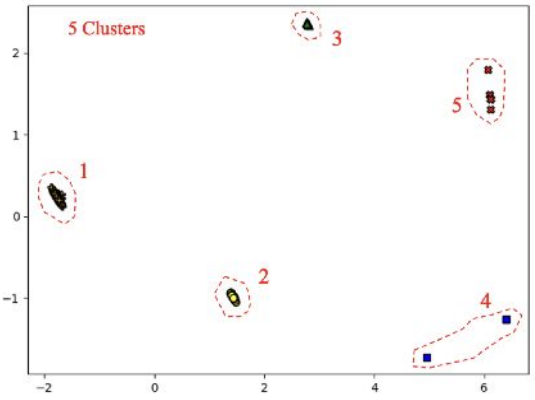


Figure 28. Clusters obtained using agglomerative clustering.

The presence of distinct clusters or groups of data inside UXP objects is further supported by Fig. 28, obtained by using agglomerative clustering. Similar to the results shown in Fig. 26, at least 5 different clusters are observed using this ML method.

Both approaches suggest that UXP objects consist of distinct patterns or structures which are evident when observed with respect to the position or index of characters. This characteristic of UXP objects exposes potential vulnerabilities that may be determined or exploited using more sophisticated ML approaches.

VII. Conclusion

This research explores the potential of the ICTO concept as a data security technology for IoT and general-purpose computing. Comparative performance with popular solutions or technologies is presented, including evaluations with TLS for network transmission and blockchain for storage and security. Based on the outcomes of this study, the ICTO concept could be an alternative to conventional security techniques, especially for IoT. Protecting data at the point of origin or the source itself with access and authorization policies embedded to the data itself seems to be the prime advantage of the ICTO concept. Most security frameworks rely on separate mechanisms to protect information when it is at-rest or in-flight, and these mechanisms have to be tightly coupled to provide comprehensive, end-to-end security. Such integrations are not only complex and costly but also introduce vulnerabilities that can exploited. The ICTO concept is a bold approach that may address many issues associated with data security.

A. Performance

System performance results in Section IV show that it takes at least 1 second of user time and around 47 MB of memory for UXP object creation. Transport of UXP objects over TCP is observed to require higher CPU and memory resources than plain TCP for small payloads. However, for larger payloads, transport of UXP objects was found to require fewer system resources compared to transport of raw user data over plain TCP. System resources used for transporting data via TLS over TCP were significantly greater than UXP object transport, requiring 50% more CPU cycles and 20 MB more memory. Automatic lossless compression of text payloads resulted in objects with total size smaller than the original payload.

The transport of user data via TLS/TCP required significantly greater transmit time and control plane overhead than the other two approaches, regardless of the payload size. In comparison, UXP/TCP required slightly more time and overhead for small payloads but outperformed other methods for large payloads.

B. Blockchain vs ICTO

Scalability is a key consideration in data security applications, and technologies such as blockchain have clear scalability issues because of the cumulative nature of the chain. In contrast, ICTO (as realized as UXP) creates an independent object for a given payload with overlapping layers of security and relatively constant overhead.

In terms of memory and storage utilization and clock cycle requirements, ICTO provides a substantial, relatively deterministic outcome. This result is in stark contrast with an implementation of blockchain in a modified distributed ledger with proof-of-work. This suggests that ICTO has the potential to replace the complex distributed ledger technologies employed in various applications.

Table 1 summarizes critical system performance criteria including literature survey discussion and experimental results using a modified Likert Scale [40] or Mean Opinion Score (MOS) [41] to quantify and rank certain qualitative results.

TABLE 1. Comparison of blockchain, ledger, and UXP

CRITERIA	Blockchain	Distributed Ledger	UXP
Complexity	5[e]	2[b]	3[b]
Single Point of Failure	1[b]	5[a]	1[b]
Scalability	3[a]	3[b]	4[a]
RAM Efficiency	4[a]	2[b]	5[a]
Cost	4[a]	2[b]	2[b]
CPU Efficiency	5[e]	3[b]	3[b]
Validation	3[c]	4[d]	5[e]
Trust	2[c]	4[d]	5[e]
Transparency	2[c]	3[d]	5[e]
Data Security	2[f]	3[f]	5[e]
Vulnerability	2[c]	2[d]	5[e]
Average	3	3	4
Scale and Legend	1 - Very poor, 2- Poor, 3 - Fair, 4 - Good, 5- Very Good a = Simple, eliminates, efficient, easy, scalable, low cost b = Complex, cannot eliminate, inefficient, difficult, high cost c = Central authority d = Trustless, multiple nodes, consensus mechanism e = User access control, multiple nested encryptions, owner ruleset f = No inherent security, depends on external security mechanisms		Best Performance

C. Security of ICTO

Basic cryptanalysis techniques including frequency analysis, index-of-coincidence and positional entropy revealed potential patterns or structures within UXP objects. Frequency analysis of UXP objects indicates that some characters occur more frequently than others and are not uniformly distributed for every position in UXP objects.

