Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

June 2022 Volume XIV Number 2 ISSN 2061-2079

Technically Co-Sponsored by

IEEE ComSoc™ IEEE Communications Society

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.
Infocommunications Journal is also included in the Thomson Reuters – Web of ScienceTM Core Collection, Emerging Sources Citation Index (ESCI)

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section

Supporters

FERENC VÁGUJHELYI – president, Scientific Association for Infocommunications (HTE)

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1051 Budapest, Bajcsy-Zsilinszky str. 12, Room: 502
Phone: +36 1 353 1027
E-mail: info@hte.hu • Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Telecommunications and Media Informatics
Phone: +36 1 463 4189, Fax: +36 1 463 3108
E-mail: pvarga@tmit.bme.hu

Subscription rates for foreign subscribers: 4 issues 10.000 HUF + postage

Publisher: PÉTER NAGY

HU ISSN 2061-2079 • Layout: PLAZMA DS • Printed by: FOM Media

The metrics of Infocommunications Journal keep improving

Pal Varga

INFOCOMMUNICATIONS Journal has received its metrics for the last year, and we are happy to report that they all show an increasing trend. The citations have almost doubled, and the SCR index has improved 1.5 times compared to the previous period. The journal has kept its Q3 place that was gained last year. Although we have not received an official impact factor yet, the impact score is 1.16, which is a very good indicator – even makes us hopeful for the near future. We keep the scientific and publication standards of the Infocommunications Journal high and keep improving these metrics. The current issue is over 100 pages long, a record number for the journal.

Let us have a brief overview of the articles included in the current, second 2022-issue of the Infocommunications Journal.

In their paper, Ákos Leiter et al. present an extensive survey on the joint domain of PMIPv6 and SDN mobility management, detailing the available SDN-integrated network-based techniques and architectures. Their paper aims to be comprehensive, taking into account all possible architectures in the field, and provides a summarizing table of the key functional indicators related to the surveyed architectures and papers.

Regarding delay-constrained optical fronthaul design for 5G and beyond, Abdulhaim Fayad and Tibor Cinkler provide a cost-effective solution. They propose an integer linear program (ILP) that simultaneously optimizes BBU (baseband unit) number and placement together with the fronthaul deployment through simulation to achieve the goal of minimum CAPEX. Among the many results, they found that free-space optic solutions can be more cost-efficient than fiber optic when weather conditions are supportive.

Kotaru Kiran and D. Rajeswara Rao present an analytical review on vertical handover management technologies in 5G heterogeneous network setup. The authors categorized the research papers in their approach on handover, distinguishing them as radio access-based, self optimization-based, SDN-based, authentication-based, eNodeB-based, neural network-based and blockchain-based approaches, and discussed the various issues and research gaps within.

Tamás Kőnig and Lajos Nagy have been motivated to decrease the noise emission of switching-mode power supplies that are used on satellites. They simulated and measured the noise reduction effect of the interleaving method in the case of a system with two DC-DC Buck converters. They found that it is a complex problem and that the phase delay of the converters can be optimized in order to improve the noise reduction.

Mohammad Bawaneh and Vilmos Simon present a unique method on time series handling called Adaptive Simulated Annealing Representation (ASAR). The root of the idea is to consider the time series representation as an optimization problem where the objective is to preserve the time series shape and reduce the dimensionality. When comparing ASAR with other methods, they found it superior in compression ratio, as well as great computational performance when used for algorithms such as 1-NN classification and K-Means clustering.

Yasir Ahmed Idris Humad and Levente Dudás present the capabilities of the wideband spectrum monitoring system for the MRC-100 3-PocketQube satellite, especially in terms of measuring radio frequency signals, with the limited size, weight, and power consumption of the designed system. They found that this architecture has enough sensitivity and dynamic range for the purpose, even at a very wide bandwidth.

Zoltán Pödör and Anna Szabó introduce a practical framework and a complex application to manage synthetic sensor data. This framework can be used for generating, storing, and visualizing of real-looking synthetic sensor data sequences that can be used in various test scenarios.

The environment of education and professional meetings has changed dramatically in the last two years due to the worldwide COVID-19 pandemic. Besides providing a survey on Quality of Experience (QoE) for various video-transmission-related services, Tushig Bat-Erdene et al. analyzed the QoE of content sharing in online education and meetings in actual real-life scenarios and shared some fascinating findings in their paper.

Fatima Es-sabery et al. provides a comparative study of various feature extractors within the sentiment analysis domain that aims at recognizing and categorizing emotions in textual data. As one of the important results, they found that the combination of CNN+FastText outperforms all other combinations in terms of accuracy, precision, recall, and F1 measure.

With this brief overview, Infocommunications Journal wishes a great summer period, as well as peace and perseverance for the rest of 2022, to all its partners.

Pal Varga received his Ph.D. degree from the Budapest University of Technology and Economics, Hungary. He is currently an Associate Professor at the Budapest University of Technology and Economics and also the Director at AITIA International Inc. His main research interests include communication systems, Cyber-Physical Systems and Industrial Internet of Things, network traffic analysis, end-to-end QoS and SLA issues – for which he is keen to apply hardware acceleration and artificial intelligence, machine learning techniques as well. Besides being a member of HTE, he is a senior member of IEEE, where he is active both in the IEEE ComSoc (Communication Society) and IEEE IES (Industrial Electronics Society) communities. He is Editorial Board member of the Sensors (MDPI) and Electronics (MDPI) journals, and the Editor-in-Chief of the Infocommunications Journal.

Survey on PMIPv6-based Mobility Management Architectures for Software-Defined Networking

Ákos Leiter[1,2], Mohamad Saleh Salah[2], László Pap[2], and László Bokor[2,3]

Abstract—**Software-Defined Networking (SDN) has changed the network landscape. Meanwhile, IP-based mobility management still evolves, and SDN affects it dramatically. Integrating Proxy Mobile IPv6 (PMIPv6) – a network-based mobility management protocol – with the SDN paradigm has created several promising approaches. This paper will present an extensive survey on the joint research area of PMIPv6 and SDN mobility management by detailing the available SDN-integrated network-based techniques and architectures that intend to accelerate handover and mitigate service disruption of mobility events in softwareized telecommunication networks. The article also provides an overview of where PMIPv6 can be used and how SDN may help reach those ways.**

Index Terms—**PMIPv6, SDN, IP-based mobility management**

I. Introduction

Software-Defined Networking (SDN) [1] [2] introduces a new concept that does not leave any network services and functionalities untouched. SDN separates control and data plane of network traffic. The control plane has been centralized, and a new entity has come to life in the network: SDN Controller. From now, a custom program can run on an SDN Controller, which can direct any traffic. This leads to programable networks where "simple" switches are controlled by a central point dynamically.

IP-based mobility management is also affected by SDN. This paper presents a survey on how Proxy Mobile IPv6 (PMIPv6) [3], a particular case of IP-based mobility management, can be adapted to the new SDN world. Several research papers have started to examine the evolution of PMIPv6 from an SDN point of view.

There are many approaches to implement efficient mobility management in SDN environments. SDN provides scalable and dynamic schema, solving many limitations of legacy architectures related to mobility processes. Using SDN and the protocols like OpenFlow [4], the mobility management processes may become more manageable. One of the expectations of integration with SDN is to solve the Single-Point-of-Failure problem (SPOF). The second is to avoid triangle routing: traffic is not needed to pass through Home Agent/Local Mobility Anchor (HA/LMA). The next one is the initial delay, and handover latency can be decreased thanks to the SDN controller's ability to manage multiple forwarding devices simultaneously and add forwarding rules to them based on the needs. Furthermore, the usage of SDN can eliminate tunnel usage to decrease overhead. Finally, in some cases, the number of control messages can be mitigated [5][6] [7]. The paper will examine all of these aspects while surveying PMIPv6+SDN architecture proposals. Looking at other types of IP-based mobility management schemas with SDN is out of scope in this paper.

The remaining sections are organized as follows. Section II introduces a general overview of IP-based mobility management. An overview of the SDN paradigm and its technological evolution is placed in Section III. The surveyed literature is presented in Section IV. Section V presents details on the future research topics of SDN-PMIPv6 integration. A conclusion has been put in Section VI, while a table summarizing our findings can be found at the end of this paper.

II. Overview of IP-Based mobility management

This section presents how IP-based mobility management has evolved and what approaches have affected the evolution, resulting in the PMIPv6 standard and its extensions.

A. MobileI Pv4 and Mobile IPv6

In the case of host-based mobility management, users are actively involved in their mobility processes. MIPv4 [8] and MIPv6 [9] are the most popular and well standardized host-based mobility protocols operating in the network layer.

MIPv4 is designed to let Mobile Nodes (MN) use two IP addresses. The first identifies the MN's Home Network where a permanent IP address exists, called the Home Address (HoA). Home Agent (HA) is located in Home Network, responsible for tracking users' mobility and providing the globally accessible address for mobile nodes. The second IP address – assigned to an MN – is the Care-of Address (CoA), whose foreign network is the network where MN moves into during its movement. Foreign Agents (FA) are placed in the Foreign Networks, responsible for assigning CoA to MN and notifying HA of the particular MN's new location. In Mobile IP's terminology, Corresponding Node (CN) is considered any node outside the mobility domain and wants to connect with MNs.

The mobility management in MIPv4 works in the following way: if an MN is located in the home network, no additional steps are required than the standard IP communication procedure. The mobility starts when the MN moves to another network (Foreign

[1] Nokia Bell Labs, Budapest, Bókay János u. 36-42, 1083 Hungary
[2] Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[3] ELKH-BME Cloud Applications Research Group, BME Informatics Building, Magyar tudósok krt 2, H-1117 Budapest, Hungary
e-mails: akos.leiter@nokia-bell-labs.com, mhd.s.salah@gmail.com, pap@hit.bme.hu, bokorl@hit.bme.hu

DOI: 10.36244/ICJ.2022.2.1

Network): each time an MN moves out from its Home Network, it takes a new CoA from the Foreign Network range. After that, MN registers its CoA by sending a Registration Request message through FA to the HA. When finishing the registration, the packets from the MN to CN are sent directly from the MN through the foreign network. Packets originated at CN and targeted to the MN go through HA. HA tunnels the packets to FA. Finally, FA processes the encapsulated packets and forwards those to the MN. Figure 1 describes the Control flow of MIPv4. Figure 2 depicts the basic architecture of MIPv4.

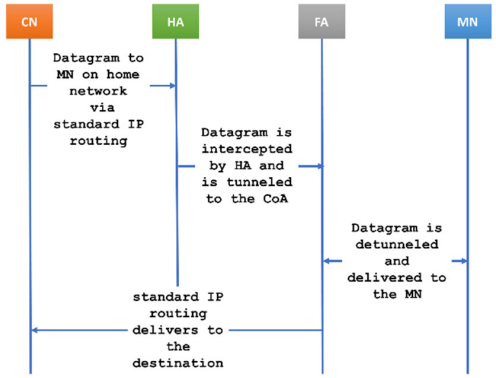


Fig. 1. The standard MIPv4 message flow[8]

The drawbacks of MIPv4 are the triangular routing which adds more latency, single point of failure (SPOF), and consumes bandwidth. In contrast, the traffic does not move directly between the sender and the receiver (CN and MN). Instead, traffic goes through the HA in the middle.

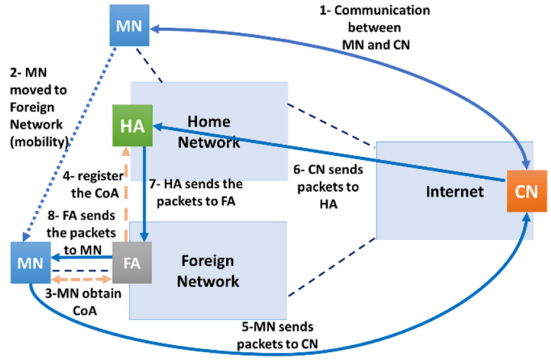


Fig. 2. The basic architecture of MIPv4 [8]

MIPv6 is similar to MIPv4, with enhancements and additional features. MIPv6 uses the Neighbor Discovery Protocol (NDP) of IPv6 [10]. NDP uses Router Solicitation (RS) and Router Advertisement (RA) messages to detect IP network prefix changes. Furthermore, NDP also deals with neighbor reachability. An IPv6 capable access router has replaced the functions of a Foreign Agent in MIPv4. This means FAs are eliminated in the context of MIPv6.

