ICN 2023

The Twenty-Second International Conference on Networks

ISBN: 978-1-68558-036-0

April 24th – 28th, 2023

Venice, Italy

ICN 2023 Editors

Eugen Borcoci, University Politehnica of Bucharest, Romania

ICN 2023

Forward

The Twenty-Second International Conference on Networks (ICN 2023), held between April 24th and April 28th, 2023, continued a series of events organized by and for academic, research and industrial partners.

We solicited academic, research, and industrial contributions. We welcomed technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard fora or in industry consortia, survey papers addressing the key problems and solutions on any of the above topics short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the ICN 2023 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to ICN 2023. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the ICN 2023 organizing committee for their help in handling the logistics of this event.

We hope that ICN 2023 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of networks.

ICN 2023 Chairs

ICN 2023 Steering Committee

Eugen Borcoci, University Politehnica of Bucharest, Romania
Pascal Lorenz, University of Haute Alsace, France
Nicola Ciulli, Nextworks, Italy
Shintaro Mori, Fukuoka University, Japan
Yantong Wang, Shandong Normal University, China
Muath Obaidat, City University of New York, USA

ICN 2023 Publicity Chairs

Laura Garcia, Universitat Politecnica de Valencia, Spain
Javier Rocher Morant, Universitat Politecnica de Valencia, Spain

ICN 2023
Committee

ICN 2023 Steering Committee

Eugen Borcoci, University Politehnica of Bucharest, Romania
Pascal Lorenz, University of Haute Alsace, France
Nicola Ciulli, Nextworks, Italy
Shintaro Mori, Fukuoka University, Japan
Yantong Wang, Shandong Normal University, China
Muath Obaidat, City University of New York, USA

ICN 2023 Publicity Chairs

Laura Garcia, Universitat Politecnica de Valencia, Spain
Javier Rocher Morant, Universitat Politecnica de Valencia, Spain

ICN 2023 Technical Program Committee

Luis F. Abanto-Leon, Technische Universität Darmstadt, Germany
Qammer H. Abbasi, University of Glasgow, UK
Khelil Abdelmajid, Landshut University of Applied Sciences, Germany
Alireza Abdollahpouri, University of Kurdistan, Sanandaj, Iran
Abdelmuttlib Ibrahim Abdalla Ahmed, University of Malaya, Malaysia
Ahmedin Mohammed Ahmed, FDRE Ministry of Innovation and Technology (MInT), Ethiopia
Francisco Airton Silva, Federal University of Piauí, Brazil
Sami Marzook Alesawi, King Abdulaziz University | Faculty of Computing and Information Technology at Rabigh, Saudi Arabia
Madyan Alsenwi, Kyung Hee University - Global Campus, South Korea
Reem Alshahrani, Kent State University, USA
Cristian Anghel, Politehnica University of Bucharest, Romania / Pentalog, France
Imran Shafique Ansari, University of Glasgow, Scotland, UK
Andrés Arcia-Moret, Xilinx, Cambridge, UK
Suayb S. Arslan, MEF University, Turkey
Mohammed A. Aseeri, King Abdulaziz City of Science and Technology (KACST), Kingdom of Saudi Arabia
Aishwarya Asesh, Adobe, USA
Michael Atighetchi, BBN Technologies, USA
Jocelyn Aubert, Luxembourg Institute of Science and Technology (LIST), Luxembourg
Marco Aurélio Spohn, Federal University of Fronteira Sul, Brazil
Omran Ayoub, Politecnico di Milano, Italy
Benjamin Aziz, University of Portsmouth, UK
Alvaro Barradas, University of Algarve, Portugal
Chemseddine Benkalfate, Quartz Laboratory of ENSEA, France
Luis Bernardo, NOVA University of Lisbon, Portugal
Robert Bestak, Czech Technical University in Prague, Czech Republic
Lucas Bondan, Research and Development Center in Information and Communication Technology (CTIC) of the Brazilian National Research and Educational Network (RNP), Brazil

Eugen Borcoci, University Politehnica of Bucharest, Romania
Fernando Boronat Seguí, Universitat Politecnica de Valencia-Campus de Gandia, Spain
Radoslav Bortel, Czech Technical University in Prague, Czech Republic
Marilisa Botte, Federico II University of Naples, Italy
Christos Bouras, University of Patras, Greece
An Braeken, Vrije Universiteit Brussel, Belgium
Arslan Brömme, Vattenfall GmbH, Berlin, Germany
Claudia Canali, University of Modena and Reggio Emilia, Italy
Batyr Charyyev, Stevens Institute of Technology, USA
Aizaz Chaudhry, Carleton University, Canada
Hao Che, University of Texas at Arlington, USA
Marc Cheboldaeff, Cognizant Technology Solutions, Germany
Adil Chekati, University of Tunis El Manar, Tunisia
Jundong Chen, Dickinson State University, USA
Sixia Chen, Central Connecticut State University, USA
Yitao Chen, Qualcomm, USA
Yuxuan Chen, Shandong University, China
Nicola Ciulli, Nextworks, Italy
Bernard Cousin, University of Rennes 1, France
Jorge Crichigno, College of Engineering and Computing | University of South Carolina, USA
Monireh Dabaghchian, George Mason University, USA
Sofiane Dahmane, University of Laghouat, Algeria
Abdulhalim Dandoush, ESME-Sudria engineering school, France
Susumu Date, Cybermedia Center - Osaka University, Japan
Babu R. Dawadi, Tribhuvan University,Nepal
Declan Delaney, University College Dublin, Ireland
Margot Deruyck, Ghent University - IMEC - WAVES, Belgium
Amir Djenna, University of Constantine, Algeria
Pengyuan Du, Facebook Inc., USA
Salahaldeen Duraibi, Jazan University, Saudi Arabia
Zakaria Abou El Houda, University of Montreal, Canada
Basem ElHalawany, Shenzhen University, China / Benha University, Egypt
Gledson Elias, Federal University of Paraíba (UFPB), Brazil
Levent Ertaul, California State University, East Bay, USA
Davide Ferraris, University of Malaga, Spain
Mário Ferreira, University of Aveiro, Portugal
Adriano Fiorese, Santa Catarina State University (UDESC), Brazil
Mathias Fischer, Universität Hamburg, Germany
Edelberto Franco Silva, Universidade Federal de Juiz de Fora, Brazil
Valerio Frascolla, Intel Deutschland GmbH, Neubiberg, Germany
Marco Furini, University of Modena and Reggio Emilia, Italy
Yu Gao, University of St. Thomas, USA
Yun Gao, Nanjing University of Posts and Telecommunications, China
Gourab Ghatak, IIIT-Delhi, India
Saptarshi Ghosh, London South Bank University, UK
Marco Giordani, University of Padova, Italy
Rita Girao-Silva, University of Coimbra & INESC Coimbra, Portugal
Shay Gueron, University of Haifa / Amazon Web Services, Israel

Tina Gui, Anheuser-Busch InBev, Belgium
Tibor Gyires, Illinois State University, USA
Nguyen Tri Hai, Chung-Ang University, Korea
Talal Halabi, University of Winnipeg, Canada
Muhammad Hanif, Hanyang University / Seoul National University of Science and Technology, South Korea
Luoyao Hao, Columbia University, USA
Esteve J. Hassan, Mohawk College of Applied Arts and Technology, Canada
William "Chris" Headley, Virginia Tech National Security Institute | Virginia Polytechnic Institute & State University, USA
Enrique Hernández Orallo, Universidad Politécnica de Valencia, Spain
Markus Hofmann, Nokia Bell Labs, USA
M. Reza Hoseinyfarahabady, University of Sydney, Australia
Md Shafaeat Hossain, Southern Connecticut State University, USA
Wen-Chen Hu, University of North Dakota, USA
Fatima Hussain, Ryerson University / Royal Bank of Canada, Toronto, Canada
Dragos Ilie, Blekinge Institute of Technology (BTH), Sweden
Pasquale Imputato, University of Naples Federico II, Italy
Gal Itzhak, Technion - Israel Institute of Technology, Israel
Faouzi Jaidi, University of Carthage - Higher School of Communications of Tunis & National School of Engineers of Carthage, Tunisia
Yong Jin, Tokyo Institute of Technology, Japan
Omprakash Kaiwartya, Nottingham Trent University, UK
Faouzi Kamoun, ESPRIT School of Engineering, Tunisia
Kyungtae Kang, Hanyang University, Korea
Binayak Kar, National Taiwan University of Science and Technology, Taiwan
Erdem Karayer, Ege University, Turkey
Kallol Krishna Karmakar, University of Newcastle, Australia
Andrzej Kasprzak, Wrocław University of Science and Technology, Poland
Sokratis K. Katsikas, Norwegian University of Science and Technology, Norway
Hakima Khelifi, Beijing Institute of Technology, China
Pinar Kirci, Istanbul University-Cerrahpasa, Turkey
Sondes Ksibi, University of Carthage | Higher School of Communications of Tunis, Tunisia
Rafael Kunst, University of Vale do Rio dos Sinos (UNISINOS), Brazil
Christo Kurisummoottil-Thomas, Eurecom, France
Mohammed Laroui, Djillali Liabes University, SBA, Algeria & Paris University, France
Vincent Latzko, Technische Universität Dresden, Germany
Riccardo Lazzeretti, Sapienza University of Rome, Italy
Piotr Lechowicz, Wroclaw University of Science and Technology, Poland
Chi-Han Lee, Academia Sinica, Taiwan
Gyu Myoung Lee, Liverpool John Moores University, UK
Jonathan Lejeune, Sorbonne Université | Inria, France
Peilong Li, Elizabethtown College, USA
Kiho Lim, William Paterson University of New Jersey, USA
Lars Lindner, Universidad Autónoma de Baja California, Mexico
Yuchen Liu, North Carolina State University, USA
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Pascal Lorenz, University of Haute Alsace, France