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Some of the positions were found to always contain characters from a fixed subset of possible characters, and positional entropy analysis revealed that UXP objects follow a fixed structure or pattern to store information within them. The presence of consistent structures throughout UXP objects indicates that they are leaking information that may serve as a starting point for a more sophisticated attack.

Unsupervised ML approaches confirmed and reiterated the fact that patterns are present in the UXP objects. ML algorithms were successful in finding distinct clusters with high degree of separation. The clusters indicate that at least 4 different patterns or structures can be found in the first 700 bytes of every UXP object.

D. Summary

UXP as an implementation or realization of ICTO technology is a complete package that focuses on data security with enhanced protection schemes. The UXP instantiation of ICTO is a proprietary implementation. As such, specific implementation details are not available for evaluation. However, as an instance of a new class of data security techniques, the contrasts and comparisons with similar and enabling technologies is valuable. Additional research through cryptanalysis is necessary to comprehend the extent of data security actually provided by the UXP implementation of ICTO. But what's clear is that it is a ready-to-use approach to data security which intrinsically supports lossless compression for implementation efficiency. It is extremely useful in cases where security and scalability are vital. Based on the results and inferences drawn from this work, the ICTO concept indeed has a potential to be a useful technology for securing data in IoT as well as general purpose computing.

Even though the UXP implementation of ICTO which was leveraged in this research may not be fully optimized, the performance statistics are compelling, especially when evaluated in context with conventional in-transit data protection schemes. For use cases in IoT, while the findings indicate the overhead of UXP or ICTO is not optimal, the constant overhead even for larger payloads does show potential. The promising trajectory of ICTO with its unique data security approach, and scalability model, justifies further exploration in IoT domain. Further research to optimize ICTO could prove beneficial especially for IoT where securing data right from the source is often difficult.

VIII. Future Work

Clearly, a wider range of ML techniques and exploration of deep learning methodologies could prove to be fruitful in the analysis of ICTO. This work has shown that basic ML methods can detect multiple structures within the secure objects, but the importance of these structures in data leakage is unclear.

Experimental analysis of the ICTO implementation suggests that it is also worth considering an open-source implementation, which could provide notable advantages including transparency, auditability, and interoperability in applications where data security is a critical requirement.

IX. References

[1] B. Thapa, B. Sharma, and S. McClellan, "Comparative Performance of TCP and MQTT," in Proc. 18th Int'l Conf. on Digital Telecom (ICDT 2023), pp.10-14, Venice, Italy, Apr. 2023. ISBN: 978-1-68558-034-6

[2] H. Dai, J. Xu, and Q. Li, "Enterprise Cloud Computing Adoption: An Empirical Study of Factors Influencing Adoption," Info. Sys. Frontiers, vol. 17, no. 2, pp. 243-257, 2015.

[3] T. Erl, Z. Mahmood, and R. Puttini, "Cloud Computing: Concepts, Technology & Architecture," Prentice Hall, 2013.

[4] M. Armbrust, A. Fox, et al., "A View of Cloud Computing," Comm. Of the ACM, v.53, n.4, pp. 50-58, 2010.

[5] C. Wang, K. Ren, W. Lou, and J. Li, "Toward publicly auditable secure cloud data storage services," in IEEE Network, vol. 24, no. 4, pp. 19–24, 2010.

[6] L. Wei, H. Zhu, et.al., "Security and privacy for storage and computation in cloud computing," Info. Sciences, vol. 258, pp. 371–386, 2014.

[7] S. Aldossary and W. Allen. "Data security, privacy, availability and integrity in cloud computing: Issues and current solutions." Int'l J. Adv. Comp. Sci. Appl. 7.4, 2016.

[8] Cloud computing-The business perspective. Decision support systems, vol. 51, pp. 176 – 189, 2011.

[9] S. V. Kartalopoulos, "Differentiating Data Security and Network Security," in IEEE Int'l Conf. Comm., Beijing, China, pp. 1469-1473, 2008.

[10] S. Turner, "Transport layer security," IEEE Internet Comp. vol. 18, n.6, pp. 60-63, 2014.

[11] K. H. Mohammed, A. Hassan, and D. Y. Mohammed, "Identity and Access Management System: a Web-Based Approach for an Enterprise," 2018.

[12] E. Biham and O. Dunkelman, "Cryptanalysis of the A5/1 GSM stream cipher," in Prog. Cryptology—INDOCRYPT 2000: 1st Int'l Conf. Crypt. In India, Calcutta, India, December 10–13, 2000. V.1. Springer, 2000.

[13] N. Novaes Neto, S. Madnick, M. G. de Paula, and N. M. Borges, "A case study of the Capital One data breach," 2020.

[14] W. Connelley and B. Gudaitis, "Architecture containing embedded compression and encryption algorithms within a data file," U.S. Patent 20030204718A1, Oct. 10, 2003.