The mobility procedure in MIPv6 works as follows. The communication between MN and CN is addressed by native/ordinary IPv6 routing when MN stays on its Home Link. If the MN moves to Foreign Network, it has a new IP address called the CoA. After that, the MN sends a registration request to the HA (Binding Update) and receives the registration reply (Binding Acknowledgment). Traffic is encapsulated between HA and MN. MN may send a BU to CN to avoid triangle routing in route optimization mode (RO). The detailed message flow of MIPv6 is illustrated in Figure 3 .

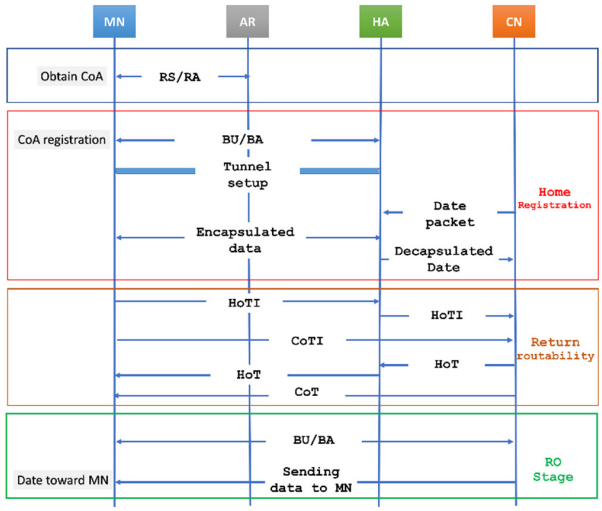


Fig. 3. The standard MIPv6 message flow [9]

Home Test Init (HoTI) and Care-of Test Init (CoTI) messages are part of the return routeability procedure. It is an authorization procedure to enable registration by a cryptographic token exchange. This procedure helps to give some assurance to CN if MN is reachable on that particular CoA. CN can securely accept BU from MN at the end of this procedure and circumvent HA (route optimization).

B. Proxy Mobile IPv6

Proxy Mobile IPv6 (PMIPv6) [3] is a network-based mobility management protocol working at the network layer. The network-based mobility management extends the network side and lets the network handle the mobility management instead of modifying the host part. Thus, MNs may not even know they are under any mobility process.

In PMIPv6 (Figure 4), the MN considers the whole PMIPv6 domain as a home network, so the MN uses just a unique HoA and different care-of addresses used by the MAGs. Mobile Access Gateway (MAG) and Local Mobility Anchor (LMA) are introduced in PMIPv6. MAG works as the access router; it detects the MN's movements and does the signaling and tunneling with the LMA, while the LMA works similarly to the HA in MIPv6 but with some additional potentials. LMA preserves accessibility to the MN's address as it travels through PMIPv6 domains. Binding Cache exists in the LMA, which is particularly a database that keeps track of the movement of MNs.

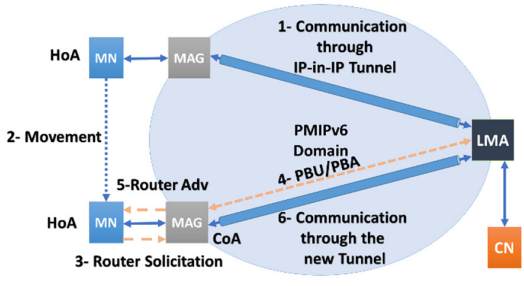


Fig. 4. The basic architecture of PMIPv6 [3]

PMIPv6 operates as follows. The MN attaches to MAG and sends Router Solicitation (RS) messages. Then MAG transmits a Proxy Binding Update (PBU) to the LMA, informing the attachment. LMA replies to the MAG via Proxy Binding Acknowledgment (PBA). MAG answers with Router Advertisement (RA) messages to the MN containing the home prefix of the MN as a reply for RS. Finally, a bidirectional tunnel between LMA and MAG is created to let the MN communicate with the CN, as depicted in Figure 5. When a handover happens (Figure 6), the new MAG detects the MN attachment. PBU/PBA messages are exchanged between the new MAG (nMAG) and LMA containing the latest information. The LMA updates its Binding Cache, and a new bi-directional tunnel will be created between the LMA and the MAG. RS/RA are also sent between MN and MAG. Meanwhile, the previous MAG (pMAG) sends PBU deregistration to the LMA, informing MN detachment. An authentication, authorization, and accounting (AAA) server implementation can supervise and manage network access procedures.

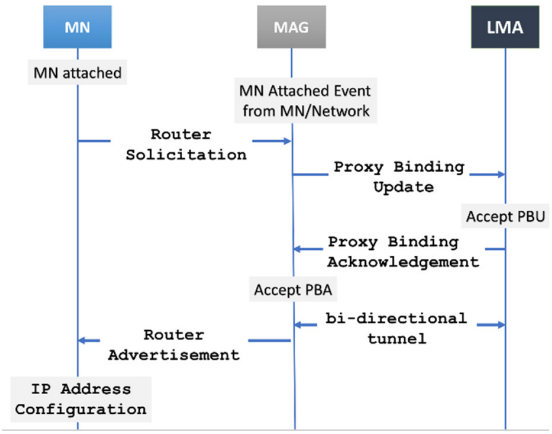


Fig. 5. PMIPv6 attachment control flow [3]

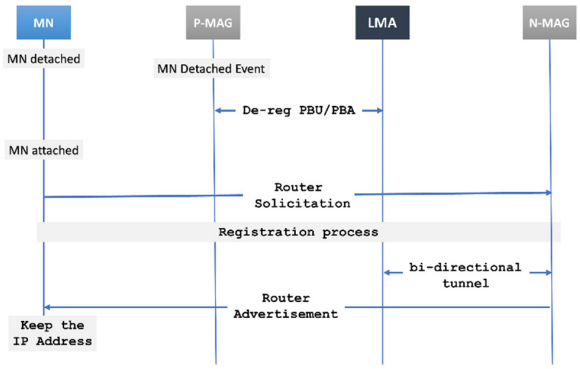


Fig. 6. PMIPv6 handover control flow [3]

PMIPv6 supports IPv4 [11] and IPv6 and doesn't require any MN side modification. The drawbacks of PMIPv6 include the tunneling overhead, the single point of failure, and high throughput in the case of LMA.

Table 2 of [12] shows a detailed comparison between IP-based mobility management protocols. The correspondent sections have been put to this paper as Table 1.

TABLE I
SUMMARY OF MIPV4, MIPV6, AND PMIPV6 [12]

Category	MIPv4	MIPv6	PMIPv6
Mobility scope	Global	Global	Local
Mobility management	Host-based	Host-based	Network-based
Network architecture	Flat	Flat	Hierarchical
Target network	IP	IP	IP
Operating layer	L3	L3	L3
Required infrastructure	HA & FA	HA	LMA & MAG
MN modification	Yes	Yes	No
Router advertisement	Broadcast	Broadcast	Unicast
Addressing model	Shared-prefix	Shared-prefix	Per-MN-prefix
MN address	HoA	HoA	CoA
Address type	IPv4	IPv6	IPv6
Address length (bits)	32	128	128

C. Flow Mobility

A flow is a set of packets matching a certain Traffic Selector (TS [13]), and flow mobility management aims to apply the mobility for each communication flow individually [14] [15]. MN can bind different CoAs (Multiple Care-of Address, MCoA [16]) for each flow to ensure individual handling of a particular flow. One of the benefits of using PMIPv6 is that MN does not have to be modified as MN does not even know that it is under mobility management. However, with PMIPv6 flow mobility extensions, end-user modification is required.

Flow mobility also supports using multiple interfaces for an MN. This leads to the traffic offload and intelligent flow handling topics: based on the interfaces' specific properties; flows can be routed differently even though they originate from the same device. For example, using 4G/5G interface can be costly because of the radio spectrum price, but 3GPP interfaces provide more reliability than ordinary Wi-Fi. Therefore, those flows, like web browsing with lower expectations from the network compared to, e.g., videoconference, can be offloaded to non-3GPP accesses. The procedure to select a suitable interface is out of the scope of this paper, but it can rely on, e.g., QoS metrics, radio resource availability, etc. Section II.H has some connection statements for this section where corresponding references can be found too.

Mobile IPv6 Flow Binding extension has the same signaling and architectural design as "ordinary" MIPv6, depicted in Figure 7, Figure 8. Flow Bindings functions are implemented with a novel Mobility Option called the Flow Identification Mobility Option, attached to the BU and BA messages. This identifies a particular flow (with Traffic Selector sub-option). It is possible to bind multiple flows by the same CoA or different CoAs with multiple interfaces too. The more flows are bound, the more BU/BA signaling messages are needed.

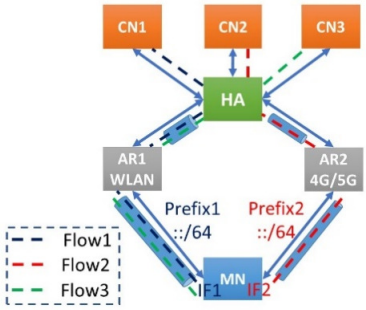


Fig. 7. Mobile IPv6 Flow Bindings architecture

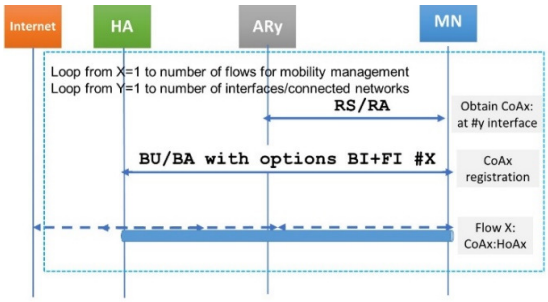


Fig. 8. Mobile IPv6 Flow Bindings signaling flow

Figure 9 and Figure 10 illustrate an example of MN shared prefix across two physical interfaces.

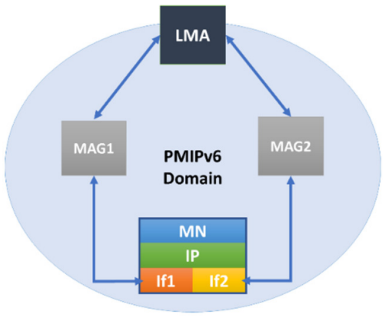


Fig. 9. Shared prefix across physical interfaces in PMIPv6 flow mobility [15]

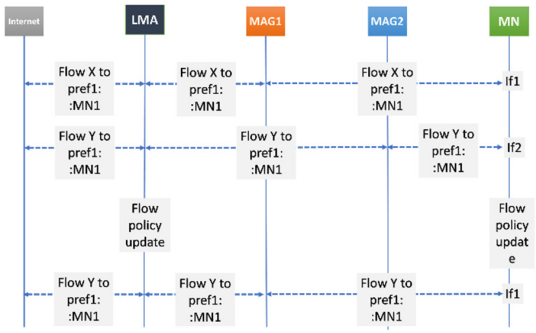


Fig. 10. Flow Mobility Message Sequence with a Common Set of Prefixes [15]

Figure 11 is an example of extended PMIPv6 where flow mobility is enabled. The MN is attached to the network by two interfaces (WLAN interface and 3G interface), and there is a different prefix for each interface. In this case, the LMA is extended to support grouping a set of mobility bindings and refer it to the same MN [17].

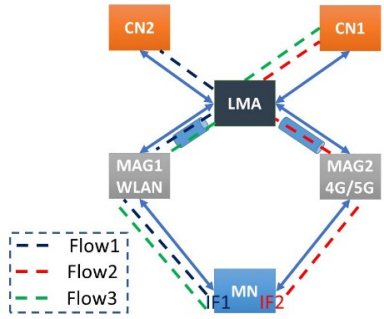


Fig. 11. PMIPv6 flow mobility with 3GPP/non-3GPP offloading

D. Dynamic Mobility Management

Dynamic Mobility Management provides the needed mobility support to a user only when required. This saves resources by reducing unnecessary mobility management signaling overhead and network cost. In the case of IP-based mobility management, e.g., it means avoiding BU/BA message exchange and eliminating tunneling overhead [18][19][20]. This works well when IP-address preservation is not a goal. For those types of traffics, when the change of IP address causes service disruption, this is not a valid path. If the IP address is changed at a VPN service, the VPN itself is broken. But for ordinary web browsing ("surfing"), it does not affect the user experience negatively. Dynamic Mobility Management can be combined with Flow Bindings too. With this combination, the individual dynamicity of particular flows can be realized. Consequently, tunneling overhead can be decreased, too, as the tunnel is not set up if mobility management is not needed.