Chitradeep Majumdar, University of Liverpool, UK
Zoubir Mammeri, IRIT - Paul Sabatier University, Toulouse, France
D. Manivannan, University of Kentucky, USA
Christopher Mansour, Mercyhurst University, USA
Tagleorge Marques Silveira, Universidade de Aveiro, Portugal
Sreekar Marupaduga, IEEE, USA
Antonio Matencio-Escolar, University of the West of Scotland (UWS), UK
Thijs Metsch, Intel Deutschland GmbH, Germany
Umair Mohammad, Florida International University, USA
Ayan Mondal, Univ. Rennes | Inria | CNRS | IRISA, France
Jordi Mongay Batalla, Warsaw University of Technology, Poland
Shafika Showkat Moni, University of Kentucky, Lexington, USA
Mario Montagud, University of Valencia & i2CAT Foundation, Spain
Manuela Montangero, Università di Modena e Reggio Emilia, Italy
Marcos Morgenstern, Federal Institute of Education, Science and Technology Farroupilha (IFFar), Rio Grande do Sul, Brazil
Shintaro Mori, Fukuoka University, Japan
Ioannis Moscholios, University of Peloponnese, Greece
Susanna Mosleh, National Institute of Standard and Technology (NIST), USA
Mort Naraghi-Pour, Louisiana State University, USA
Giovanni Nardini, University of Pisa, Italy
Galymzhan Nauryzbayev, Nazarbayev University, Kazakhstan
Anselme Ndikumana, Kyung Hee University, South Korea
Hien Quoc Ngo, Queen’s University Belfast, UK
Quang Ngoc Nguyen, Waseda University, Tokyo, Japan
Uyen Trang Nguyen, York University, Toronto, Canada
Maciej Nikodem, Wroclaw University of Science and Technology, Poland
Boubakr Nour, Beijing Institute of Technology, China
Muath Obaidat, City University of New York, USA
Olusola Odeyomi, Wichita State University, USA
Lidia Ogiela, AGH University of Science and Technology, Krakow, Poland
Marek R. Ogiela, AGH University of Science and Technology, Krakow, Poland
Urszula Ogiela, AGH University of Science and Technology, Krakow, Poland
Timothy O'Shea, Virginia Tech University & DeepSig Inc., USA
Madhurananda Pahar, University of Stellenbosch, South Africa
Constantin Paleologu, University Politehnica of Bucharest, Romania
Shashi Raj Pandey, Kyung Hee University - Global Campus, South Korea
Rahul Paropkari, Sprint, USA
Edoardo Persichetti, Florida Atlantic University, USA
Ferdous Pervej, North Carolina State University, Raleigh, USA
Vitaly Petrov, Nokia Bell Labs, Helsinki, Finland
Paulo Pinto, Universidade Nova de Lisboa, Portugal
Agnieszka Piotrowska, Silesian University of Technology, Poland
Ravi Prakash, TU Delft, Netherlands
Cong Pu, Marshall University, USA
Abdellatif Rahmoun, Ecole Supérieure en Informatique, Sid Bel-Abbes, ESI-SBA, Algeria
Shankar Raman, Indian Institute of Technology Madras, India
Kurdman Rasol, Universitat Politècnica de Catalunya (UPC), Spain

Adib Rastegarnia, Purdue University, USA
Claudina Rattaro, Universidad de la República, Montevideo, Uruguay
Danda B. Rawat, Howard University, USA
Yenumula B. Reddy, Grambling State University, USA
Ghaya Rekaya, Telecom Paris, France
Eric Renault, IMT-TSP, France
Farhad Rezazadeh, Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Spain
Imad Rida, University of Technology of Compiegne, France
Elisa Rojas, University of Alcala, Madrid, Spain
Gerardo Rubino, INRIA, Rennes, France
Rukhsana Ruby, Shenzhen University, China
Marina Ruggieri, University of Roma Tor Vergata, Italy
Abdulhakim Sabur, Arizona State University, USA
Amit Samanta, IIT Kharagpur, India / Max Planck Institute for Software Systems, Germany
Rodrigo Sanches Miani, Universidade Federal de Uberlândia, Brazil
Masahiro Sasabe, Graduate School of Science and Technology - Nara Institute of Science and Technology, Japan
Qi Shi, Liverpool John Moores University, UK
Yuankun Shi, Intel, China
Megumi Shibuya, The University of Electro-Communications, Japan
Kostas Stamos, University of Patras, Greece
Cristian Lucian Stanciu, University Politehnica of Bucharest, Romania
Prasad Talasila, Aarhus University, Denmark
Ashis Talukder, Kyung Hee University, South Korea / University of Dhaka, Bangladesh
Sudeep Tanwar, Institute of Technology | Nirma University, Ahmedabad, India
Giorgio Terracina, Università della Calabria, Italy
Vasileios Theodorou, Intracom Telecom, Greece
Florian Tschorsch, Technische Universität Berlin, Germany
Eirini Eleni Tsiropoulou, University of New Mexico, USA
Abu Barkat Ullah, University of Canberra, Australia
Dalton C. G. Valadares, IFPE, Brazil
Rob van der Mei, Centre for Mathematics and Computer Science (CWI), Amsterdam, Netherlands
Costas Vassilakis, University of the Peloponnese, Greece
Quoc-Tuan Vien, Middlesex University, UK
César Viho, IRISA - ISTIC/Université Rennes 1, France
Calin Vladeanu, University Politehnica of Bucharest, Romania
Dmitriy Volkov, eQualit.ie, Canada
Xianzhi Wang, University of Technology Sydney, Australia
Yantong Wang, Shandong Normal University, China / King's College London, UK
Bernd E. Wolfinger, University of Hamburg, Germany
Longfei Wu, Fayetteville State University, USA
Hong Yang, Nokia Bell Labs, Murray Hill, USA
Daqing Yun, Harrisburg University, USA
Habib Zaidi, Geneva University Hospital | Geneva University, Department of Radiology & Medical Informatics, Switzerland
Mariusz Żal, Poznan University of Technology, Poland
Pavol Zavarsky, Framatome, Canada
Sherali Zeadally, University of Kentucky, USA

Tengchan Zeng, Virginia Tech, Blacksburg, USA
Shengzhi Zhang, Boston University | MET College, USA
Shuai Zhang, Aalborg University, Denmark
Zhu Zhengyu, Zheng Zhou University, China
Taieb Znati, University of Pittsburgh, USA
Doukha Zouina, University of Science and Technology Houari-Boumediene (USTHB), Algeria

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Higher-order Statistics of Series of Packet Losses 1
Blazej Adamczyk and Andrzej Chydzinski

A Study on Zero-touch-design Information-centric Wireless Sensor Networks 7
Shintaro Mori

Personalized Protocols for Data Hiding in Cloud-to-Things Applications 10
Marek R. Ogiela and Urszula Ogiela

Use Cases and 6G Architecture: New Needs and Challenges 12
Francine Cassia de Oliveira, Gustavo Iervolino de Morais, Joao Victor Menino e Silva, and Ramon Magalhaes Nogueira

Strategies for Minimization of Energy Consumption in Data Centers 17
Simon Friis, Line Larsen, and Sarah Ruepp

Higher-order Statistics of Series of Packet Losses

Blazej Adamczyk and Andrzej Chydzinski
Department of Computer Networks and Systems
Silesian University of Technology
Gliwice, Poland
email: blazej.adamczyk@polsl.pl, andrzej.chydzinski@polsl.pl

***Abstract*—We analyze higher-order statistics of the length of the series of consecutive packet losses at the router's output buffer. So far, only the average length of the series of losses has been studied, in the context of the quality of real-time multimedia transmissions. In this paper, we compute the coefficient of variation, skewness and excess kurtosis of the length of the series of losses, using a complex traffic model. Then we study the influence of properties of the queue and the traffic on these higher-order statistics. In particular, we study the impact of the autocorrelation of packet interarrival times, the batch structure of the traffic, the buffer capacity and the load of the queue on the coefficient of variation, skewness and kurtosis and discuss their potential impact on the quality of multimedia transmissions perceived by end users.**

***Index Terms*—Internet; buffer overflows; series of losses; higher-order statistics**

I. INTRODUCTION

One of the consequences of the best-effort design of the Internet are packet losses, which happen at the router's output buffer, when the temporary arrival rate from all input interfaces exceeds the capacity of the output link. The mechanism of these losses is simple - the buffer gets full and newly arriving packets are deleted.

When analyzing and describing the packet loss process, the most obvious and useful characteristic is the loss ratio, i.e., the ratio of the number of lost packets to the total number of packets, considered in some, usually not short, time interval. This characteristics has been widely studied using measurements [1]–[7] and mathematical models [8]–[15].

The second well known and useful characteristic of the loss process is the burst ratio [16]. This characteristic is just the average length of the series of consecutive packet losses, properly normalized. Therefore, it describes not the bare frequency of occurrences of losses, but their statistical structure – tendency to cluster together in series. Such series are known to impair significantly the quality of real-time audio and video transmissions, via unpleasant pausing of freezing [17]. There is also a substantial literature on the burst ratio, with direct measurements [18]–[21] and mathematical models [22]–[28].

As it was said, the burst ratio is proportional to the average length of the series of losses. Hence, it does not contain any detailed information about the distribution of the series of losses, but its average value only. In particular, it does not contain any information about the variability of the series of losses and how heavy is the tail of this distribution. Intuitively, such information may be of some value when analyzing the impairment of real-time audio and video transmissions. For instance, a heavy tail of this distribution indicates that we can expect occasionally a very long series of losses.

In this paper, we compute and analyze higher-order statistics of the distribution of the series of losses, i.e., the coefficient of variation, skewness and excess kurtosis. They are based on higher moments (the second, third and fourth moment, respectively) and, combined together, contain much more detailed information about this distribution, than the bare burst ratio.

In calculations, we use a mathematical model of the buffer fed by an aggregated traffic. The Batch Markovian Arrival Process (BMAP) is used to model the traffic. It is perhaps one of the most versatile and useful models, due to it broad modeling capabilities. In particular, using BMAP we can model an arbitrary interarrival time distribution and the autocorrelation of packet interarrival times [29], [30], batch arrivals (useful in TCP modeling [31]) and several other features of traffic. Some of these features will be exploited in numerical examples. The BMAP process is widely used in the performance evaluation of networks, see [32]–[35] and the references there.

To the best of the authors' knowledge, there are no published results on higher-order statistics of the series of losses. The only results we are aware of are those devoted to the first-order statistic, i.e., the burst ratio, published in the papers mentioned above.

The rest of the paper is structured as follows. In Section II, the model of the buffer fed by the BMAP process is described. In Section III, formulas for the coefficient of variation, skewness and excess kurtosis of the series of packet losses are presented. In Section IV, numerical results are given. In particular, five different parameterizations of traffic are used and accompanied by different buffer sizes, loads of the queue and service times. Influence of all these factors on higher-order statistics of the series of losses is discussed. Concluding remarks are given in Section V.

II. THE MODEL

We use a finite-buffer queueing model with a single server. Namely, the buffer size is K, including the service position. If upon a packet arrival there are K packets present in the buffer, a newly arriving packet is deleted. The service time has general distribution given by distribution function F with the average value of $\overline{F}$. The load of the queue, ρ, is defined as

$$\rho = \Lambda\overline{F}, \tag{1}$$

where Λ is the rate of the arrival process.

The packet arrival process is modeled by the BMAP process [36]. BMAP is a Markov process denoted by $(N(t), J(t))$, $t \geq 0$, where $N(t)$ is the cumulative number of packets that arrived in $(0, t)$, while $J(t)$ is the state of some Markov chain (continuous-time type), called the modulating chain, with the state space $\{1, \ldots, s\}$. The infinitesimal matrix of $(N(t), J(t))$ is:

$$\begin{bmatrix} D_0 & D_1 & D_2 & D_3 & \cdot\cdot \\ & D_0 & D_1 & D_2 & \cdot\cdot \\ & & D_0 & D_1 & \cdot\cdot \\ & & & \cdot & \cdot\cdot \end{bmatrix},$$

where each D_i, $i \geq 0$, constitutes an $s \times s$ matrix. In addition, each $D_i, i \geq 1$, is nonnegative, while D_0 is negative on its diagonal elements and nonnegative outside the diagonal. Finally, $D = \sum_{i=0}^{\infty} D_i$ has to be an irreducible infinitesimal matrix and has to differ from D_0.