[15] R. Patawaran and G. Chapman, "Secure storage and retrieval of confidential information," U.S. Patent 20110252480A1, Jul. 16, 2013.

[16] "What is Blockchain Technology", IBM Blockchain https://www.ibm.com/topics/what-is-blockchain

[17] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System". 2009. Bitcoin.org

[18] H.-N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for Internet of Things: A Survey," IEEE Internet Things J., vol. 6, no. 5, pp. 8076–8094, Oct. 2019, doi: 10.1109/JIOT.2019.2920987.

[19] M. Samaniego, U. Jamsrandorj, and R. Deters, "Blockchain as a Service for IoT," in 2016 IEEE Int'l Conf. Internet of Things (iThings), Green Comp. Comm.(GreenCom), Cyber, Phys. Social Comp. (CPSCom), and Smart Data (SmartData), Dec. 2016, pp. 433–436. doi: 10.1109/iThings-GreenCom-CPSCom-SmartData.2016.102.

[20] P. Arul and S. Renuka, "Blockchain technology using consensus mechanism for IoT- based e-healthcare system," IOP Conf. Series: Materials Sci. and Eng., Feb. 2021. doi: 10.1088/1757-899X/1055/1/012106.

[21] S. Kim, G. Chandra Deka, and P. Zhang, "Role of Blockchain Technology in IoT Applications" in Advances in Computers, v. 115, Academic Press, Sep. 2019. [Online]. Available:

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

https://learning.oreilly.com/library/view/role-of-blockchain/9780128171929/S0065245819300397.xhtml

[22] R. Li, T. Song, B. Mei, H. Li, X. Cheng, and L. Sun, "Blockchain for Large-Scale Internet of Things Data Storage and Protection," IEEE Trans. Serv. Comput., vol. 12, no. 5, pp. 762–771, Sep. 2019, doi: 10.1109/TSC.2018.2853167.

[23] G. Smith, M. Weed, D. Fischer, and E. Ridenour, "System and methods for using cipher objects to protect data," U.S. Patent 11,093,623 B2, 2021.

[24] Dierks and E. Rescorla. "The Transport Layer Security (TLS) Protocol Version 1.2." IETF RFC 5246, 2008.

[25] Canonical, "Forkstat - a tool to show process activity", https://manpages.ubuntu.com/manpages/xenial/en/man8/forkstat.8.html (accessed May 29, 2023).

[26] "perf stat," Linux manual page, https://man7.org/linux/man-pages/man1/perf-stat.1.html (accessed May 29, 2023).

[27] alexmgr, "tinyec." May 13, 2023. Accessed: May 20, 2023. [Online]. Available: https://github.com/alexmgr/tinyec

[28] "scipy.stats.t — SciPy v1.10.1 Manual." https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html (accessed Jun. 10, 2023).

[29] "ASCON." Accessed: Apr. 25, 2023. [Online]. Available: https://ascon.iaik.tugraz.at/index.html

[30] B. Carter and T. Magoc, "Classical ciphers and cryptanalysis," in Space 1000, vol. 1, pp. 1, 2007.

[31] J. Dooley, "History of Cryptography and Cryptanalysis: Codes, Ciphers, and Their Algorithms," in History of Computing, Springer 2018.

[32] L. Knudsen and M. Robshaw, "Brute force attacks," in The Block Cipher Companion, 1st ed., New York, NY, USA: Springer, 2011, pp. 95-108.

[33] E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," J. Crypt., vol. 4, no.1, pp. 3-72, 1991.

[34] N. Courtois and W. Meier, "Algebraic attacks on stream ciphers with linear feedback," in Proc. Int'l Conf. Theory and Appl. Crypto. Tech. - EUROCRYPT 2003, Warsaw, Poland, May 4-8, 2003, v.22, pp. 345-359, Springer, 2003.

[35] D. J. Bernstein et al., "Factoring RSA keys from certified smart cards: Coppersmith in the wild," in Proc. Int. Conf. Theory and Appl. of Crypt. and Info. Sec. - ASIACRYPT 2013, Bengaluru, India, December 1-5, 2013, vol.19, pp. 341-360, Springer 2013.

[36] K. S. McCurley, "The discrete logarithm problem," in Proc. Symp. Applied Math, vol. 42, pp. 49-74, 1990.

[37] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," MIT press, pp. 73, 2016.

[38] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," in Annals of Math. Stat., vol. 27, no. 3, pp. 832-837, 1956.

[39] E. Parzen, "On estimation of a probability density function and mode," in Annals of Math. Stat., vol. 33, no. 3, pp. 1065-1076, 1962.

[40] R. Likert, "A Technique for the Measurement of Attitudes. Archives of Psychology," 140, 1–55, 1932.

[41] R. Streijl, S. Winkler and D. Hands. "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives." Multimedia Sys., vol. 22, pp. 213–227, 2016. https://doi.org/10.1007/s00530-014-0446-1

2023, © Copyright by authors, Published under agreement with IARIA - www.iaria.org