Figure 12 is an example of dynamic mobility management where the MN moved to a new network, and a new IP (IP2) was assigned to it. Many applications establish connections after the mobility and do not need mobility support. In contrast, the applications that require mobility support forward the traffic with the old IP (IP1) to the AR1 (the AR1 works as HA in this case). To sum up, the system will provide mobility only based on the need [18].

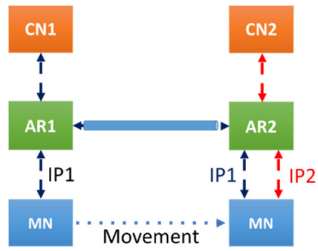


Fig. 12. Dynamic Mobility Management [18]

E. Distributed Mobility Management

Distributed Mobility Management (DMM) refers to the idea of using multiple mobility management functions instead of a centralized one and distributing them over different locations [21]. The nearest mobility function probably serves the MN, as in Figure 13. Distributed Mobility Management could be across different levels: core level, access router level, access level, and host level, and this distribution could be organized partially or fully. By using the distributed mobility management architec-

ture, unnecessarily long routes are avoided, and more scalability is provided to the network. Furthermore, the SPOF issue can be solved, which is one of the biggest problems for HA and LMA to make them carrier-grade.

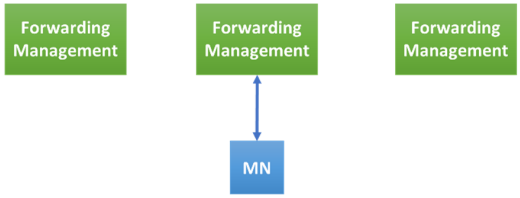


Fig. 13. Distributed mobility management approach [21]

An example of fully distributed PMIPv6 is illustrated in Figure 14. In this case, the standard PMIPv6 signaling messages are used. However, when the CN sends a packet to its MAG, this MAG sends Proxy Binding Query (PBQ) [22] to all MAGs in the domain. Only the MAG that connected the MN will reply with Proxy Binding Query Acknowledgment (PBQA). Then the data packets will be forwarded to the MN through the related MAG.

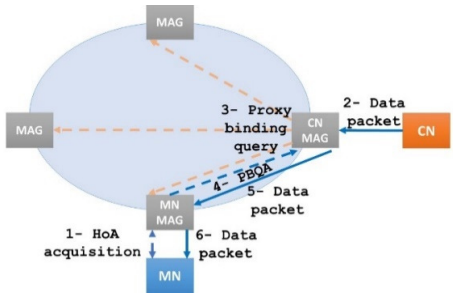


Fig. 14. Distributed PMIPv6 [23]

Dynamic Mobility Management and On-Demand Mobility Management should be differentiated slightly in the context of Distributed Mobility Management. DMM can mean dynamic anchor selection for a particular MN [24]. However, On-Demand Mobility Management pertains to having session continuity requirements [25].

F. *Types of Handovers*

One of the promises of SDN in the IP-based mobility context is to simplify and/or accelerate handovers. There are several types of handovers related to PMIPv6 [26]:

- Intra-domain handovers: the same MAG is used, but a user switches between different access links. Even the type of access network can be changed.
- Inter-MAG handovers: a user changes its corresponding MAG. In this case, MAG will initialize a new binding (PBU/PBA).
- Inter-LMA handovers: LMA and MAG are changed at the same time. A completely new registration process with their occasional AAA dialog is executed. External routing is also readjusted as the anchor point to the outside world (i.e., the LMA) is changed. So, there is a new path ultimately to reach a particular MN.

The L1 and L2 technology-agnostic design of IP and IP-based mobility management are supporting technologies of inter-technology handovers.

G. *Supporting technologies*

With the introduction of IP-based mobility management, IP addresses enjoy the separation of identification and locator roles (ID/Loc separation). End-user devices used to be always on the exact location, so a well-defined IP prefix allocation scheme determined the location and identity of nodes. This is where ID/Loc separation comes into the picture. The MIPv6 address family is a good example, but not the only one. The Host Identity Protocol (HIP) [27] introduces a new layer between the IP and the Transport layer for solving the problems mentioned above. A new identifier, Host Identifier (HI), is created to give unique keys to mobile nodes. In the context of MIPv6 and PMIPv6, HIP has been started to get experimented with and standardized [27][28][29][30].

The Media Independent Handover (MIH) [31] was introduced in IEEE 802.21 and aimed to permit handovers between different heterogeneous technologies like Wi-Fi and cellular technologies without disruption of service, leading to enhanced user experience. The MIH defines a set of Service Access Points (SAP) in the link layer that maps to a generic interface between the different link-layer technologies and the upper layers. MIH also provides a group of mobility functions to support the upper layers mobility protocols [31].

H. *3GPP support*

PMIPv6 has also been incorporated by the 3GPP standards [32] [33]. There are two basic approaches to use PMIPv6 in 3GPP concepts:

- Relying on the connection of non-3GPP access networks to 3GPP networks: S2 interface family. In most cases, this means connecting a Wi-Fi network to a service provider network (SP-Wi-Fi). There are two subdomains differentiated by the owner of the Wi-Fi network devices. These devices can either belong to an SP or a 3rd party company.
- Using PMIPv6 on S5/S8 interface instead of the GPRS Tunneling Protocol (GTP).

Several papers and standards deal with traffic offloading of mobile networks with the help of Wi-Fi networks and PMIPv6, e.g.: [34][35][36].

III. Overview of Software-Defined Networks and corresponding mobility management

Software-Defined Networking is one of the most popular topics these days in terms of networking technologies. It is considered as the evolution of legacy networking architecture (Figure 15). Before SDN, each networking device is a fix- device with a built-in control plane and data plane. Then SDN came up with the idea of separating the control plane for all networking devices, making it a centralized controlling programable unit "SDN controller" and keeping the data plane inside the networking devices. In other words, the networking devices just forward data based on flow tables, while the SDN controller manages the preferences [1][2][4].

SDN architecture consists of three layers [37]:

- Infrastructure Layer, where the data plane and networking devices work;
- Control Layer, where the network services and the controlling process are kept;

- Application Layer that connects to the Controlling layer via APIs and contains the business applications where policies and rules could be applied to the controller.

SDN commonly uses OpenFlow protocol as a connection channel between the SDN controller and the forwarding devices. Using OpenFlow, SDN controllers can put the flow table entities inside the forwarding devices, and these devices forward the data based on the flow table information.

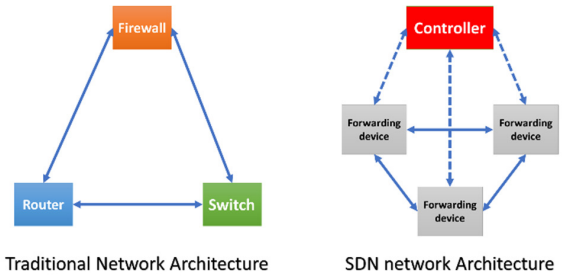


Fig. 15. Overview of traditional and software-defined network architectures

The concept of integrating SDN with mobility management has come up several times in the literature. In this part, we focus on the overall paradigm of mobility management and SDN combination. However, some references deal with IPv6-based mobility management as well.

D Liu et al. [38] present problem statements regarding SDN integration of IP-based mobility management.

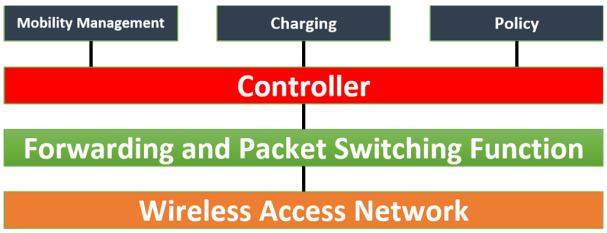


Fig. 16. SDN-based mobile core network [38]

As Figure 16 envisions, they see the mobile network core on SDN-bases, where the SDN controller has networking functions (charging, policy, etc.). Mobility management is also one of the SDN controller-integrated network functions. This approach promises to simplify mobility management in general. As the SDN controller has the full view of the network, it can optimize traffic routing/forwarding and catch mobility events. This leads to Figure 17, which shows an architecture for MIPv4/MIPv6 integration with the SDN controller functions.

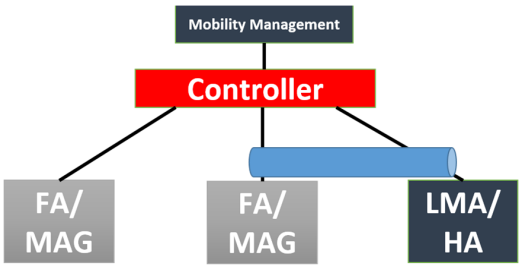


Fig. 17. Enhance SDN to support mobility tunnel processing [38]

Forwarding mechanisms can catch mobility events and notify the Mobility Management application on the top of the SDN controller. Based on the traffic directions, it can calculate optimal routing paths also from MN/HA and CN points of view. This paper does not present detailed architecture proposals and signaling flows on those results; however, it shows and lists open research questions on this topic.

SDN has also influenced Distributed Mobility Management. Hanuel Ko et al. [39] propose an SDN-DMM architecture. They state there are several problems in the context of IP-based mobility management which can be solved or mitigated by using SDN. These identified problems are:

1. lack of dynamic mobility support;
2. suboptimal routing;
3. scalability issues;
4. single-point-of-failure.

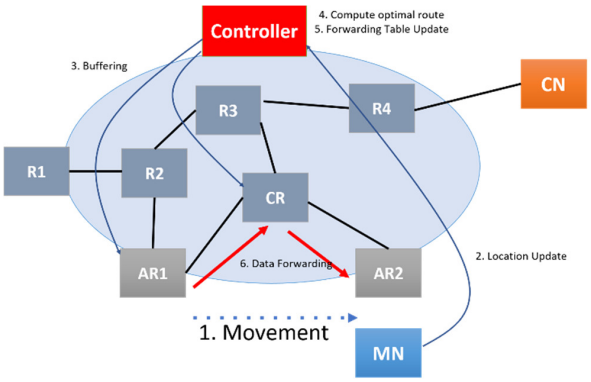


Fig. 18. Distributed Mobility Management with SDN (SDN-DMM) [39]

Figure 18 shows how handover can be executed in the context of SDN-DMM. Their assumption is that there is continuous data traffic between MN and CN. Step 1 pertains to the MN movement where MN changes its attachment point. MN must report its new location (Step 2). At Step 3, the Controller sends Openflow buffering messages to Crossover Router (CR) and Access Router1 (AR1) entities. Meanwhile, in Steps 4 and 5, the Controller computes a new optimal path for the CN-MN traffic through AR2. Buffered traffic is sent to AR2 and MN in Step 6. With this new route calculation, the usage of AR1 is left out, and only AR2 is used. Furthermore, with proper Openflow rules, tunneling may be eliminated.

Tien-Thinh-Nguyen et al. [40] have also dealt with SDN and IP-based mobility management integration. Their proposal's name is S-DMM pertains to SDN and Distributed Mobility Management. They measured that similar performance can be achieved in S-DMM compared to the traditional DMM in terms of handover latency and end-to-end delay. However, the complexity of the overall system control plane is reduced.

In the upcoming sections, we believe the above-introduced SDN concepts can serve as a solid base of understanding for discussing SDN and PMIPv6 integration – a more specific and focused integration case.

IV. The analyzed literature

This section intends to present the studied papers in a timely order. However, at the end of the paper, a summary table (Table 3) indicates the key functional properties of architectural proposals. With the help of Table 3, we believe that the Readers can get a quick but enough broad view and ensure suitable grouping of the most critical aspects of the proposals. Furthermore, this section goes into the details of the SDN-PMIPv6 integration approaches paper-by-paper deeply.