In the analysis of BMAP, the function $P_{i,j}(n, t)$ is used frequently:

$$P_{n,m}(k, t) = \Pr\{N(t) = k, J(t) = m | N(0) = 0, J(0) = n\}. \tag{2}$$

In what follows, $\boldsymbol{e}$ is the column vector of length s of 1's, I is the $s \times s$ identity matrix, $\mathbf{0}$ is the $s \times s$ matrix of 0's and $\mathbf{1}$ is the $s \times s$ matrix of 1's.

III. HIGHER-ORDER STATISTICS

The higher-order statistics of the length of the series of losses can be obtained in a similar way, as the burst ratio parameter was obtained in [27]. Namely, following the proof of Theorem 1 of [27], we can see that the probability that the series of consecutive packet losses is of length k is:

$$P_k = \frac{\boldsymbol{u}\boldsymbol{r}(k)}{1 - \boldsymbol{u}\boldsymbol{r}(0)}, \qquad k = 1, 2, \ldots, \tag{3}$$

where vector $\boldsymbol{u}$ of size s contains the distribution of the modulating chain at the moment ending the buffer overflow period in the stationary regime. It can be computed as the stationary vector of stochastic $s \times s$ matrix V, i.e., the vector fulfilling the set of equations:

$$\begin{cases} \boldsymbol{u}\boldsymbol{e} = 1, \\ \boldsymbol{u}V = \boldsymbol{u}, \end{cases} \tag{4}$$

where matrix V has the following form:

$$V = W^{-1}\left(Z + \sum_{i=1}^{K} R_{K-i}\overline{A}_i - \sum_{i=1}^{K}\sum_{j=1}^{i} Y_{K-i}R_{i-j}\overline{A}_j\right), \tag{5}$$

with

$$A_k = \left[\int_0^{\infty} P_{n,m}(k, t)dF(t)\right]_{n,m}, \tag{6}$$

$$Y_k = \left[\frac{-(D_0)_{nn}p_n(k, m)}{\lambda_n}\right]_{n,m}, \tag{7}$$

$$R_0 = \mathbf{0}, \quad R_1 = A_0^{-1},$$

$$R_{j+1} = A_0^{-1}\left(R_j - \sum_{i=0}^{j} A_{i+1}R_{j-i}\right), \quad j \geq 1, \tag{8}$$

$$Z = \sum_{i=K}^{\infty} Y_i\overline{A}_0, \tag{9}$$

$$\overline{A}_i = A_0 - \sum_{j=0}^{i-1} A_j, \tag{10}$$

and

$$W = \sum_{i=0}^{K} R_{K-i}A_i - \sum_{i=1}^{K}\sum_{j=0}^{i} Y_{K-i}R_{i-j}A_j. \tag{11}$$

Moreover, $p_n(k, m)$ in (7) is the probability that in the arrival process there will be a change of the modulating chain to m together with an arrival of a batch of size k, if the modulating chain is currently in state n. This probability is equal to:

$$p_n(0, n) = 0, \quad \text{for every } n, \tag{12}$$

$$p_n(0, m) = \frac{1}{-(D_0)_{nn}}(D_0)_{nm}, \quad n \neq m, \tag{13}$$

$$p_n(k, m) = \frac{1}{-(D_0)_{nn}}(D_m)_{nm}, \quad k \geq 1. \tag{14}$$

On the other hand, vectors $\boldsymbol{r}(k)$, which are present in (3), are defined as follows. The n-th entry of vector $\boldsymbol{r}(k)$ is the probability, that during the first buffer overflow period the number of lost packets equals k, assuming $X(0) = K - 1$ and $J(0) = n$. Vectors $\boldsymbol{r}(k)$ have the following form, see [27]:

$$\boldsymbol{r}(k) = W^{-1} \cdot \left(\sum_{i=1}^{K} R_{K-i}A_{i+k} - \sum_{i=1}^{K}\sum_{j=1}^{i} Y_{K-i}R_{i-j}A_{j+k} + \sum_{i=K}^{K+k} Y_i A_{K+k-i}\right)\boldsymbol{e}. \tag{15}$$

Matrices A_k, which are present in (8), (10), (11) and (15), can be computed using the uniformization method of [36]. Exploiting this method we get:

$$A_i = \sum_{j=0}^{\infty} \frac{T_{i,j}}{j!} \int_0^{\infty} e^{-\theta t}(\theta t)^j dF(t), \tag{16}$$

with

$$\theta = \max_n\{(-D_0)_{nn}\}, \tag{17}$$

and:

$$\begin{aligned} T_{0,0} &= I, & (18) \\ T_{k,0} &= \mathbf{0}, \quad k \geq 1, & (19) \\ T_{0,j+1} &= T_{0,j}(I + \theta^{-1}D_0), & (20) \\ T_{k,j+1} &= \theta^{-1}\sum_{i=0}^{k-1} T_{i,j}D_{k-i} + T_{k,j}(I + \theta^{-1}D_0). & (21) \end{aligned}$$

Now, using (3) with (4), (5) and (15), we can obtain higher-order statistics of the series of losses. Namely, the coefficient of variation, C_v, of the length of the series of losses is:

$$C_v = \frac{S}{G}, \tag{22}$$

where G is the average length of the series of losses:

$$G = \frac{\sum_{k=1}^{\infty} k\boldsymbol{ur}(k)}{1-\boldsymbol{ur}(0)}, \tag{23}$$

while S is the standard deviation:

$$S = \sqrt{\frac{\sum_{k=1}^{\infty} k^2\boldsymbol{ur}(k)}{1-\boldsymbol{ur}(0)} - G^2}. \tag{24}$$

The skewness M of the length of the series of losses is:

$$M = \frac{\sum_{k=1}^{\infty}(k-G)^3\boldsymbol{ur}(k)}{S^3(1-\boldsymbol{ur}(0))}. \tag{25}$$

Finally, the excess kurtosis N of the length of the series of losses is:

$$N = \frac{\sum_{k=1}^{\infty}(k-G)^4\boldsymbol{ur}(k)}{S^4(1-\boldsymbol{ur}(0))} - 3. \tag{26}$$

IV. EXAMPLES

In these examples, we will use the same parameterizations of the system that were used in [27] to study the first-order statistic. In particular, the following five BMAP parameterizations will be used:

$BMAP_1$: $D_0 = -6.66666666 \cdot I,\ D_1 = 6.66666666 \cdot I.$
$BMAP_2$:

$$D_0 = \begin{bmatrix} -2.66407491 & 0.21318153 & 0.06876823 \\ 0.28194977 & -4.12978130 & 0.28194977 \\ 0.07564506 & 0.07564506 & -14.6406033 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} 1.30324098 & 0.60086202 & 0.47802213 \\ 0.28641400 & 2.92428551 & 0.35518224 \\ 0.80716673 & 0.28882659 & 13.3933198 \end{bmatrix}.$$

$BMAP_3$: $D_0 = -I,\ D_2 = 0.02222222 \cdot \mathbf{1},\ D_4 = 0.07777778 \cdot \mathbf{1},\ D_8 = 0.23333333 \cdot \mathbf{1}.$
$BMAP_4$:

$$D_0 = \begin{bmatrix} -0.39961124 & 0.03197723 & 0.01031523 \\ 0.04229246 & -0.61946720 & 0.04229246 \\ 0.01134675 & 0.01134675 & -2.19609050 \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 0.14544482 & 0.01134675 & 0.02166199 \\ 0.01134675 & 0.03197723 & 0.02166199 \\ 0.02166199 & 0.03197723 & 0.05260770 \end{bmatrix},$$

$$D_4 = \begin{bmatrix} 0.01134675 & 0.02166199 & 0.01134675 \\ 0.01134675 & 0.33111906 & 0.01134675 \\ 0.04229246 & 0.01134675 & 0.02166199 \end{bmatrix},$$

$$D_8 = \begin{bmatrix} 0.03869456 & 0.05712054 & 0.03869456 \\ 0.02026858 & 0.07554653 & 0.02026858 \\ 0.05712054 & 0.00000000 & 1.93472829 \end{bmatrix}.$$

$BMAP_5$:

$$D_0 = \begin{bmatrix} -45.5935855 & 1.95261616 & 0.19526161 \\ 0.01952616 & -4.55935855 & 0.19526161 \\ 0.00195261 & 0.01952616 & -0.45593586 \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 0.06508720 & 0.52069762 & 5.20697622 \\ 0.52069762 & 0.00065087 & 0.05792761 \\ 0.05076801 & 0.00650872 & 0.00065087 \end{bmatrix},$$

$$D_4 = \begin{bmatrix} 0.06508720 & 0.52069762 & 5.20697622 \\ 0.52069762 & 0.00065087 & 0.05792761 \\ 0.05076801 & 0.00650872 & 0.00065087 \end{bmatrix},$$

$$D_8 = \begin{bmatrix} 0.35797962 & 2.86383692 & 28.6383692 \\ 2.86383692 & 0.00357979 & 0.31860186 \\ 0.27922410 & 0.03579796 & 0.00357979 \end{bmatrix}.$$

On purpose, all of these arrival processes have the same arrival rate, $\Lambda = 20/3$, but quite different internal statistical properties. In particular, $BMAP_1$ is in fact a simple Poisson process, so it has no autocorrelation of interarrival times, nor batch arrivals. $BMAP_2$ is positively, strongly autocorrelated, but has no batch arrivals. On the other hand, $BMAP_3$ is not autocorrelated, but has batch arrivals. Finally, both $BMAP_4$ and $BMAP_5$ are strongly autocorrelated and have batch arrivals. The difference is that the autocorrelation of $BMAP_4$ is positive, while $BMAP_5$ has an oscillating autocorrelation, with positive and negative signs. It is important that when batch arrivals are involved (in $BMAP_3$, $BMAP_4$ and $BMAP_5$), the same batch sizes are used in every case. Similarly, in two cases with positive autocorrelation ($BMAP_2$ and $BMAP_4$), exactly the same autocorrelation function is used.

By default, we will use $K = 50$, $\rho = 1$ and exponential service time with the mean of 3/20. There will be some exceptions, but they will be clearly stated.

In Figure 1, the distribution P_k of the length of the series of losses is presented for all considered arrival processes. As we can see, for $BMAP_1$ and $BMAP_2$, which do not have the batch structure, the distribution is regular and monotonic. A heavier tail can be observed in the correlated case, $BMAP_2$. For $BMAP_3, \ldots, BMAP_5$, which do have the batch structure, this distribution has an irregular form, with multiple spikes.

TABLE I
HIGHER-ORDER STATISTICS OF THE SERIES OF LOSSES FOR DIFFERENT ARRIVAL TRAFFIC. $K = 50$ AND $\rho = 1$.

traffic	C_v	M	N
$BMAP_1$	0.7071	2.1213	6.5000
$BMAP_2$	0.8030	2.0786	6.3582
$BMAP_3$	0.7414	1.8390	5.6675
$BMAP_4$	0.7552	2.0258	6.3752
$BMAP_5$	0.7378	1.6039	4.5386

In Table I, the higher-order statistics of the series of consecutive packet losses for $BMAP_1, \ldots, BMAP_5$ are presented. As we can see, the coefficient of variation assumes moderate

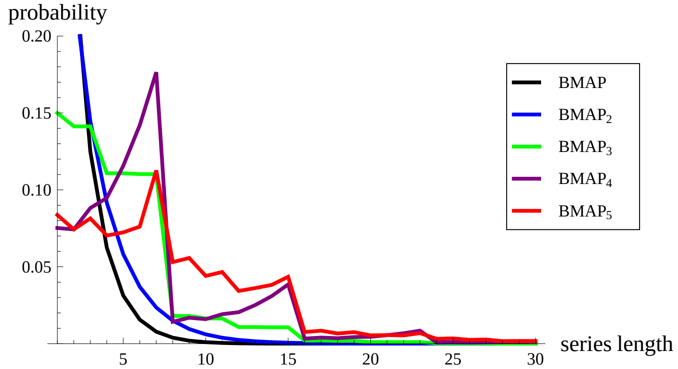


Figure 1. Distribution of the length of the series of losses for $BMAP_1, \ldots, BMAP_5$. $K = 50$ and $\rho = 1$.

values in the range 0.7-0.8. The autocorrelated structure seems to elevate C_v slightly (compare $BMAP_2$ with $BMAP_1$). This is not so clear about the batch structure. If we compare $BMAP_3$ with $BMAP_1$, it seems that the batch structure makes the coefficient of variation greater. On the other hand, C_v is less in the case of $BMAP_4$, than in the case of $BMAP_2$, even though they share the same autocorrelation function, $BMAP_4$ has the batch structure, while $BMAP_2$ does not.

The skewness in Table I is positive in all the cases, with the values around 2. This is consistent with Figure 1, in which all the tails are on the right side.

The most interesting statistic in Table I is the excess kurtosis. As we can see, it assumes rather high, positive values in all the cases, which indicate fat tails of the distributions of the series of losses (much fatter than in the case of normal distribution which has $N = 0$).

Contrary to C_v, the excess kurtosis seems to be less when a positive autocorrelation or batch structure is involved - compare again $BMAP_2$ with $BMAP_1$ and $BMAP_3$ with $BMAP_1$, respectively. In the case of the oscillating autocorrelation, $BMAP_5$, the smallest value is observed, while still rather high.

So far, only the buffer of size 50 was considered. Now we will check the dependence of the higher-order statistics on the buffer size.

In Figures 2, 3 and 4, the coefficient of variation, skewness and excess kurtosis as functions of the buffer size are presented, for all the considered arrival streams. As we can see, all three statistics are practically independent on the buffer size, when the arrival process has no batch structure - the curves are flat for $BMAP_1$ and $BMAP_2$. On the other hand, a high dependence of the three statistics on the buffer size can be observed when the arrival process does have the batch structure and the buffer is rather small - see the spikes in Figures 2-4 for $BMAP_3, \ldots, BMAP_5$. However, for a relatively small K, about 25, all three statistics stabilize and do not change anymore, when the buffer grows. Therefore $K = 50$ used herein as a default value is already in the stable regime.

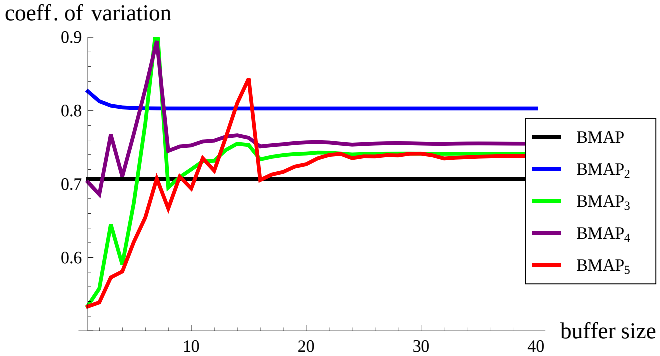


Figure 2. Coefficient of variation versus the buffer size for $BMAP_1, \ldots, BMAP_5$, $\rho = 1$.

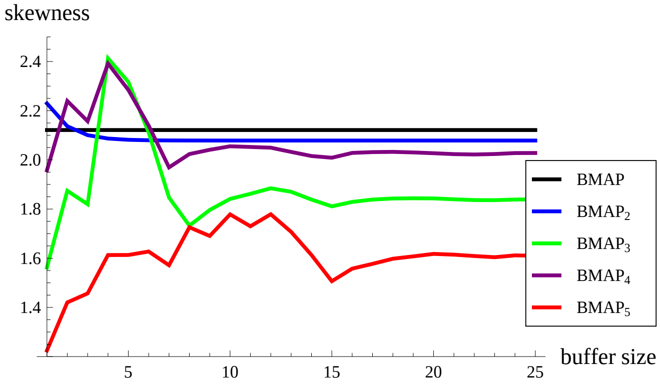


Figure 3. Skewness versus the buffer size for $BMAP_1, \ldots, BMAP_5$, $\rho = 1$.

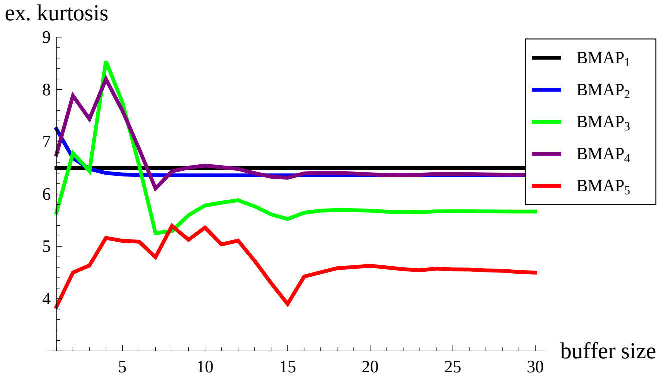


Figure 4. Excess kurtosis versus the buffer size for $BMAP_1, \ldots, BMAP_5$, $\rho = 1$.

Increasing the buffer e.g. to $K = 100$ would have a negligible effect on the three statistics.

Now we get back to $K = 50$ and vary the load of the queue

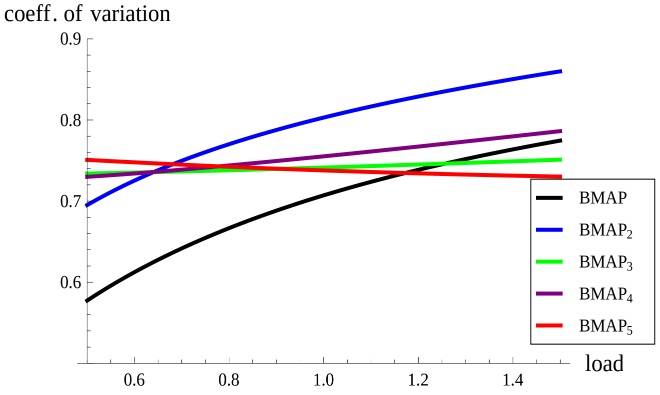


Figure 5. Coefficient of variation versus the system load for $BMAP_1, \ldots, BMAP_5$, $K = 50$.

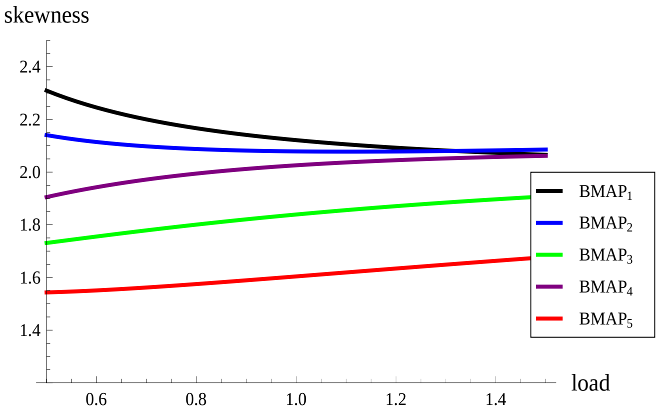


Figure 6. Skewness versus the system load size for $BMAP_1, \ldots, BMAP_5$, $K = 50$.

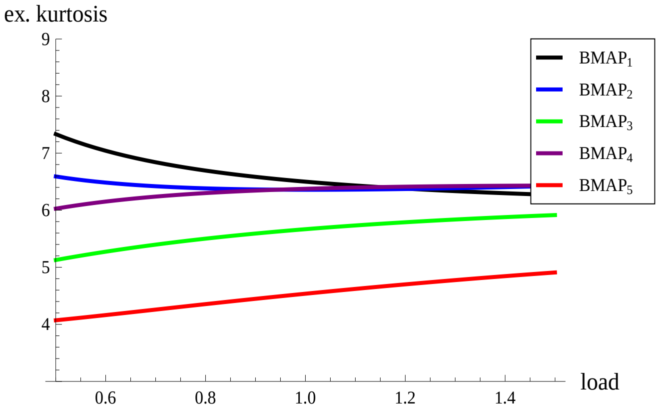


Figure 7. Excess kurtosis versus the system load for $BMAP_1, \ldots, BMAP_5$, $K = 50$.

(it was unaltered so far, $\rho = 1$).

In Figures 5, 6 and 7, the coefficient of variation, skewness and excess kurtosis as functions of the load of the queue are presented for all the considered arrival streams. As we can see in Figure 5, the behaviour of C_v depends strongly on the presence of batches. Namely, C_v grows significantly with the queue load if the arrivals are single ($BMAP_1$, $BMAP_2$), no matter if the arrival process is correlated or not. C_v changes much slower with load when the arrival process has the batch structure.

In Figure 6, we can notice that the skewness is different for all the arrival processes, when the load is low, $\rho = 0.5$. When the load is high, $\rho = 1.5$, the skewness is almost identical for $BMAP_1$, $BMAP_2$ and $BMAP_4$, but different for remaining BMAPs. A similar situation is in the case of excess kurtosis, which can be observed in Figure 7.

TABLE II
HIGHER-ORDER STATISTICS OF THE SERIES OF LOSSES FOR DIFFERENT SERVICE TIME DISTRIBUTIONS. $BMAP_5$, $K = 50$ AND $\rho = 1$ WERE USED.

service time	C_v	M	N
F_1	0.6729	1.1257	2.1510
F_2	0.6954	1.2963	2.8960
F_3	0.7090	1.4159	3.5801
F_4	0.7378	1.6039	4.5386
F_5	0.9512	2.9188	14.2978

Now we get back to $\rho = 1$ and vary the service time distribution, which was exponential in all the examples so far. Namely, we use now the following five distribution functions of the service time:

$$F_1(x) = 0 \text{ if } x < \frac{3}{20}, \quad F_1(x) = 1 \text{ otherwise}, \tag{27}$$

$$F_2(x) = \frac{20}{6}x, \quad 0 \leq x < \frac{6}{20}, \tag{28}$$

$$F_3(x) = 1 - \frac{40}{3}xe^{-\frac{40}{3}x} - e^{-\frac{40}{3}x}, \quad x \geq 0, \tag{29}$$

$$F_4(x) = 1 - \frac{20}{3}e^{-\frac{20}{3}x}, \quad x \geq 0, \tag{30}$$

$$F_5(x) = 1 - 0.95e^{-9.5x} - 0.05e^{-x}, \quad x \geq 0. \tag{31}$$

All of these distributions have the mean of 3/20, therefore all of them produce the load of 1. They differ, however, in the standard deviation, which is 0, 0.086, 0.107, 0.150 and 0.314 for F_1-F_5, respectively.