Paper #1 by Seong-Mun Kim et. al (2014 Jan)[41]

Seong-Mun Kim et al. [41] propose a solution where PMIPv6 is integrated with Openflow. Their proposal (OPMIPv6) separates the mobility management functions from PMIPv6 components. Furthermore, this allows the removal of PMIPv6 tunnels with their overheads. The LMA function could be located at the controller; the MAG function could be placed either into the controller or the access switch (Figure 19). There is a case when LMA and MAG are on the same centralized controller node (OPMIP6-C); thus, PMIPv6 signaling should not be used. Openflow is possible to avoid IP tunneling because the data path is set up by Openflow protocol through flow tables in the forwarding devices (switches). Signaling between controller and switches is transferred via a secured channel. In this proposal, the data plane is configured by LMA, located on the controller. It is possible to add multiple controllers to provide redundancy to the system. They proved that OPMIPv6 performs better than PMIPv6 in means of tunneling overhead, the resiliency of failures, and handling capacity.

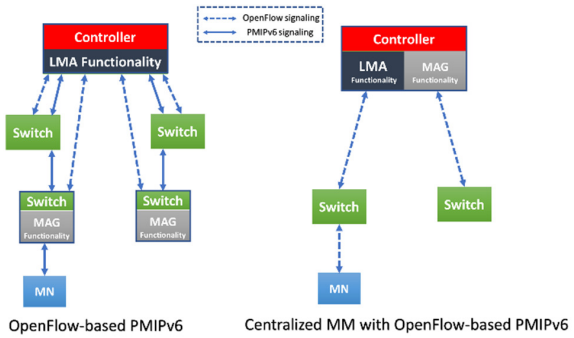


Fig. 19. Seong-Mun Kim et al.'s deployment scenarios[41]

Paper #2 by Kyoung-Hee-Lee (2014 Jan) [42]

Kyoung-Hee Lee [42] extends PMIPv6 with Routeflow to make it SDN-ready. Routeflow[43] is an SDN deployment framework used to handle IP routing protocols (Figure 20). Using a centralized server containing several VMs where each VM represents a programable switch and a routing protocol working between these VMs to create a forwarding information base (FIB), the central server collects the needed data (IP and ARP tables) to build OpenFlow rules. It translates those data to OpenFlow rules installed after processing the forwarding devices. The solution keeps the PMIPv6 signaling and concepts. LMA and MAGs are installed into separate VMs. Each VM carries out the mobility role for a particular mobility node. If an inter-MAG handover happens, standard PMIPv6 messages are exchanged, but the Routeflow server translates the Binding Cache of LMA into Openflow rules. These new Openflow rules are distributed to the switches then. So PMIPv6 Control plane is kept, but the Routeflow server maintains another control channel to manage switches. The routing table of VMs of MAGs is also continuously translated to the underlying Openflow switches by the Routeflow server. In this case, Routeflow is a controller. This scenario suggests some modification on the control message flow of PMIPv6 to fit the underlying SDN architecture.

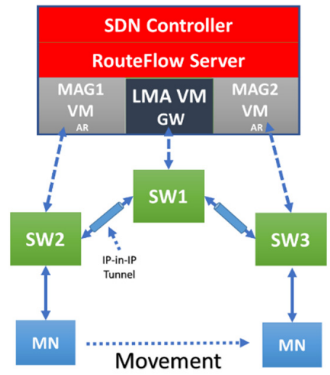


Fig. 20. Kyoung-Hee Lee's proposed architecture [42]

Paper #3 By Syed M. Raza et al. (2014 Jan)[44] [45]

Syed M. Raza et al. [44][45] introduce the OF-PMIPv6 concept as an integration of PMIPv6 and Openflow (Figure 21). In their proposal, the signaling path and the data path are separated. OpenFlow Mobile Access Gateways OMAG entity is introduced, which is only responsible for L2 functionalities and manages the IP tunneling. Meanwhile, L3 (signaling) messages are handled by the controller. PMIPv6 is not modified at all; messages and tunnels are still used in this solution. Their main goal is to decrease handover latency. OMAG communicates with the controller via Openflow.

They introduce three new message types to the Openflow protocol:

- The controller sends Tunnel_init to the next OMAG (nOMAG) to create an IP tunnel from the nOMAG to the anchor.
- S_Report is used for reporting link states of Mobile Nodes to the Controller.
- L_Report is sent by OMAG to report the loss of a Mobile Node to the Controller after a specific time.

This is an extra step of control messaging, but the authors state that it means the minimal effect on handover latency and packet loss based on their performance evaluations.

The OF-PMIPv6 controller communicates with OMAGs and LMAs as well. On the controller, there are three modules proposed:

- OpenFlow Module uses the Controller's built-in functions to communicate the PMIPv6 control and mobility-related messages with OMAG.
- PMIPv6 Module takes responsibility for performing standard PMIPv6 control messages with the anchor and the AAA system.

- Mobility Management Module has a connectivity database (C-DB) that stores the Mobile Nodes' information (MN ID, LMA ID, attached OMAG ID, and MN link-state values from OMAGs).

The C-DB is a kind of Binding Cache. The authors propose two types of handover:

- Reactive: When RS message is received, OMAG forward OF-RS message to controller, not PBU to LMA.
- Proactive: OMAG monitors the link continuously; as soon as the values drop below the lower threshold, it reports it to the controller. Then the C-DB is updated by the controller.

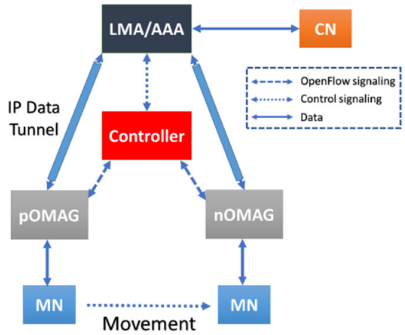


Fig. 21. Syed M. Raza et al.'s architecture proposal[44]

Paper #4 By Hoa Yu (2014 May) [46]

Hao Yu [46] suggests a solution for inter-domain heterogeneous vertical handover in SDN environments using PMIPv6. The SDN controller handles the tunnel creation between MAGs and LMAs from different domains in the proposal. This is required because the original LMA maintains home Network Prefix. If an MN moves to another domain, the traffic is tunneled back to the original LMA to maintain connectivity.

By using SDN, the vertical handover between different domains is possible. When an MN is moved to a new domain, the MAG in the new domain discovers the MN attachment and performs the mobility signaling, depicted in Figure 22. Meanwhile, the old LMA deregisters the MN from the old domain. When the new LMA receives a PBU from the new MAG, it discovers from the PBU that the MN moved from another domain. There is communication between the LMAs from different domains.

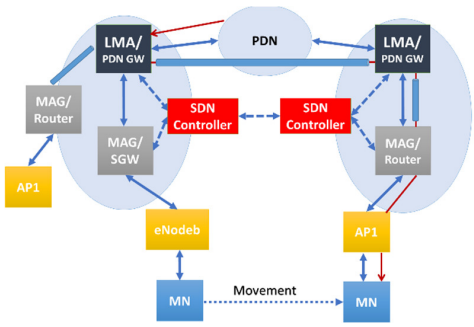


Fig. 22. Hao Yu's architecture proposal [46]

Paper #5 by You Wang et al. (2014 Aug) [47]

You Wang et al. [47] present an Openflow-based architecture for IP-based mobility and discuss how SDN can help evolve mobility management. Their proposal is not a PMIPv6 enchantment and integration to the SDN world. Instead, it is a new mobility approach that behaves similarly to PMIPv6 with taking advantage of SDN. The MN does not own the MN's CoA; the first-hop Openflow switch maintains it. This means MN does not need to take care of address reconfiguring like at PMIPv6. The MN has a non-routable HoA used to look up for MN's current location. One additional role for the controller is to maintain a binding cache that matches the HoA and CoA of the MN.

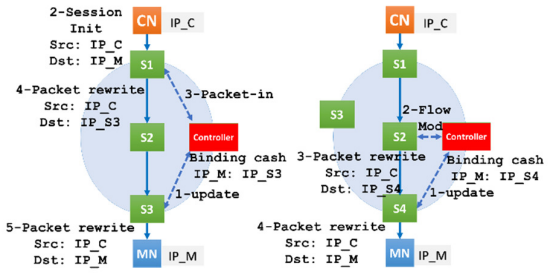


Fig. 23. You Wang et al.'s architecture proposal (left is communication initialization; right is movement handling) [47]

In Figure 23, the IP_M is the HoA, and IP_C is the CN address.

When the first attachment occurs, the switch which detects the MN arrivals (S3) sends a BU to the controller containing IP_M and IP_S3. As the controller has the binding cache locally, it will store the new updates and sends a new rule to the switch to replace all packet's destination addresses to IP_S3 to IP_M.

When CN sends a packet to the HoA of MN (IP_M), the switch (S1) does not know where to forward it, and the switch will ask the controller. The controller checks its local binding cache. Then it will send a flow rule to the first-hop-switch of CN to replace all the packets' destination addresses to IP_S3 from IP_M, which are directed to MN.

During handover, almost the same process happens. The new switch (S4) sends BU to the controller; in this case, the controller sends updates to all switches in the path between CN and MN. This changes all the flow directed to IP_S3 to be IP_S4.

Thus, the network takes care of binding caching, no need to have it on MN, and that's why it is similar to PMIPv6. Also, triangle routing is solved because packets do not need to pass through HA/LMA/Controller. Binding Cache can also be put in several parts of the network. The authors propose an algorithm to find the optimal place. The paper discusses the case of using multiple controllers and Dual mobility.

Paper #6 Yuta Watanabe et al (2015 Jan) [48]

Yuta Watanabe et al. [48] address the problem where LMA is overloaded because every traffic is directed to go through it. With the help of Openflow, they construct a path that avoids LMA to CN. Also, they get rid of tunneling and propose to use only the Openflow toolset without tunneling.

The OpenFlow switch represents the MAG in this solution.

The paper does not include any architecture diagrams, and the messages sequence flow diagram is miss ordered, so we re-created the figure with the corrections we believe represent the target of the paper (Figure 24). The paper concludes that their Openflow-based path optimization has higher throughput than ordinary PMIPv6 or a PMIPv6 route optimization proposal (PRO) [49].

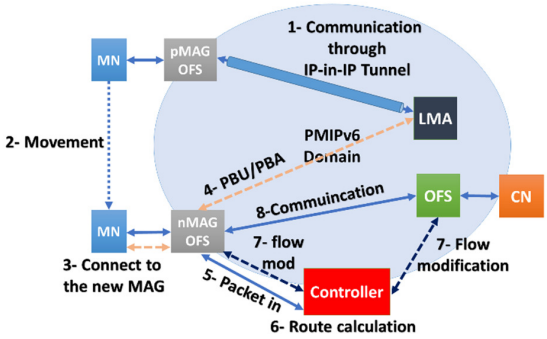


Fig. 24. Yuta et al. 's architecture proposal by the authors of this paper [48]

Paper #7 Sakshi Chourasia et al. (2015 Apr) [50]

Sakshi Chourasia et al. [50] present an Openflow-based improvement for EPC networks (Figure 25). They focus on decreasing signaling overhead during handovers (intra-LTE, inter-RAT). They have a logically centralized controller for the EPC control plane to manage mobility, called EPC Controller. SGWs and PGWs are replaced with Openflow switches, controlled by EPC Controller. EPC controller also has connections to eNodeB-s via TCP links. EPC Controller is also responsible for end-user authentication and IP address allocation. Furthermore, it also supports charging procedures. The OpenFlow switches handle the IP mobility and forward the packets based on preferences provided by the controller. They state that tunneling overhead can be eliminated with the usage of Openflow.

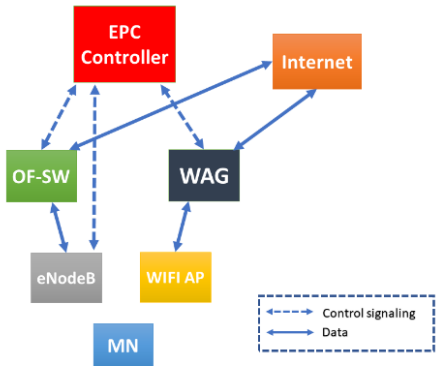


Fig. 25. Sakshi Chourasia et al.'s Openflow integration model to EPC [50]

Paper #8 Seong-Mun Kim et. al (2015 Apr) [51]

Seong-Mun Kim et al. [51] have a broader scope in this paper about their previous proposal, already presented at Paper #1. This paper has not given any novel architectural elements compared to the previously presented one because it mainly focused on evaluations. However, a summarized table comparing ordinary PMIPv6 and their OPMIPv6 solution is worth mentioning (Table 2). It is a good basepoint for making architectural conclusions.