The results are shown in Table II. They can be summarized in two points. First, the variation of the service time influences significantly all three higher-order statistics. A particularly great influence can be observed in the case of the excess kurtosis. Second, the dependence is monotonic in each case, i.e., the higher the variation of the service time, the higher C_v, M and N.

V. CONCLUSIONS

We analyzed higher-order statistics of the length of the series of packet losses at a router's output buffer, using a queueing system with a complex, flexible traffic model. In particular, we presented formulas and numerical examples for the coefficient of variation, skewness and excess kurtosis of the length of the series.

A few observations were made. For instance, all three statistics depended on the buffer size in a very complicated way, but only for small buffer sizes. For a moderate buffer size, they all stabilized and did not change anymore when the buffer grew. All three statistics were significantly influenced by the variance of the service time, in a monotonic manner. The coefficient of variation of the series of losses grew rather quickly with the system load when the arrival process did not have the batch structure, and much slower, when it did.

Perhaps the most important observation made was that the distribution of the series of losses was strongly leptokurtic in all the considered examples. It means that this distribution has usually a rather fat tail, so occasionally a long series of losses can be expected. Naturally, such series may influence badly the quality of real-time multimedia transmissions. Unfortunately, no quantitative measure of such influence has been proposed so far in terms of a higher-order statistic. For instance, the impairment of voice transmission is estimated using the first-order statistic only (see [17]). An interesting future work would be proposing a formula for this impairment, taking into account the kurtosis.

ACKNOWLEDGEMENT

This research was funded by National Science Centre, Poland, grant number 2020/39/B/ST6/00224.

REFERENCES

[1] J. Bolot, “End-to-end packet delay and loss behavior in the Internet”, Proc of ACM SIGCOMM’93, pp. 289–298, 1993.

[2] M. Yajnik, S. Moon, J. Kurose, and D. Towsley, “Measurement and modelling of the temporal dependence in packet loss”, Proc. IEEE INFOCOM, vol. 1, pp. 345–352, 1999.

[3] N. G. Duffield, F. Lo Presti, V. Paxson, and D. Towsley, “Inferring link loss using striped unicast probes”, IEEE INFOCOM, pp. 915–923, 2001.

[4] J. Sommers, P. Barford, N. Duffield, and A. Ron, “Improving accuracy in end-to-end packet loss measurement”, Computer Communication Review, no. 35(4), pp. 157–168, 2005.

[5] J. Sommers, P. Barford, N. Duffield, and A. Ron, “A geometric approach to improving active packet loss measurement”, IEEE/ACM Transactions on Networking, no. 16(2), pp. 3 – 320, 2008.

[6] Z. Hu and Q. Zhang, “A new approach for packet loss measurement of video streaming and its application”, Multimedia Tools and Applications, no. 77(10), pp. 11589–11608, 2018.

[7] H. Lan, “Strengthening packet loss measurement from the network intermediate point”, KSII Transactions on Internet and Information Systems, no. 13(12), pp. 5948–5971, 2019.

[8] I. Cidon, A. Khamisy, and M. Sidi, “Analysis of packet loss processes in high-speed networks”, IEEE Transactions on Information Theory, no. 39(1), pp. 98–108, 1993.

[9] X. Yu, J. W. Modestino, and X. Tian, “The accuracy of Gilbert models in predicting packet-loss statistics for a single-multiplexer network model”, IEEE INFOCOM, pp.2602–2612, 2005.

[10] A. Chydzinski, R. Wojcicki, and G. Hryn, “On the Number of Losses in an MMPP Queue”, Lecture Notes in Computer Science, no. 4712, pp. 38–48, 2007.

[11] G. Hasslinger and O. Hohlfeld, “The Gilbert-Elliott model for packet loss in real time services on the Internet”, Proc. of MMB 2008, pp. 1–15, 2008.

[12] M. Bratiychuk and A. Chydzinski, “On the loss process in a batch arrival queue”, Applied Mathematical Modelling, no. 33, iss. 9, pp. 3565–3577, 2009.

[13] A. Chydzinski, and B. Adamczyk, “Transient and stationary losses in a finite-buffer queue with batch arrivals”, Mathematical Problems in Engineering, vol. 2012, ID 326830, pp. 1–18, 2012.

[14] M. Ellis, D. P. Pezaros, T. Kypraios, and C. Perkins, “A two-level Markov model for packet loss in UDP/IP-based real-time video applications targeting residential users”, Computer Networks no. 70, pp. 384–399, 2014.

[15] P. Mrozowski and A. Chydzinski, “Queues with dropping functions and autocorrelated arrivals”, Methodology and Computing in Applied Probability, no. 20(1), pp. 97–115, 2018.

[16] J. W. McGowan, “Burst ratio: a measure of bursty loss on packet-based networks”, 16 2005. US Patent 6,931,017, 2005.

[17] ITU-T Recommendation G.107: “The E-model, a computational model for use in transmission planning”, Technical report, 2014.

[18] D. Samociuk, A. Chydzinski, and M. Barczyk, “Experimental measurements of the packet burst ratio parameter”, Communications in Computer and Information Science, no. 928, pp. 455–466, 2018.

[19] D. Samociuk, M. Barczyk, and A. Chydzinski, “Measuring and analyzing the burst ratio in IP traffic”, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, no. 303, pp. 86–101, 2019.

[20] A. Chydzinski and D. Samociuk, “Burst ratio in a single-server queue”, Telecommunication Systems, no. 70(2), pp. 263–276, 2019.

[21] M. Barczyk and A. Chydzinski, “AQM based on the queue length: A real-network study”, PLoS ONE, no. 17(2): e0263407, 2022.

[22] J. Rachwalski and Z. Papir, “Burst Ratio in Concatenated Markov-based Channels”, Journal of Telecommunications and Information Technology, no. 1, pp. 3–9, 2014.

[23] J. Rachwalski and Z. Papir, “Analysis of Burst Ratio in Concatenated Channels”, Journal of Telecommunications and Information Technology, no. 4, pp. 65–73, 2015.

[24] A. Chydzinski, M. Barczyk, and D. Samociuk, “The Single-Server Queue with the Dropping Function and Infinite Buffer”, Mathematical Problems in Engineering, vol. 2018, ID 3260428, pp. 1–12, 2018.

[25] A. Chydzinski, D. Samociuk, and B. Adamczyk, “Burst ratio in the finite-buffer queue with batch Poisson arrivals”, Applied Mathematics and Computation, no. 330, pp. 225–238, 2018.

[26] A. Chydzinski, “Impact of the Dropping Function on Clustering of Packet Losses”, Sensors, no. 22(20), 2022.

[27] A. Chydzinski, “Burst ratio for a versatile traffic model”, PLoS ONE, no. 17(8): e0272263, 2022.

[28] A. Chydzinski, “Per-flow structure of losses in a finite-buffer queue”, Applied Mathematics and Computation, no. 428, 127215, pp. 1-15, May 2022.

[29] A. Klemm, C. Lindemann, and M. Lohmann, “Modeling IP traffic using the batch Markovian arrival process”, Performance Evaluation, vol. 54, issue 2, 2003.

[30] P. Salvador, A. Pacheco, and R. Valadas, “Modeling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs”, Computer Networks, no. 44, pp. 335-352, 2004.

[31] M. Garetto and D. Towsley, “An efficient technique to analyze the impact of bursty TCP traffic in wide-area networks”, Performance Evaluation, no. 65, pp.181–202, 2008.

[32] H. W. Lee, N. I. Park and J. Jeon, “A new approach to the queue length and waiting time of BMAP/G/1 queues”, Computers & Operations Research, vol. 30, pp. 2021–2045, 2003.

[33] T. Hofkens, K. Spaey and C. Blondia, “Transient Analysis of the D-BMAP/G/1 Queue with an Application to the Dimensioning of a Playout Buffer for VBR Video”, Proceedings of Networking’04. Lecture Notes in Computer Science, vol. 3042, pp. 1338–1343, 2004.

[34] A. N. Dudin, A. A. Shaban, and V. I. Klimenok, “Analysis of a queue in the BMAP/G/1/N system”, International Journal of Simulation: Systems, Science and Technology. vol. 6, no. 1–2, pp. 13–23, 2005.

[35] A. D. Banik and M. L. Chaudhry, “Efficient Computational Analysis of Stationary Probabilities for the Queueing System BMAP/G/1/N With or Without Vacation(s)”, INFORMS Journal on Computing, no. 29(1), pp. 140–151, 2017.

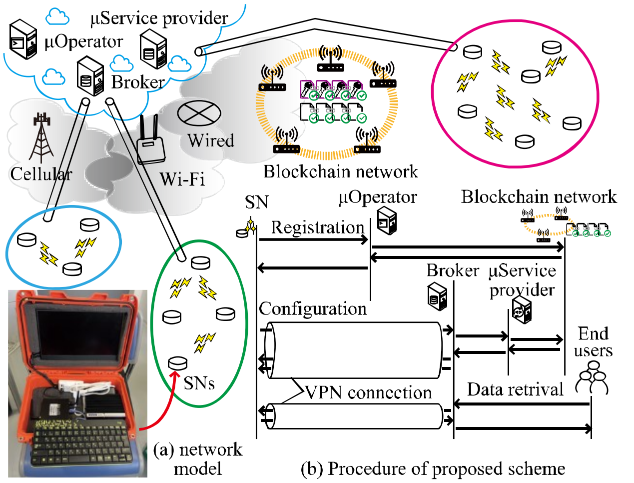
[36] D. M. Lucantoni, “New results on the single server queue with a batch Markovian arrival process”, Commun. Stat., Stochastic Models, vol. 7, no. 1, pp. 1–46, 1991.

A Study on Zero-touch-design Information-centric Wireless Sensor Networks

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 describes a novel zero-touch-design information-centric wireless sensor network for smart-city applications. To promote self-growing in the autonomous-distributed environment, the proposed scheme adopts a mechanism of zero-touch technology focusing on the lower layers in which sensor nodes join the network. The results of preliminary computer simulations demonstrate the effectiveness of the proposed scheme in terms of energy consumption. This study is a part of our ongoing research project to develop an ecosystem that enables a smart-city-as-a-service platform, where we are currently focused on the development and experimental trials in on-site testing.**

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



(a) network model (b) Procedure of proposed scheme

Figure 1. Overview of proposed scheme

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, which remain major challenges. Further, in next-generation wireless networks, such as beyond Fifth-Generation (5G), massive Sensor Nodes (SNs) might 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-based one that provides access to various end-users and applications ranging from individuals to enterprises or governments [1].

In light of this background, we focus on two key techniques: zero-touch and data-centric. Zero-touch technology aims at completely automating the network-management process to minimize operating costs and set up individual execution environments. A zero-touch-design system was utilized in the first Linux operating system and has since led to the demand for service deployments that are versatile and flexible in cloud-native micro-services. As for the data-centric techniques, an Information-Centric Network (ICN) (e.g., a content-centric network or named-data network) can renovate current network protocols, such as 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 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 Networks (ICWSNs).

In our previous study [2], an ICWSN-based ecosystem with a blockchain for smart-city applications was investigated on the basis of a scheme that achieves efficient and reliable caching. In our ongoing research project, which we call the Decentralized Digital twins' Ecosystem (D^{2}EcoSys), we have developed a portable testbed device, and evaluated its applicability to mmWave-band Wireless Fidelity (Wi-Fi), e.g., IEEE 8012.11 ad/ay, towards deployment for on-site testing. The current paper provides a blueprint of zero-touch-design ICWSNs to promote self-growing and ensure a reliable sensing-data distribution in which multiple players actively participate and exchange data. The computer simulation was conducted to investigate energy consumption, as the potential waste involved in the use of ICN and blockchain is significant.

The remainder of this paper is organized as follows. Section II describes the proposed scheme and Section III presents the numerical results. In Section IV, we provide related works and. We conclude in Section V with a brief summary and findings.

II. PROPOSED SCHEME

In zero-touch-design ICWSNs, the first step is to enable automatic participation in the network, i.e., the network trusts an individual SN through the device owner when the SN joins a member of ICWSNs. This procedure requires comprehensive involvement from device owners, micro-(network) operators, and micro-service (roaming) providers, as shown in Figure 1. For the management of terminal information, the proposed scheme utilizes blockchain-based ledgers. Specifically, it obtains the SN identification information from the blockchain network, which plays an instrumental role in the execution of smart contracts. The scheme guarantees the trustworthiness of the SN during an initial process, i.e., it considers that the data generated by a reliable SN can be trusted without receiving any verification from the blockchain network [2]. For this reason, the blockchain-based storage for the data no longer requires traditional computation-intensive mining, and the blockchain can simply select alternative consensus schemes, such as proof-of-authority or proof-of-elapsed-time algorithms.

When an SN device is turned on, it sends a registration request to the uOperator 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 (L2) layer, and secure data exchanges. After joining the network, the SN downloads and installs a configuration setting and application software from the uService provider.

III. NUMERICAL RESULTS

The difficulty in deploying the proposed scheme is how to ensure any 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 2 shows the computer simulation we ran to compare the cumulative energy consumption with the conventional scheme (current application-programming-interface-based IoT platform). The computer simulator was implemented using the C++ language. The simulation condition is that the generation and transmission of sensing data respectively correspond to the status of calculation and wireless communication. In addition, 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 [3].

As shown in Figure 2(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 2(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 node was determined by a Poisson distribution, which is a more realistic calculation than the one in Figure 2(a). As we can see from the figures, the proposed scheme can reduce energy consumption by 3.85% and be advantageous until 138 retrieval attempts.

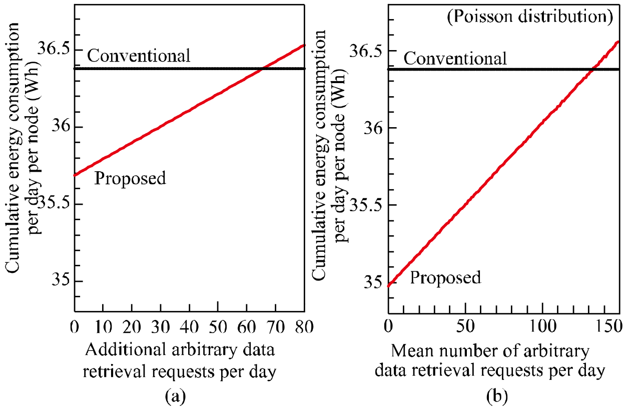


Figure 2. Simulation results: (a) additional data retrieval requests per day vs. cumulative energy consumption, and (b) mean number of additional requests according to a Poisson distribution vs. cumulative energy consumption per unit for 1,000 SNs.

IV. RELATED WORK

Togou et al. [4] developed a distributed blockchain-enabled network slicing framework. Nour et al. [5] introduced a network slice and resource provider utilizing a blockchain in which the scheme could shift from a network-operator-oriented architecture to a more open system with multiple actors. Rathi et al. [6] provided a blockchain-based management and orchestration technique for multi-domain networks.

V. CONCLUSION

This paper described a zero-touch-design ICWSN for a smart-city-as-a-service platform along with a scheme that promotes self-growing in autonomous-distributed conditions. The results of a preliminary numerical investigation indicated that the proposed scheme could effectively reduce energy consumption. Testbed development, deployment, and on-site demonstration of the scheme in realistic smart cities are currently working in progress.

ACKNOWLEDGMENT

This work was partly supported by NICT Japan, Grant Number 05601 and JSPS KAKENHI Grant Number JP19K 20261.

REFERENCES

[1] 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.

[2] S. Mori, “Information-centric wireless sensor networks for smart-city-as-a service: Concept proposal, testbed development, and fundamental evaluation,” *Proc. IEEE CCNC*, Jan. 2023, pp. 945–946.

[3] S. Mori, “A preliminary analysis of data collection and retrieval scheme for green information-centric wireless sensor networks,” *Proc. ACM SigComm WS Net4us*, Aug. 2022, pp. 1–6, doi: https://dl.acm.org/doi/10.1145/3538393.3544932.

[4] 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.

[5] 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.

[6] 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.

Personalized Protocols for Data Hiding in Cloud-to-Things Applications

Marek R. Ogiela
AGH University of Krakow
30 Mickiewicza Ave, 30-059 Kraków, Poland
e-mail: mogiela@agh.edu.pl

Urszula Ogiela
AGH University of Krakow
30 Mickiewicza Ave, 30-059 Kraków, Poland
e-mail: ogiela@agh.edu.pl

***Abstract*— In this article, we will introduce the idea of creating personalized security protocols for hiding and distributing secret information. In particular, a new paradigm combining Cloud Computing with IoT will be considered as an application area for the proposed hiding technologies. Such protocols will take advantage of selected user characteristics in protocols dedicated to secret data transmission.**

Keywords-Cloud-to-Things protocols; data hiding; personalized security protocols.

I. INTRODUCTION

Information hiding algorithms play a huge role in modern computer systems and IT security. Such methods are categorized as steganography, which deals with a variety of techniques for hiding data or creating invisible communication channels. Among the most popular methods are techniques for transmitting secret data placed in image container [1]. The way the information is hidden in the container is the key, which must also be used to reconstruct the data from the media [2].

This paper will describe the idea of using personal data to create a key that allows to distribute secret information in the selected visual container. Since cryptographic keys usually have a fixed length, the proposed method will also use hash functions to encode the personal characteristics of a particular user in the form of hash sequence with particular length.

The rest of the paper is structured as follows. In Section II, we introduce the hash-based personalized hiding protocols. In Section III applications of hiding protocols will be described. Finally, we conclude the work in Section IV.

II. HASH-BASED PERSONALIZED DATA HIDING PROTOCOLS

Information hiding techniques, which exploit users' personal characteristics can use various unique personal biometric or behavioral traits. Popular biometric can be acquired using sensors, or motion capture devices that allow analysis of selected gestures or movements [3]. The resource of individual features acquired in this way makes it possible to select some of them and apply them to an appropriate algorithm for hiding the secret. Selected personal features regardless of their characteristics can be encoded using hash functions. Such an operation allows one to generate a hash of a certain length depending on the selected hash function. Since the resulting hash will be used as a key to place the secret data in the information carrier, it makes sense for it to be as long as possible, which can be achieved by choosing hash functions that generate hashes about 512 bits long.

After generating a personal key for hiding the data in the information carrier, one can proceed to encode the secret so that it is invisible in the container. The key sequence can be used here in various ways. The first is that the key bits can be interpreted as offsets indicating the next pixels of the data where we place the secret information. These offsets can be determined taking into account individual bit values, i.e., 0, 1, or bit blocks containing a larger number of bits and determining larger offset values when determining the next points (pixels) of the carrier to place the secret information in Figure 1.

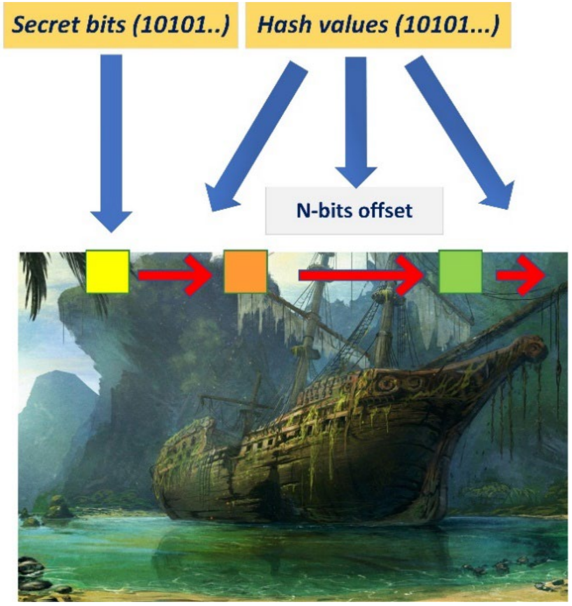


Figure 1. Secret hiding procedure based on offset values.

The second way to use the key can be to use successive bit values to determine the color components of successive image points where part of the secret is to be placed [4]. Thus, in this method, all the points of the container are considered consecutively, and only the bits of the key decide in which color component we place the secret bit of information. The idea of this method is illustrated in Figure 2.

ISBN: 978-1-68558-036-0

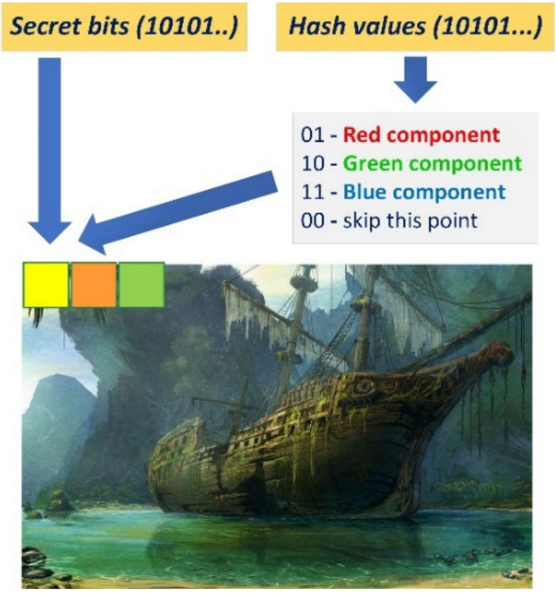


Figure 2. Secret hiding procedure based on color components.

This method can also be modified and place secret information in the color components of consecutive points determined in an irregular way, i.e., not a raster bitmap. Here, it is possible to use a column order of viewing the image and placing the secret instead of a row order, or to use special bit masks of specific sizes (e.g. 3x3, 5x5), which will indicate the neighboring points for the currently considered point, and just in them will allow placing the next bits of secret data.

III. HIDING PROTOCOLS IN CLOUD-TO-THINGS APPLICATIONS

Methods for hiding secret information in visual containers have a number of practical applications. Among them are guaranteeing copyrights in digital works, transferring strategic information or splitting secrets to create multiple keys.

Such applications are of great importance in Cloud-to-Things computing technologies, where there is a need to use particularly important information processed in the computer cloud in the task of guaranteeing confidentiality in the IoT area [5]. IoT-related protocols and services require the use of personalized cryptographic keys, which can be generated and transmitted using the described secrecy hiding protocols [6]. Personal keys, obtained using hash functions, can also be used for user authentication tasks when accessing remote services and data in distributed systems [7].

The presented protocol for generating personal keys in the form of a hashed string is very versatile and can be used wherever the collected information resources are processed using cloud computing resources, and then transmitted and used in Internet of Things devices.