TABLE II
COMPARISON OF PMIPV6 AND OPENFLOW-BASED PMIPV6 [51]

	PMIPv6	OPMIPv6
LMA	A dedicated router	A normal switch with LMA function and a controller
Multiple LMA	Limited to gateways with LMA functions	Any gateway router connected to the LMA function with a controller
Resilience to failure	Difficult	Easier than PMIPv6 by replicating LMA controller architecture
Flexibility to workload increase	Rigid	Modular installment of the LMA controller is possible
Changing the primary LMA gateway	Not supported	No LMA change is required in a domain; a gateway router can be easily changed
MAG	A dedicated router	A normal IP router with LMA function
Separation of control and data plane	Not supported	Supported
Tunneling	IP-in-IP tunneling	No tunneling with Openflow architecture
Handover delay	PMIPv6 signaling delay	PMIPv6 signaling and flow table setup time for all routers on the flow path
Dual role agent	Not practical	LMA and MAG can be combined into a single mobility management controller

The comparison and the numerical results of the paper show clearly that the cost of tunnel elimination comes with increased handover delay. This is due to the added Openflow signaling (new TCP connections) to the existing PMIPv6 signaling. But Table 2 presents new added features and resiliency by Openflow, which are worth the additional cost from an architectural point of view.

Paper #9 Abbas Bradai et al. (2015 Jun) [52]

Abbas Bradai et al. [52] proposed a solution called Software-Defined Mobility Management (SDMM) influenced by the PMIPv6 architecture. In this solution, the mobility management entities are virtual machines (V_LMA and V_MAG) collocated with the SDN controller, depicted in Figure 26. The SDN controller is responsible for creating optimal tunnels for mobility. V_LMA and V_MAG are accountable for receiving control messages (e.g., router solicitation), and they take action of tunnel creation. The tunnel is created between particular Openflow-enabled switches. They eliminate the usage of PBU/PBA. A separate Mobility Database (MD) is connected to the SDN controller, where bindings are tracked. MD has a well-defined interface to the SDN controller, which immediately translates mobility events to Openflow rules for corresponding switches.

V_LMA VM determines the most convenient forwarding entity to work as a mobility anchor point in the network. At the same time, V_MAG is mapped to forwarding functions to make them work as MAGs. After choosing the best anchor point, a tunnel will be created between the LMA and MAG.

They suggest having dynamic anchor entities in the network for each set of flows to achieve the best results in terms of delay and throughput balancing. A study and modeling of choosing the best anchor point are introduced in the paper.

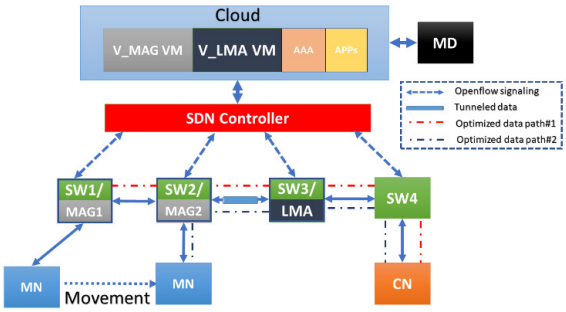


Fig. 26. Abbas Bradai et al. 's Software-Defined Mobility Management architecture [52]

Paper #10 *Wen-Kang Jia (2015 Nov) [53]*

Wen-Kang Jia [53] has proposed an SDN-based PMIPv6 architecture extension for Evolved Packet Core (EPC). The paper also deals with inter-domain handover with the help of the architecture proposal. Also, route optimization is presented to the CN from MN. The overview of the system architecture is depicted in Figure 27.

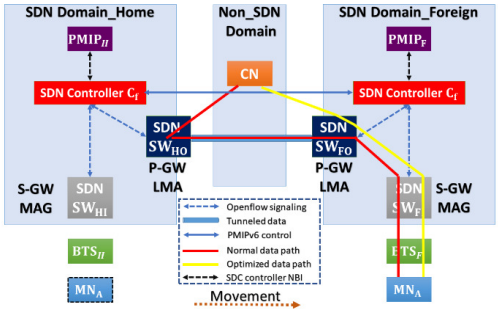


Fig. 27. Wen-Kang Jia's system architecture [53]

MAG is placed to SGW, collocated with an SDN switch (SW home input, SWHI) as an SDN application. LMA is on PGW, which is also collocated with an SDN switch (SW home output, SWHO). There is a central SDN controller (CH) in the domain. When an MN attaches to the radio link (Base Transceiver Station (BTS)), the RS messages go to the central SDN controller forwarded by SWHI. The PMIP function of CH generates the corresponding RA to the MN. It also verifies if that particular MN can join the domain or not. Capabilities also can be set up. A virtual PBU is sent to the LMA function to update its BC on SDN SWHO. Then the SDN Controller updates the forwarding rules of switches via Openflow. Meanwhile, MN acquires its HoA. Attaching to the home network does not use any kind of PBU/PBA message in this architecture proposal.

When the MN visits a foreign network, the foreign SDN controller (CF) should update the home SDN controller (CH) about the MN's attachment. Between controllers, ordinary PBU/PBA is used to exchange information. Data is tunneled back to the home network, but SDN controllers can negotiate route optimization to avoid the home network when reaching a CN from the visited network.

Paper #11 *A. Aissioury et al. (2015 Dec) [54]*

A. Aissioui et al. [54] extend the Follow-Me-Cloud (FMC) concept with PMIPv6. FMC is about using the available nearest data center to the user. User sessions are always moved to that data center, and VMs in this data center provides the user's services. In this case, not just users move, connected services too. This work assumes that clouds are federated, so there is a control level connection between each cloud. There are two global entities: Inter-Domain Mobility Database (IDMD) and Follow Me Cloud Controller (FMCC). IDMD is responsible for maintaining PMIPv6-related actions (registrations and movement details) while FMCC can select the appreciated data center. Particularly, Binding Cache is outsourced to IDMD. After PMIPv6 registration procedures, FMCC is notified about the movement of MNs. This can be the trigger to relocate service between clouds. FMCC is the SDN controller, and actually IDMD is a supportive database to handle mobility. Service movement decision is made via Decision Making Application Module (DMAM) and Mapping Information Gateway (MIGW). DMAM is responsible for concluding whether service movement is required even while MIGW keeps and maintains the mapping between PMIPv6 entities and the underlying SDN-ready network. The overall architecture can be seen in Figure 28.

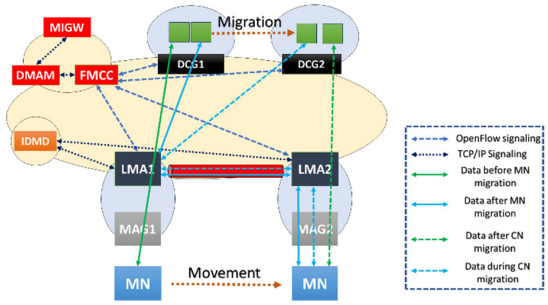


Fig. 28. A. Aissioui et al.'s proposed architecture for PMIPv6-based FMC [54]

Paper #12 *Pill-Won Park et al. (2016 Feb) [55]*

Pill-Won Park et al. [55] proposed an OpenFlow-Based Mobility Management (OMM) implementation using the PMIPv6 manner to locate mobility management requirements. Their architecture proposal consists of Mobility Management Entity (MME), which existed in the SDN controller and handles all mobility management functions (Figure 29). Three layers of switches are considered:

- Access Switches (ASs),
- Intermediate Switches (ISs),
- Gateway Switches (GWs).

A A A server connects to MME. Binding Cache, Flow Matrix, and GW-HNP mapping table are the relevant data structures of MME for supporting mobility management. GW-HNP mapping table is used for mapping the HNP (Home Network Prefix) with the actual gateway.

Flow Matrix is responsible for saving flow paths of the MNs; also, it saves previous AS, CS, and list of HNPs. The flow paths are saved as pairs of upstream and downstream. After selecting the GW, the MME search for the flow path between this GW and the AS is connected to the MN. When the handover occurs, the

flow tables are updated in the switches, as the flow path of the new AS stored in the flow matrix.

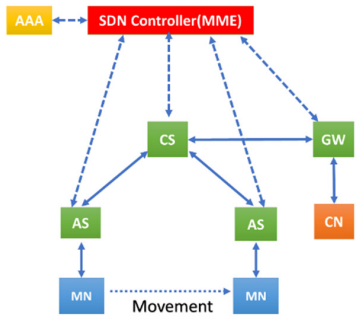


Fig. 29. System Architecture by Pill-Won Park et al. [55]

Paper #13 Kuljaree Tantayakul et. al (2016 Mar) [56]

Kuljaree Tantayakul et al. [56] main goal is to prove mobility management can be done without using PMIPv6 in an SDN-ready network (Figure 30). They propose SDN Mobility Service introducing new SDN signaling-based methods. Only two components are considered: controller and access routers (AR). Consequently, there is no tunnel usage.

Two new SDN operations were introduced: MN registration and MN handover. Both rely on exiting Openflow implementation, and they take the place of the mobility management signaling role.

The performance of PMIPv6 and their SDN-based proposal have been measured in terms of UDP throughput, TCP, and packet loss ratio. In every field, the proposal performed better.

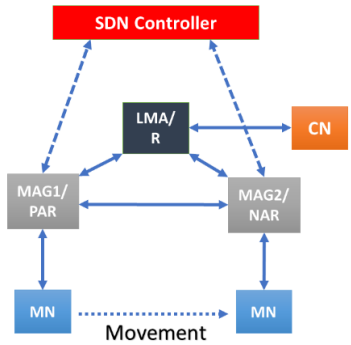


Fig. 30. Kuljaree Tantayakul et al. 's SDN mobility architecture proposal [56]

Paper #14 Ce Chen et al. (2016 Apr) [57]

Ce Chen et al. [57] propose a Mobility SDN scheme (M-SDN) to reduce the handover latency. Their proposal eliminates IP tunneling referred to in the paper by Raza et al., (Paper #3). They introduce the N-casting phase to accelerate handover, a preparation phase of handovers where every possible target is considered and prepared. Besides ordinary Openflow components, they use a Location Server to keep track of users' movements.

They assume that there are two types of handovers based on SDN domains (each domain has its controller, Figure 31):

- intra-domain handover
- inter-domain handover

There is an application called Mobility Application on the controller, which consists of two parts: Core module and REST interface. The Core module handles mobility-related maintenance: location tracing, flow redirection. The REST interface is the bridge between the core module and all other components. New signaling messages have been introduced to take care of mobility management.

One thing to highlight: there is no mention of using IPv6. One of the core components is DHCP. From this, we assume this is an IPv4-based solution, but the underlying ideas come from OF-PMIPv6.

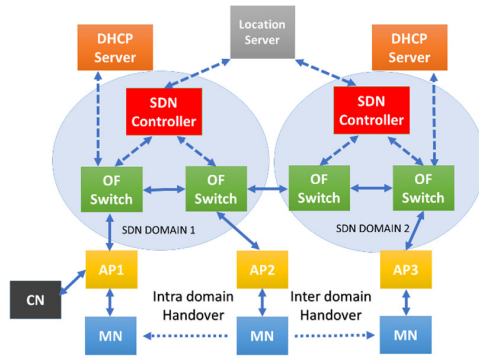


Fig. 31. Ce Chen et al.'s M-SDN proposed architecture [57]

Paper #15 Syed M. Raza et al. (2016 Jun) [58]

Syed M. Raza et al.[58] proposed a solution for inter- domain IP mobility with route optimization using PMIPv6 based on the SDN environment. Their work has a solution to the limitation of serving only one domain in PMIPv6 and SDN-PMIPv6, summarized in Paper #3.

They suggest having a communication channel between the controllers in different domains. When the handover occurs, the new controller checks the MN ID with the other controllers to confirm that this MN is in a new device or attached to another domain before, so the controller can decide to register the MN as a new device or trigger the mobility operation.

The architecture of this proposal is illustrated in Figure 32.