The main advantage of the described approach is the ability to create personalized cryptographic solutions that are oriented to a specific system user, and at the same time allow only authorized users for whom personalized keys have been generated to run procedures for data transmission, device control, resource access or data analysis.

IV. CONCLUSIONS

This paper describes methods of using individual user characteristics and personal features in the creation of personalized cryptographic protocols. In particular, a method for creating personal keys using hash functions has been presented. Such keys are in the form of a hashed bit sequences of a certain length, which was created from encoding selected personal characteristics using a hash function. The resulting keys can then be used in authentication protocols or hiding data procedures. They can also be used in cloud-based services and applications that are dedicated to performing tasks and communicating with devices in IoT.

The techniques presented develop methods categorized as cognitive and personalized cryptography. Future work will attempt to extend them toward creating multiple personalized keys that can be used interchangeably by a single user. Such a solution would provide opportunities for a selected user to simultaneously implement multiple protocols using his personal characteristics. Such protocols would be quite independent and would be implemented using different personal characteristics and the keys produced for them by means of hash functions.

ACKNOWLEDGMENT

This work has been partially supported by the funds of the Polish Ministry of Education and Science assigned to AGH University of Krakow. The research project was supported by the program „Excellence initiative – research university” for the AGH University of Krakow.

REFERENCES

[1] L. Ogiela, “Transformative computing in advanced data analysis processes in the cloud,” Inf. Process. Manage. Vol. 57(5), paper 102260, 2020.

[2] S. Zapechnikov, “Privacy-Preserving Machine Learning as a Tool for Secure Personalized Information Services,” Procedia Computer Science, Vol. 169, 2020, pp. 393-399, doi: 10.1016/j.procs.2020.02.235.

[3] M. R. Ogiela, L. Ogiela, and U. Ogiela, “Biometric methods for advanced strategic data sharing protocols,”. In: 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing IMIS 2015, pp. 179–183, 2015, doi: 10.1109/IMIS.2015.29.

[4] M. R. Ogiela and U. Ogiela, “Secure information splitting using grammar schemes,” Studies in Computational Intelligence, Vol. 244, pp. 327–336. Springer, 2009, doi: 10.1007/978-3-642-03958-4_28.

[5] C. Guan, J. Mou, and Z. Jiang, “Artificial intelligence innovation in education: a twenty-year data-driven historical analysis,” Int. J. Innov. Stud. 4(4), pp. 134–147, 2020.

[6] N. Ferguson and B. Schneier, “Practical Cryptography,” Wiley, 2003.

[7] S. J. H. Yang, H. Ogata, T. Matsui, and N.-S. Chen, “Human-centered artificial intelligence in education: seeing the invisible through the visible,” Comput. Educ.: Artif. Intell. Vol. 2, paper 100008, 2021.

Use Cases and 6G Architecture: New Needs and Challenges

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

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

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

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

***Abstract*— The cellular network standard is stepping to the sixth generation, 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 uses cases that poses 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 that will need the main 6G functionalities to work correctly and meet the expected expectations with the main 6G requirements so that it is possible to identify relevant points in the construction of a robust and flexible network architecture.**

Keywords - 6G; use cases; architecture; applications.

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 aspects of the International 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, and 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 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 continued need for increased coverage, capacity and extremely high user data rates, focus on 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 [1]. 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 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 about the 6G network. For example, new 6G applications and requirements are discussed in [2], 6G enabling technologies are mentioned in [3] and 6G enablers to drive Industry 5.0 are discussed in [4]. 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 remainder of this paper is organized as follows. Section II summarizes the related work. Section III brings some use cases that can be explored in 6G. In Section IV, some target indicators related to these use cases that will bring possible changes in the 6G network architecture. Section V concludes the paper and suggests future works.

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 [5], 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. [6], 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. [7] 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 [8], 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 to fulfill 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 stringent requirements 5G can not meet, its stringent requirements, such as Augmented Reality (AR), Virtual Reality (VR), haptic internet, 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. Three 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, which 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 [9]. 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 [10]. 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 [11].

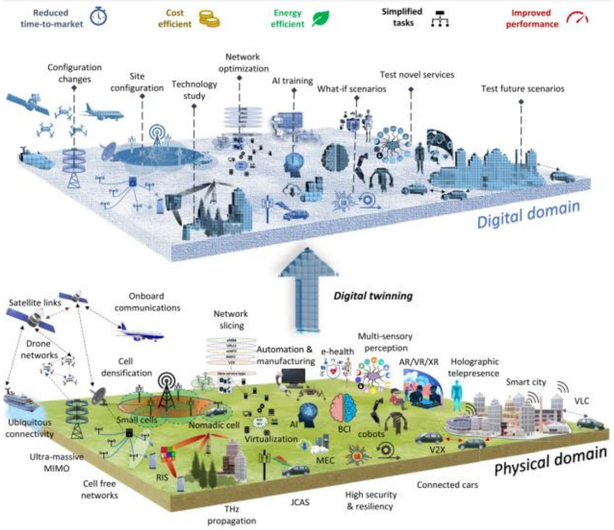


Figure 1. Representation of the virtual elements in relation to a physical system [11].

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 [8], 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 [12], 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 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 [13].

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 our smartphones, every year we see the screen resolution being 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 feedbacks in real time [11].

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, 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, Bio-inspired Technologies, among others.

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

IV. TARGET INDICATORS FOR 6G

Each new use case presents extremely specific advanced requirements and that, in the current scenario, the 5G network does not have the capacity to meet and work with all these requirements. Figure 2 illustrates the comparison between 5G and 6G requirements. Therefore, 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 network.

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, 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.

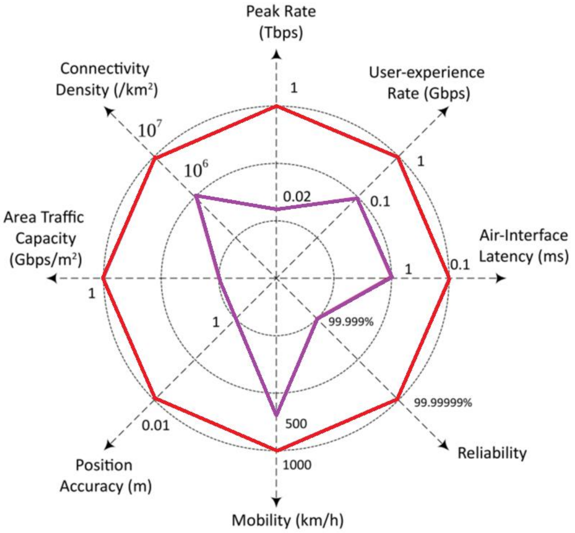


Figure 2. Comparison between 5G and 6G requirements [14].

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 [14]. E2E latency is trickier to manage due to the myriad network elements involved, but 6G should overcome this challenge.

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 [11].

C. Terahertz Communications

Communications in Terahertz work between 100GHz and 10THz and, compared to millimeter waves, they brings a 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 [15].

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 [6].

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

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. comparable [16].

A good comparison of the key THz propagation characteristics and their impact on THz systems, is depicted in Table 1 [17].

V. 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. Finally, the Industry 5.0 use case presents the requirements already mentioned by the previous use cases as a necessity. After studying these use cases, the requirements, also known as target indicators, were discussed bringing confirmation of the need for a new network architecture.

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

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

[3] 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.

[4] 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.

[5] 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.

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

[7] M Giordani et al., “Towards 6G Networks: Use Cases and Technologies”, IEEE Communications Magazine, vol. 58, no. 3, 2020.

[8] 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.

[9] L. U. Khan, W. Saad, D. Niyato, Z. Han and C. S. Hong, “Digital Twin Enabled 6G: Vision, Architecture Trends, and Future Directions,” IEEE Communications Magazine, vol. 60, no. 1, 2022.

[10] A. Rasheed, O. San and T. Kvamsdal, “Digital Twin Values, Challenges And Enablers From A Modeling Perspective,” IEEE Access, vol. 8, 2020.

[11] W. Tong and P. Zhu, “6G The Next Horizon,” Cambridge University Press, 2021.

[12] N. P. Kuruvatti, M. A. Habibi, S. Partani, B. Han, A. Fellan and H. D. Schotten., “Empowering 6G Communication Systems With Digital Twin Technology: A Comprehensive Survey,” IEEE Access, vol. 10, pp. 112158–112186, 2022.

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

[14] 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, 2021.

[15] 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.

[16] L. Yan, C. Han and J. Yuan., “Hybrid Precoding for 6G Terahertz Communications_Performance Evaluation and Open Problems,” 2020 2nd 6G Wireless Summit, 2020.

[17] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjoland and F. Tufvesson., “6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities”, Proceedings of the IEEE, 109.7, 2021, pp. 1166-1199.

Strategies for Minimization of Energy Consumption in Data Centers

Simon Friis*, Line M. P. Larsen*† and Sarah Ruepp*
* Department of Electrical and Photonics Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark
† TDC Net, Copenhagen, Denmark
Corresponding author: email: s175405@student.dtu.dk

***Abstract*—Data Centers (DCs) consume a significant amount of energy, making overall energy consumption a major concern. The scale of a DC affects its energy efficiency, with larger centers having more resources for energy-saving measures but at the same time different challenges than those faced by DCs of smaller scale. This work analyses and compares energy efficiency of small, medium, and large DCs, analysing factors such as resource and server utilization and design. Finally, the energy minimisation techniques are evaluated for their potential impact on DC energy consumption, as well as their contribution to DCs of different sizes.**

Keywords-Data center; cloud; green; low power; scalability; energy efficiency.

I. INTRODUCTION

Energy efficiency in Data Centers (DCs) is a crucial topic of modern DC operations, as it can help to reduce energy costs and environmental impact of the Information and Communications Technology (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 1.5% [1] of global carbon emissions. With the increase in data volume, DCs will consume more energy, thus; there is a need to find new and more efficient ways to run them.

Two types of DCs include private entreprise DCs and public cloud DCs. As illustrated in Fig. 1, the end-users gain access to the DCs to store and process data through a network of computers, wireless Access Points (APs), switches/routers, firewalls, and the Internet. The computers are connected to both data centers 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. 1 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 count:

- The cloud DC enables remote, on-demand access to data and application services from other providers through the Internet.

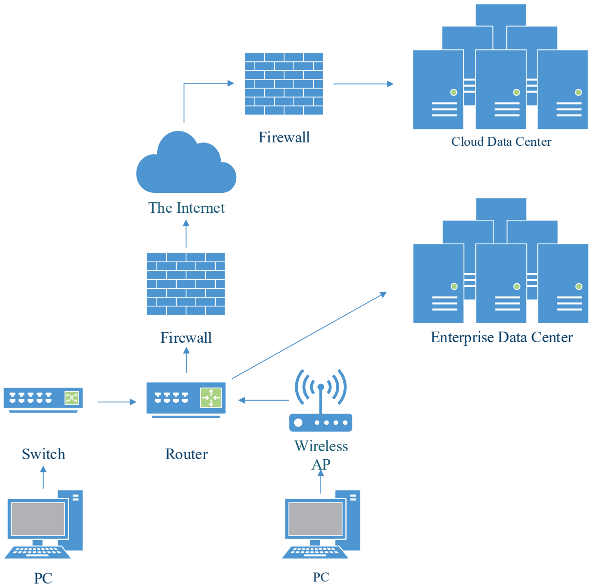


Figure 1. A basic network with computers connecting to both a cloud data center and an enterprise data center.