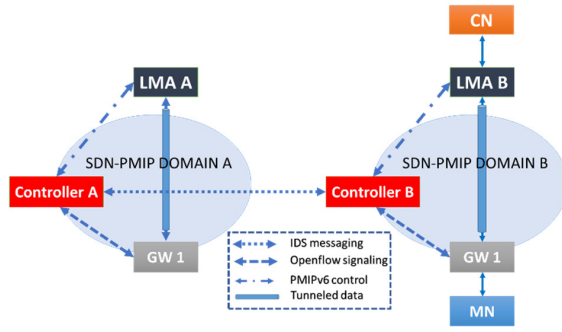


Fig. 32. Syed M. Raza et al.'s architecture [58]

This solution is an extension of Paper #3. They together solve inter-domain handover and addresses triangle routing problems in the context of PMIPv6 and SDN integration. This paper uses the new naming of the solutions: SDN-PMIPv6, instead of OF-PMIPv6 at Paper #3.

Paper #16 *Walla F Elsadek et. al (2016 Aug) [59]*

Walla F Elsadek et al.[59] aim to propose an LTE- independent inter-domain handover. They utilize the SDN concept inspired by PMIPv6. Their proposal creates a virtual path to the MN's home network with an SDN mobility overlay called "Three Tier Mobility Overlay". They proposed to hide the L3 complexity by using these virtual paths.
SDN controller maintains users' profile data. There is an L4 selective breakout possibility to the Internet to avoid core network overload at this framework. This can be seamlessly integrated into the existing CAPWAP or PMIPv6 deployments. New mobility entities introduced in this paper (Figure 33):

1. Mobility Access Switch Tier (AS): provides the connection between the access aggregation layer and a broadcast domain.
2. Mobility Detector Switch Tier (DS): two types of this switch are introduced. The Foreign DS and this switch work as a foreign agent that provides services to the MN at the foreign network, and Home DS work as the home agent and guarantees security.
3. Mobility Gateway (MG) and Relay Switch (RS) Tier: those entities take care of connecting overlay in intra-domain mobility. The difference between MG and RS is the connected overlay managed by a different or same SDN controller.

Their proposal pushes toward a complete SDN solution for handling mobility management.

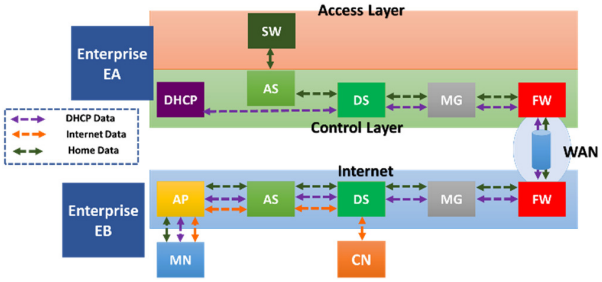


Fig. 33. Walla F Elsadek et al. 's proposed mobility scenario [59]

Paper #17 *Syed M Raza et al. (2017 Jan) [60]*

Syed M. Raza et al. [60] proposed a solution for inter- domain mobility in the SDN PMIPv6 environment using an on-demand mobility concept: Inter-Domain SDN-PMIP (IDS- PMIP). Using multiple prefixes for the MN makes it possible to have a seamless handover between two domains. The main idea is to keep the old prefix for the already active sessions, and all new sessions will build based on the new prefix. A new field called proxy info is added to the Controller Binding Cache (CBC). This field contains proxy tuples; those proxy tuples have the MNs prefixes and the controllers' addresses, which assign those prefixes. There is a communication (Border Gateway Protocol (BGP) -based) between the controllers from different domains for discovering the previous domain for the MN.

When an MN attaches to an SDN-PMIP domain, it sends RS to the first Openflow-enabled gateway (OGW), depicted in Figure 34. OGW forwards RS to the domain controller, which parses the MN-ID to conclude AAA. If AAA is successful, the controller sends Discovery Request (DISC- REQ) to all the neighboring controllers. Controllers answer with Discover Response (DISC-RESP): if there is a matching entry for the particular MN in their CBC, they respond with that entry (MN-ID, used prefixes). Otherwise, it is left empty. This is also a solution for first-time registration: if the field is empty, it implicitly tells it is not an inter-domain handover. After having this discovery procedure completely, PBU can be sent to the actual LMA, which belongs to the particular domain. When PBA is received from LMA, the controller sets up tunnels between OGW and LMA. This means LMA is not responsible for setting up the tunnel even though one of the tunnel endpoints is LMA.

When inter-domain handover happens (Figure 35), DISC-RESP is not empty. The previous controller (controller A) deregisters the MN from its domain. But two prefixes are assigned at this moment to the MN. The new LMA (domain 2) starts advertising both prefixes after successfully executing the recent PBU/PBA events. New connections are assigned to the new prefix (Prefix B). The usage of double IPv6 prefixes is the method or cost to make handover smother.

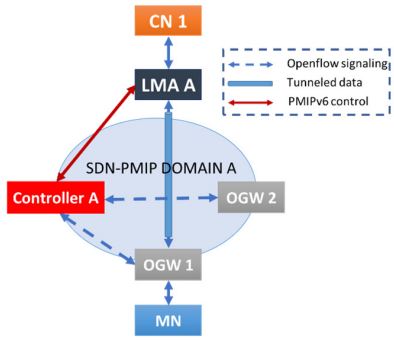


Fig. 34. Mobile Node registration [60]

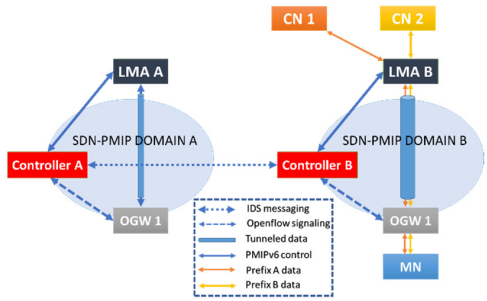


Fig. 35. Inter-domain handover [60]

Paper #18 *N. Omheni et al. (2018 Jan) [61]*

N. Omheni et al. [61] propose a partially distributed mobility management in the SDN context using PMIPv6.

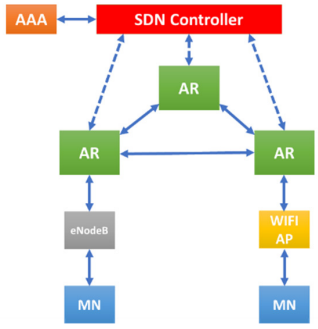


Fig. 36. N. Omheni et al. 's architecture proposal for DMM [61]

They have a three-layer architecture, shown in Figure 36. The first is the access layer where SDN-enable WIFI access points and LTE interfaces are placed, the devices in the access layer can be programmed. Distributed Mobility Management – Access Router (DMM-AR) comes at the second layer, particularly Openflow capable switches that provide connectivity to the access layer and contribute to the mobility management process. The third layer is dedicated to the controller and other management-related entities like AAA. The SDN controller act as PMIPv6 LMA and the DMM-AR Act as the MAG. Their proposal divided the operation into two stages, the preparation and registration Stage and the handover execution stage.

Paper #19 *Syed M. Raza et al. (2019 Oct) [62]*

Syed M. Raza et al. [62] extended the previously solutions introduced on (Paper #15; Paper #17) with evolved architecture called on-demand inter-domain SDN-PMIPv6 (OIS-PMIPv6). Some advantages came with this solution, like decreased handover latency, improved resources utilization, and added scalability to the system. System architecture can be seen in Figure 37.

New Controller-to-Controller Communication Protocol (C3) was introduced in the paper. It enables session establishment between two controllers to facilitate prefix discovery. Also, a prefix retrieval technique is used to release and return IPs in OnDemand mobility.

The Data Plane Gateway(DP-G) refers to the MAG. CBC mess ages contain new fields to control the mobility in multiple domains; they refer to this field as a "Proxy", including tuples defined as prefix and the home controller of the prefix.

The MN has a new prefix in each domain, and all new sessions establish with this prefix. After inter-domain handover, the IPv6 address with the old prefix remains active in the new domain while the sessions using that IPv6 address stay active while all-new sessions establish with the newly assigned prefix.

The old sessions keep working as follows: the uplink traffic toward CN goes through the new anchor as the destination is the same, while the downlink traffic goes through the old anchor. This uplink/downlink traffic helps in improving the overall delay

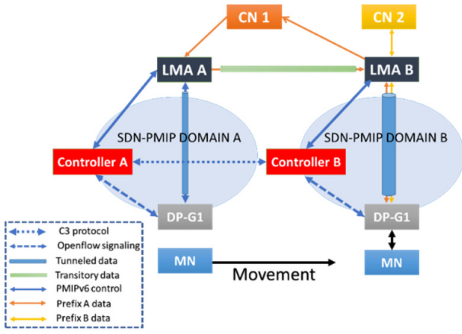


Fig. 37. Syed M. Raza et al.'s testbed for handover tests [62]

V. OUTLOOK ON FURTHER PMIPV6 EVOLUTION

PMIPv6 is not just examined in the narrow context of mobile telecommunication. One of the foreseeable future research topics and application possibilities is V ehicular communication (V2X). IPv6 Wireless Access in Vehicular Environment (IPWAVE) [63] [64] is a concept to facilitate vehicular communication in IP networks. Here, communication between the infrastructure and vehicles must be maintained to avoid service disruption. IP-Road Side Unites (IP-RSU) are placed next to the roads; they can act like MAGs and communicate with IP OnBoard Unites (IP-OBU), which are built into the cars. Mobility Anchors (MA) in this architecture behave like an LMA. When a vehicle moves to an area where another IP-OBU serves, then IP-based mobility management can avoid service disruption. IPW A VE architecture is depicted in Figure 38.

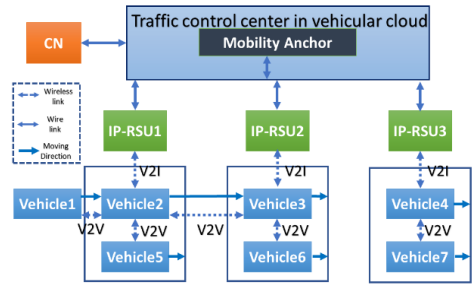


Fig. 38. Example of vehicular network architecture for V2I and V2V [63]

The aviation industry has also been examining the usage of IP- based mobility management. Aeronautical Telecommunication Network with IP (ATN/IP) may also utilizes PMIPv6 [65] [66][67][68], main concepted can be seen in Figure 39.

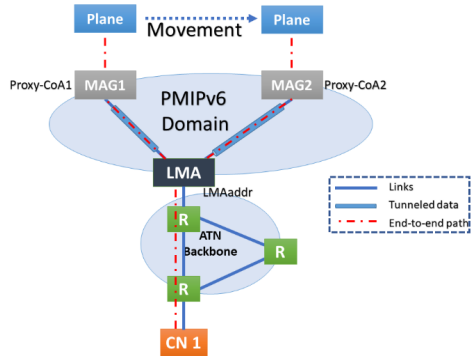


Fig. 39. Aeronautical telecommunication handover [66]

PMIPv6, as a network-based approach, is also intended to solve a mobility problem, specialized when an airplane moves from one aviation control system to another. There were even experiments to change IPv6 in IPv6 tunnel of PMIPv6 to MPLS in aeronautical networks [68].

PMIPv6 has been started to be examined for integrating with 5G networks. Kyoungjae Sun et al.[69] present a distributed PMIPv6 integration to 5G User Plane Functions (UPF) where core and edge sites are considered, presented in Figure 40. They also follow the 5G system design guidelines to separate control and user plane. Distributed LMA (DLMA) is located on the edge side, which receives control messages of PMIPv6; meanwhile, an edge UPF handles the data plane. Their goal is to exchange session handling and QoS based on specific types of IPv6 addresses, defined in RFC 8653 [25].

But SDN brings in a new approach for PMIPv6 too. Including the above-mentioned specialized (emerging) applicabilities and the "ordinary" telecommunications usage (working with 5G, non-3GPP traffic offload). Thus, the domain of PMIPv6 presents

several new research topics with SDN. This will not just contain the architecture redesign of PMIPv6 in the SDN context but examine the dynamicity, flow-based, and distributed manner of the SDN-PMIPv6.

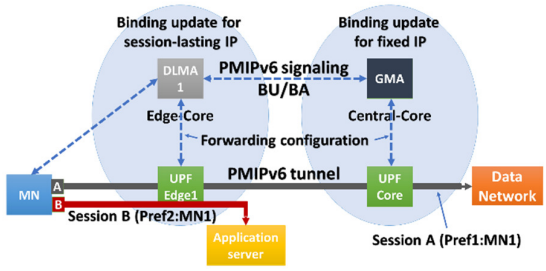


Fig. 40. PMIPv6 based distributed session mobility management [69]

5G is accelerating the cloudification of network services, and Mobile IPv6 and Proxy Mobile IPv6 are not exceptions in this trend [70] [71]. But it highlights new aspects of networking like energy consumption or Total Cost of Ownership (TCO) questions. To our best knowledge, these questions have not yet been addressed so far in the overall IP-based mobility management context. However, with 5G, there are several publications to deal with it (e.g., [72][73][74][75][76]). The ones above present new ways for further research directions of IP-based mobility management. This paper has also concluded that flow mobility has not been considered together so far in a PMIPv6 + SDN environment where multiple interfaces are used. This also opens further research possibilities.

VI. Conclusion

To the best of our knowledge, this paper surveyed all the architectures focusing on PMIPv6 – SDN integration solutions available in the literature.

In Table 3, we summarized the essential characteristics of all the analyzed architectures. The biggest differentiators are whether the architecture proposal keeps PMIPv6 control plane signaling or relies on tunneling elimination. Most of the papers also use Openflow for SDN implementation. From the point of view of standard compatibility and interoperability with legacy systems, keeping the PMIPv6 control plane is crucial.

Deployment of SDN or Openflow controllers has brought in a new type of Single-Point-of-Failure problem. Even though LMA SPOF weakness is solved with SDN, the network operator should solve the SDN controller problem. The reliability of SDN controllers is out of the scope of this paper, but a full-scale, deployment-ready solution must deal with that issue. The surveyed studies clearly show that standardization work must also be considered as inter- (administrative)domain handover needs a common base.

Interesting, but 4G and 5G network compatibility are not widely examined and considered in the surveyed papers. Furthermore, cloud computing compatibility has rarely been mentioned in the available literature. We think this definitely will be an important future resource direction.

References

[1] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, 2015, DOI: 10.1109/JPROC.2014.2371999.

[2] S. Sezer et al., "Are we ready for SDN? Implementation challenges for software-defined networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 36–43, 2013, DOI: 10.1109/MCOM.2013.6553676.

[3] S. Gundavelli (Ed.), K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, *Proxy Mobile IPv6*. Fremont, CA, USA: RFC Editor, 2008. DOI: 10.17487/RFC5213.

[4] N. McKeown et al., "OpenFlow: Enabling Innovation in Campus Networks," *SIGCOMM Comput Commun Rev*, vol. 38, no. 2, pp. 69–74, Mar. 2008, DOI: 10.1145/1355734.1355746.

[5] H. Modares, A. Moravejosharieh, J. Lloret, and R. B. Salleh, "A Survey on Proxy Mobile IPv6 Handover," *IEEE Syst. J.*, vol. 10, no. 1, pp. 208–217, 2016, DOI: 10.1109/JSYST.2013.2297705.

[6] W.-W. Chen, L.-H. Yen, C.-L. Chuo, T.-H. Heish, and C.-C. Tseng, "SDN-enabled session continuity for wireless networks," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6. DOI: 10.1109/ICC.2017.7997199.

[7] M. Al-Khalidi, N. Thomos, M. J. Reed, M. F. AL-Naday, and D. Trossen, "Anchor Free IP Mobility," *IEEE Trans. Mob. Comput.*, vol. 18, no. 1, pp. 56–69, 2019, DOI: 10.1109/TMC.2018.2828820.

[8] C. Perkins (Ed.), "IP Mobility Support for IPv4, Revised." RFC Editor, Fremont, CA, USA, Nov. 2010. [Online]. Available: https://www.rfc-editor.org/rfc/rfc5944.txt

[9] C. Perkins (Ed.), D. Johnson, and J. Arkko, "Mobility Support in IPv6." RFC Editor, Fremont, CA, USA, Jul. 2011. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6275.txt

[10] T. Narten, E. Nordmark, W. Simpson, and H. Soliman, *Neighbor Discovery for IP version 6 (IPv6)*. Fremont, CA, USA: RFC Editor, 2007. DOI: 10.17487/RFC4861.

[11] R. Wakikawa and S. Gundavelli, "IPv4 Support for Proxy Mobile IPv6." RFC Editor, Fremont, CA, USA, May 2010. [Online]. Available: https://www.rfc-editor.org/rfc/rfc5844.txt

[12] H. Tuncer, S. Mishra, and N. Shenoy, "Review: A Survey of Identity and Handoff Management Approaches for the Future Internet," *Comput Commun*, vol. 36, no. 1, pp. 63–79, Dec. 2012, DOI: 10.1016/j.comcom.2012.07.017.

[13] G. Tsirtsis, G. Giarreta, H. Soliman, and N. Montavont, "Traffic Selectors for Flow Bindings." RFC Editor, Fremont, CA, USA, Jan. 2011. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6088.txt

[14] S. Gundavelli (Ed.), K. Leung, G. Tsirtsis, and A. Petrescu, "Flow-Binding Support for Mobile IP." RFC Editor, Fremont, CA, USA, Aug. 2015. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7629.txt

[15] C. Bernardos (Ed.), "Proxy Mobile IPv6 Extensions to Support Flow Mobility." RFC Editor, Fremont, CA, USA, May 2016. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7864.txt

[16] R. Wakikawa (Ed.), V. Devarapalli, G. Tsirtsis, T. Ernst, and K. Nagami, "Multiple Care-of Addresses Registration." RFC Editor, Fremont, CA, USA, Oct. 2009. [Online]. Available: https://www.rfc-editor.org/rfc/rfc5648.txt

[17] T. Melia, C. J. Bernardos, A. de la Oliva, F. Giust, and M. Calderon, "IP Flow Mobility in PMIPv6 Based Networks: Solution Design and Experimental Evaluation," *Wirel. Pers. Commun.*, vol. 61, no. 4, pp. 603–627, Dec. 2011, DOI: 10.1007/s11277-011-0423-3.

[18] H. A. Chan, H. Yokota, J. Xie, P. Seite, and D. Liu, "Distributed and Dynamic Mobility Management in Mobile Internet: Current Approaches and Issues," *JCM*, vol. 6, pp. 4–15, 2011.

[19] Chang Woo Pyo, Jie Li, and H. Kameda, "Simulation studies on dynamic and distributed domain-based mobile IPv6 mobility management," in *The 11th IEEE International Conference on Networks, 2003. ICON2003.*, Oct. 2003, pp. 239–244. DOI: 10.1109/ICON.2003.1266196.

[20] M. Kalyanaraman, S. Seetharaman, and S. Srikanth, "Dynamic and integrated approach for proxy-Mobile-IPv6 (PMIPv6) based IP Flow Mobility and offloading," in *2015 International Conference on Computing and Communications Technologies (ICCCT)*, Feb. 2015, pp. 356–361. DOI: 10.1109/ICCCT2.2015.7292775.

[21] H. Chan (Ed.), D. Liu, P. Seite, H. Yokota, and J. Korhonen, "Requirements for Distributed Mobility Management." RFC Editor, Fremont, CA, USA, Aug. 2014. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7333.txt

[22] H. JUNG, M. GOHAR, J.-I. KIM, and S.-J. KOH, "Distributed Mobility Control in Proxy Mobile IPv6 Networks," *IEICE Trans. Commun.*, vol. E94.B, no. 8, pp. 2216–2224, 2011, DOI: 10.1587/transcom.E94.B.2216.

[23] J. I. Kim and S. J. Koh, "Distributed mobility management in proxy mobile IPv6 using hash function," in *The International Conference on Information Networking 2013 (ICOIN)*, 2013, pp. 107–112. DOI: 10.1109/ICOIN.2013.6496360.

[24] P. Bertin, S. Bonjour, and J. Bonnin, "A Distributed Dynamic Mobility Management Scheme Designed for Flat IP Architectures," in *2008 New Technologies, Mobility and Security*, 2008, pp. 1–5. DOI: 10.1109/NTMS.2008.ECP.8.

[25] A. Yegin, D. Moses, and S. Jeon, "On-Demand Mobility Management." RFC Editor, Fremont, CA, USA, Oct. 2019. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8653.txt

[26] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia, "Fast Handovers for Proxy Mobile IPv6." RFC Editor, Fremont, CA, USA, Sep. 2010. [Online]. Available: https://www.rfc-editor.org/rfc/rfc5949.txt

[27] R. Moskowitz, P. Nikander, P. Jokela (Ed.), and T. Henderson, *Host Identity Protocol*. Fremont, CA, USA: RFC Editor, 2008. DOI: 10.17487/RFC5201.

[28] R. Moskowitz (Ed.), T. Heer, P. Jokela, and T. Henderson, "Host Identity Protocol Version 2 (HIPv2)." RFC Editor, Fremont, CA, USA, Apr. 2015. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7401.txt

[29] T. Henderson (Ed.), C. Vogt, and J. Arkko, "Host Mobility with the Host Identity Protocol." RFC Editor, Fremont, CA, USA, Feb. 2017. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8046.txt

[30] G. Iapichino and C. Bonnet, "Host Identity Protocol and Proxy Mobile IPv6: A Secure Global and Localized Mobility Management Scheme for Multihomed Mobile Nodes," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, 2009, pp. 1–6. DOI: 10.1109/GLOCOM.2009.5426039.

[31] A. D. L. Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An overview of IEEE 802.21: media-independent handover services," *IEEE Wirel. Commun.*, vol. 15, no. 4, pp. 96–103, 2008, DOI: 10.1109/MWC.2008.4599227.

[32] "3GPP TS 23.402: Architecture enhancements for non-3GPP accesses." https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=850

[33] "3GPP TS 29.275: Proxy Mobile IPv6 (PMIPv6) based Mobility and Tunnelling protocols; Stage 3." https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1693

[34] Á. Leiter and L. Bokor, "A Flow-Based and Operator-Centric Dynamic Mobility Management Scheme for Proxy Mobile IPv6," *Wirel. Commun. Mob. Comput.*, vol. 2019, p. 4567317, Apr. 2019, DOI: 10.1155/2019/4567317.

[35] D. R. Purohith, K. M. Sivalingam, and A. Hegde, "Network architecture for seamless flow mobility between LTE and WiFi networks: testbed and results," *CSI Trans. ICT*, vol. 7, no. 1, pp. 45– 59, Mar. 2019, DOI: 10.1007/s40012-019-00214-1.

[36] S. Gundavelli (Ed.), X. Zhou, J. Korhonen, G. Feige, and R. Koodli, "IPv4 Traffic Offload Selector Option for Proxy Mobile IPv6." RFC Editor, Fremont, CA, USA, Apr. 2013. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6909.txt

[37] W. Braun and M. Menth, "Software-Defined Networking Using OpenFlow: Protocols, Applications and Architectural Design Choices," *Future Internet*, vol. 6, no. 2, pp. 302–336, 2014, DOI: 10.3390/fi6020302.

[38] D. Liu, "RFC expired draft – Mobility Support in Software Defined Networking," Jul. 08, 2013. https://datatracker.ietf.org/doc/html/draft-liu-sdn-mobility-00 (accessed Feb. 01, 2022).

[39] H. Ko, I. Jang, J. Lee, S. Pack, and G. Lee, "SDN-based distributed mobility management for 5G," in *2017 IEEE International Conference on Consumer Electronics (ICCE)*, 2017, pp. 116–117. DOI: 10.1109/ICCE.2017.7889250.

[40] T.-T. Nguyen, C. Bonnet, and J. Harri, "SDN-based distributed mobility management for 5G networks," in *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–7. DOI: 10.1109/WCNC.2016.7565106.

[41] S. Kim, H. Choi, P. Park, S. Min, and Y. Han, "OpenFlow-based Proxy mobile IPv6 over software defined network (SDN)," in *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Jan. 2014, pp. 119–125. DOI: 10.1109/CCNC.2014.6866558.

[42] K. Hee Lee, "Mobility Management Framework in Software Defined Networks," *Int. J. Softw. Eng. Its Appl.*, vol. 8, no. 8, pp. 1–10.

[43] *Routeflow: an open source project virtualized IP routing services over OpenFlow enabled hardware*. Accessed: Feb. 01, 2022. [Online]. Available: https://routeflow.github.io/RouteFlow/

[44] S. M. Raza, D. S. Kim, and H. Choo, "Leveraging PMIPv6 with SDN," New York, NY, USA, 2014. DOI: 10.1145/2557977.2558056.