- 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 [2]:

- *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.
- *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. Reduces the probability of network failures by lowering workload of overwhelmed servers.
- *Access Routers:* A secure connection point for external network traffic.

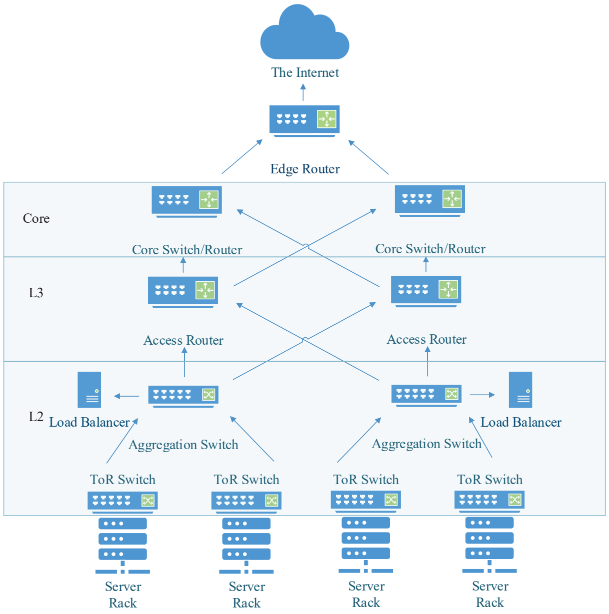


Figure 2. Key components of a basic data center architecture separated into different layers.

- *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.

These devices 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. 2 illustrates the basic elements of a data center 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 serves as the bridge between the internal DC network and the external network, being the Internet.

There are many different types of DCs and they will therefore be categorized into three different sizes in this paper ranging from small, medium and large.

- *Small DCs* are categorized as usually having less than 1,000 servers, as well as less complex infrastructure, limited storage and consume less power compared to larger DCs.
- *Mid-scale DCs* have a larger number of servers and usually range between 1,000 to 10,000 [3] with more complex infrastructure.
- *Large DCs* are defined as usually having more than 10,000 servers with an even more complex infrastructure [3]. Large DCs are typically used by large companies or governments.

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. Section II presents other related papers and features our contribution to the topic. Section III goes in-depth with modern and commonly used strategies. 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.

II. RELATED WORK

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 [4], which presents an overview of energy-aware resource management approaches with focus on basic architecture of cloud DCs and virtualization technology. The survey in [5] investigates the green energy aware power management problem for Megawatt-scale DCs and classifies work that considers renewable energy and/or carbon emission. In [6], they 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. In [7], they conduct an in-depth study of the existing literature on DC power modeling, covering more than 200 models.

Recent related work includes [8], where the approaches moving towards green computing are investigated and categorized to help researchers and specialists in cloud computing expand green cloud computing and improve the environment quality. The work in [9], gives a brief overview of the state-of-the-art in green cloud computing. They examine existing research in the area and categorize it into different themes. They also discuss the challenges and opportunities in the field, and provide insights into future directions for research. This paper provides valuable background information and a significant understanding of the current landscape of green cloud computing. In the survey [10], the authors discuss different mechanisms for lowering the power utilization in DCs. It 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. Table I lists relevant research in the field of energy improved DCs categorized into relevance regarding the different DC sizes.

A. Our contribution

Our work contributes to the field by exploring numerous strategies for improving energy efficiency in DCs, while taking the different sizes of DCs into consideration. We aim to provide an in-depth overview of key challenges, opportunities and methods for improving energy efficiency in all types of DCs. The contributions of this paper are not only an overview of current research directions but also an overview of how proposed technologies and techniques can be realised in modern DCs, including:

- 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.

TABLE I. DC SIZE REFERENCES

DC Size	References
Large scale	[1] [2] [4] [5] [6] [7] [8] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19]
Mid scale	[2] [7] [8] [10] [13] [14] [15] [16] [17] [19] [20]
Small scale	[2] [8] [10] [13] [14] [15] [16] [17] [19] [20] [21] [22]

- Analyzing previous surveys and papers on energy efficiency in DCs, and provide a overview of their key findings and limitations. Our work highlights that there has barely been performed any research on how effective certain energy efficiency strategies have been for different size DCs.

III. 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 [13]. A common tool for measuring energy efficiency in DCs are the Power Usage Effiectiveness (PUE) metric. It is calculated by dividing the total amount of energy used by the DC, including all systems and components, by the energy used by the IT equipment within the DC [23].

This section will give a brief overview of multiple state-of-the-art 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. 3 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, dark blue for cooling systems optimization, red for DCIM tools and yellow is for Sleep State methods.

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
- Data Center Infrastructure Management (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 [22]. Another relevant method in this category is *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. This results in energy savings and reduced carbon emissions [24].

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 [14].

DCIM tools monitor, measure, manage and/or control data center 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 [25].

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 [15].

AI-driven DC Management is a method for automating control and monitoring of DC resources. By improving DC operations, energy efficiency improves as well [26].

Quantum computing have 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 is still in its early stages [27].

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. 3 illustrates what components in a DC that can be impacted by sleep state methods. 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 [13]. The devices and nodes that are involved when utilizing this method are highlighted in Fig. 3, marked by yellow. Modern processors support multiple types of sleep states, primarily:

- Core C-states
- Package C-states
- P-states
- DRAM (Dynamic Random-Access Memory) 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. C0 is the active state, with no CPU power savings. C1 is the state with the least power savings but with the shortest exit latency whereas C6 is having the longest exit latency at a $133\mu s$ transition time [13].

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 [16]. However, the concept is that additional power is saved compared to the power saved with the sub-components individually [16]. 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 [13]. 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) 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 [13].

P-states changes the frequency and voltage of a part of the system. This being the cores or other components such as a shared L3 cache [16]. P-state is a module state affecting a collection of cores that share resources [16]. 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 [16].

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. The MC is also behind the Self-refresh function as it sends the refresh signal to DRAM to ensure that the data is valid. DRAM does have the ability to start the Self-refresh process itself which can reduce the power consumption in the MC [28].

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. 3 illustrates resource utilization techniques within a DC, highlighted with as green.

1) Load balancing: 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 [11]. Irregular arrivals of requests impact the accuracy of the expected workloads, so 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 [17]. 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 [11].

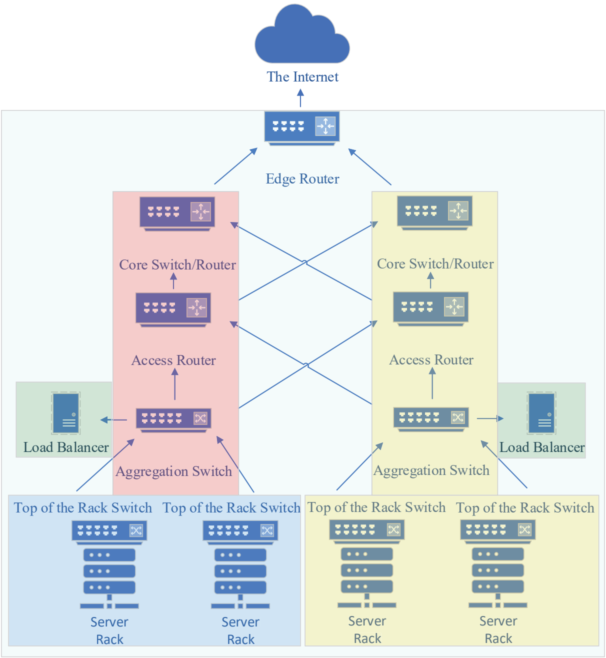


Figure 3. Resource utilization techniques highlighted in color for involved key components of a data center.

2) Scheduling: A cloud system uses virtualization technology to provide cloud resources such as CPU and memory to users in the form of virtual machines. 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 [18]. 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 a active state and resource utilization is higher. This method has been tested and was able to achieve up to 46% energy savings [18].

ISBN: 978-1-68558-036-0

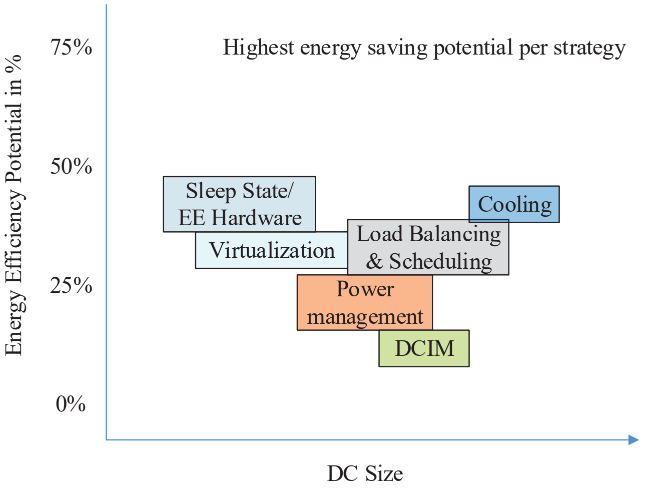


Figure 4. The graph compares the energy efficiency improvement percentages achieved through different energy efficiency strategies in DCs of different scale.

IV. 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 [20]. However, most attention has been directed towards the largest DCs, which are only responsible for a small portion of the overall energy consumption of DCs in general, since small-/mid-scale sized DCs being responsible for approximately 50% of the energy consumption [20]. 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 [20], 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 objective in place at all [20]. 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 three different sizes/categories. 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 DCs because of price and other circumstances. On the other hand, small-scale DCs might see greater gains 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.

A. Strategies at Different Scale

Many different factors are decisive for how effective certain strategies are when it comes to the energy efficiency for data centers 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 [20]. Fig. 4 shows potential power savings of different strategies for varying DC sizes.

Small-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 however, 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% [22], virtualization has proven possibilities of 30% [14], sleep state methods can provide up to 34% energy savings [19], and resource utilization methods can reduce energy consumption by up to 46% [18]. 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. 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. For smaller DCs it’s the complete opposite and operators of such facilities should therefore experience the implementation of this technology as less challenging. These technologies are new and are therefore still being researched, so it is not yet possible to provide potential power savings.

V. CONCLUSION

In conclusion, the scale of a DC can impact its energy efficiency. Small and medium-sized DCs can achieve notable energy-savings by improving design and infrastructure, as well as improving resource utilization, while large-scale DCs can make use of the greater amount of available resources to increase energy efficiency in the same and other ways. This is important since numerous factors impact how energy efficiency is achieved in each DC. This paper highlights the importance of considering DC-scale when estimating the potential impact of energy-saving strategies, as well as suggesting various methods for energy efficiency improvements for DCs of different sizes.

ACKNOWLEDGMENT

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.

REFERENCES

[1] 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.

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

[3] S. Moss. (18/05/2022) 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/

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] L. M. P. Larsen, H. Christiansen, S. Ruepp, and M. Berger, "Towards greener 5g and beyond radio access networks - a survey," *IEEE Open journal of the Communications Society*, 2023.

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

[26] 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.

[27] 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

[28] 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.