[45] S. M. Raza, D. S. Kim, D. Shin, and H. Choo, "Leveraging proxy mobile IPv6 with SDN," *J. Commun. Netw.*, vol. 18, no. 3, Art. no. 3, Jun. 2016, DOI: 10.1109/JCN.2016.000061.

[46] Y. Hao, "Softward-Defined Networking in Heterogeneous Radio Access Networks," presented at the TNC14.

[47] Y. Wang and J. Bi, "A solution for IP mobility support in software defined networks," in *2014 23rd International Conference on Computer Communication and Networks (ICCCN)*, 2014, pp. 1–8. DOI: 10.1109/ICCCN.2014.6911783.

[48] Y. Watanabe, Y. Nakamura, and O. Takahashi, "A method to improve network performance of Proxy Mobile IPv6," in *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, 2015, pp. 74–75. DOI: 10.1109/ICMU.2015.7061036.

[49] M. Liebsch, L. Le, and J. Abeille, "Route Optimization for Proxy Mobile IPv6," Internet Engineering Task Force, Internet-Draft draft-abeille-netlmm-proxymip6ro-01, Nov. 2007. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-abeille-netlmm-proxymip6ro-01

[50] S. Chourasia and K. M. Sivalingam, "SDN based Evolved Packet Core architecture for efficient user mobility support," in *Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft)*, Apr. 2015, pp. 1–5. DOI: 10.1109/NETSOFT.2015.7116148.

[51] S. M. Kim, H. Y. Choi, Y. H. Han, and S.-G. Min, "An adaptation of proxy mobile IPv6 to openflow architecture over software defined networking," *IEICE Trans. Commun.*, vol. E98B, no. 4, pp. 596–606, Apr. 2015, DOI: 10.1587/transcom.E98.B.596.

[52] A. Bradai, A. Benslimane, and K. D. Singh, "Dynamic anchor points selection for mobility management in Software Defined Networks," *J. Netw. Comput.* Appl., vol. 57, pp. 1–11, 2015, DOI: 10.1016/j.jnca.2015.06.018.

[53] W.-K. Jia, "Architectural Design of an Optimal Routed Network- Based Mobility Management Function for SDN-Based EPC Networks," in *Proceedings of the 11th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, New York, NY, USA, 2015, pp. 67–74. DOI: 10.1145/2815317.2815326.

[54] A. Aissioui, A. Ksentini, and A. Gueroui, "PMIPv6-Based Follow Me Cloud," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015, pp. 1–6. DOI: 10.1109/GLOCOM.2015.7417766.

[55] P.-W. Park, S.-M. Kim, and S.-G. Min, "OpenFlow-Based Mobility Management Scheme and Data Structure for the Mobility Service at Software Defined Networking," *Int. J. Distrib. Sens. Netw.*, vol. 12, no. 3, Art. no. 3, 2016, DOI: 10.1155/2016/3674192.

[56] K. Tantayakul, R. Dhaou, and B. Paillassa, "Impact of SDN on Mobility Management," in *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Mar. 2016, pp. 260–265. DOI: 10.1109/AINA.2016.57.

[57] C. Chen, Y. Lin, L. Yen, M. Chan, and C. Tseng, "Mobility management for low-latency handover in SDN-based enterprise networks," in *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–6. DOI: 10.1109/WCNC.2016.7565105.

[58] S. M. Raza, P. Thorat, R. Challa, H. Choo, and D. S. Kim, "SDN based inter-domain mobility for PMIPv6 with route optimization," in *2016 IEEE NetSoft Conference and Workshops (NetSoft)*, Jun. 2016, pp. 24– 27. DOI: 10.1109/NETSOFT.2016.7502436.

[59] W. F. Elsadek and M. N. Mikhail, "Inter-domain Mobility Management Using SDN for Residential/Enterprise Real Time Services," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, 2016, pp. 43– 50. DOI: 10.1109/W-FiCloud.2016.25.

[60] S. M. Raza, P. Thorat, R. Challa, and H. Choo, "On demand inter domain mobility in SDN based Proxy Mobile IPv6," in *2017 International Conference on Information Networking (ICOIN)*, Jan. 2017, pp. 194–199. DOI: 10.1109/ICOIN.2017.7899503.

[61] N. Omheni, F. Zarai, B. Sadoun, and M. S. Obaidat, "New Approach for Mobility Management in Openflow/Software-Defined Networks," in *Proceedings of 8th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, Setubal, PRT, 2018, pp. 25–33. DOI: 10.5220/0006847800250033.

[62] S. M. Raza, P. Thorat, R. Challa, S. Jeon, and H. Choo, "Design and experimental evaluation of OnDemand inter-domain mobility in SDN supported PMIPv6," *Wirel. Netw.*, vol. 26, no. 1, pp. 603–620, Jan. 2020, DOI: 10.1007/s11276-019-02169-2.

[63] E. Jeong, *IPv6 Wireless Access in Vehicular Environments (IPWAVE): Problem Statement and Use Cases draft-ietf-ipwave-vehicular-networking-20*. 2021. [Online]. Available: https://tools.ietf.org/html/draft-ietf-ipwave-vehicular-networking-20

[64] S. Cespedes, N. Lu, and X. Shen, "VIP-WAVE: On the Feasibility of IP Communications in 802.11p Vehicular Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 82–97, 2013, DOI: 10.1109/TITS.2012.2206387.

[65] European Union SESAR project, *Future Communications Infrastructure (FCI) Functional Requirements Document (FRD)*. 2018.

[66] S. Ayaz, C. Bauer, M. Ehammer, T. Gräupl, and F. Arnal, "Mobility Options in the IP-based Aeronautical Telecommunication Network," 2008.

[67] S. Ayaz et al., "Architecture of an IP-based aeronautical network," in *2009 Integrated Communications, Navigation and Surveillance Conference*, 2009, pp. 1–9. DOI: 10.1109/ICNSURV.2009.5172851.

[68] L. Zhao, J. Zhang, X. Zhang, and Q. Li, "A Mobility Management Based on Proxy MIPv6 and MPLS in Aeronautical Telecommunications Network," in *2009 First International Conference on Information Science and Engineering*, 2009, pp. 2452– 2455. DOI: 10.1109/ICISE.2009.66.

[69] K. Sun and Y. Kim, "PMIP-Based Distributed Session Mobility Management for 5G Mobile Network," in *2019 International Conference on Information Networking (ICOIN)*, 2019, pp. 453–456. DOI: 10.1109/ICOIN.2019.8718131.

[70] Á. Leiter, L. Bokor, and I. Kispál, "An Evolution of Mobile IPv6 to the Cloud," in *Proceedings of the 18th ACM Symposium on Mobility Management and Wireless Access*, New York, NY, USA, 2020, pp. 137–141. DOI: 10.1145/3416012.3424633.

[71] Á. Leiter, N. Galambosi, and L. Bokor, "An Evolution of Proxy Mobile IPv6 to the Cloud," in *Proceedings of the 19th ACM International Symposium on Mobility Management and Wireless Access*, New York, NY, USA: Association for Computing Machinery, 2021, pp. 107–115. [Online]. Available: 10.1145/3479241.3486684

[72] S. Rizvi, A. Aziz, M. T. Jilani, N. Armi, G. Muhammad, and S. H. Butt, "An investigation of energy efficiency in 5G wireless networks," in *2017 International Conference on Circuits, System and Simulation (ICCSS)*, 2017, pp. 142–145. DOI: 10.1109/CIRSYSSIM.2017.8023199.

[73] M. Fiorani, S. Tombaz, J. Martensson, B. Skubic, L. Wosinska, and P. Monti, "Modeling energy performance of C-RAN with optical transport in 5G network scenarios," *IEEEOSA J. Opt. Commun. Netw.*, vol. 8, no. 11, pp. B21–B34, 2016, DOI: 10.1364/JOCN.8.000B21.

[74] X. An, W. Kiess, J. Varga, J. Prade, H.-J. Morper, and K. Hoffmann, "SDN-based vs. software-only EPC gateways: A cost analysis," in *2016 IEEE NetSoft Conference and Workshops (NetSoft)*, 2016, pp. 146–150. DOI: 10.1109/NETSOFT.2016.7502461.

[75] W. Kiess, M. R. Sama, J. Varga, J. Prade, H.-J. Morper, and K. Hoffmann, "5G via evolved packet core slices: Costs and technology of early deployments," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–7. DOI: 10.1109/PIMRC.2017.8292691.

[76] J. Varga, A. Hilt, C. Rotter, and G. Járó, "Providing Ultra-Reliable Low Latency Services for 5G with Unattended Datacenters," in *2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, 2018, pp. 1–4. DOI: 10.1109/CSNDSP.2018.8471756.

[77] R. J. Atkinson and S. N. Bhatti, *Identifier-Locator Network Protocol (ILNP) Engineering Considerations*. IETF, 2012. [Online]. Available: http://www.ietf.org/rfc/rfc6741.txt

Ákos Leiter has graduated as a Computer Engineer MSC in Department of Networked Systems and Services (HIT), Budapest University of Technology and Economics (BME) in 2015, specialized in Computer Networks. His theses was about proposing an operator-centric, dynamic flow mobility protocol with IP in Evolved Packet Core. Currently he is a PhD candidate at the Department of Networked Systems and Services in the Multimedia Networks and Services Laboratory (MEDIANETS) and a Research Engineer at Nokia Bell Labs. His main research field is Network Function Virtualization and Software Defined Networking including Orchestration and Network Automation. His work-in-progress Phd theses is about the cloudification of Mobile IPv6 protocol family on the top of Kubernetes.

Mohamad Saleh Salah has BSs and MSc in telecommunication engineering. He received his M.Sc. degree from the Department of Networked Systems and Services, Budapest University of Technology and Economics with excellence with the highest honor grade in 2021. He has been actively working in the telecom domain since 2016 focusing on network transformation technologies with the main interest in telecom cloud, cloud-native, virtualization, NFV/SDN, and mobile networks. Mohamad worked in leading ICT companies like SAP and Nokia Bell Labs and covered many engineering positions such as planning, operations, and researching; right now he is a technical leader at ng-voice company driving cloud-native IMS designing and architecting.

László Pap graduated from the Technical University of Budapest, Faculty of Electrical Engineering, Branch of Telecommunications in 1967. He became Dr. Univ. and Ph.D. in 1980, and Doctor of Sciences in 1992. In 2001 and 2007 he has been elected as a Correspondent and Full Member of the Hungarian Academy of Sciences. His main fields of the research are the electronic systems, nonlinear circuits, synchronization systems, modulation and coding, spread spectrum systems, CDMA, multiuser detection and mobile communication systems. His main education activity has covered the fields of electronics, modern modulation and coding systems, communication theory, introduction to mobile communication. Professor Pap had been Head of the Dept. of Telecommunications, the Dean of the Faculty of Electrical Engineering at Budapest University of Technology and Economics, and Vice Rector of the University.

László Bokor received the M.Sc. degree in computer engineering from the Department of Telecommunications, Budapest University of Technology and Economics (BME) in 2004, later the M.Sc.+ degree in bank informatics from the Faculty of Economic and Social Sciences, BME, and the Ph.D. degree from the Doctoral School of Informatics, BME. He has researched in multiple EU-funded and national research and development projects for several years. He is currently with the Department of Networked Systems and Services (HIT) as an Associate Professor. He leads the Commsignia-BME HIT Automotive Communications Research Group at BME and the Vehicle Communication Working Group of the Mobility Platform at KTI Institute for Transport Sciences. He is a member of the HTE, the Hungarian Standards Institution's T echnical Committee for ITS (MSZT/MB 911), the TPEGoverC-ITS Task Force within the TPEG Application Working Group of TISA, the ITS Hungary Association, and the BME's MediaNets Laboratory, where he currently participates in different R&D projects. In recognition of his professional work and achievements in mobile telecommunications, he received the HTE Silver Medal (2013), the HTE Pollák-Virág Award (2015, 2022), and the HTE Gold Medal (2018). He was a recipient of the UNKP-16-4-I. Post-Doctoral Fellowship in 2016 from the New National Excellence Program of the Ministry of Human Capacities of Hungary. In 2018 he was awarded the Dean's Honor (BME VIK) for education and research achievements in the field of communication of autonomous vehicles; in 2020, he received the BME HIT Excellence in Education Award.

TABLE III
SUMMARY OF THE SURVEYED PAPERS WITH KEY FUNCTIONAL INDICATORS

#	The approach	Publication date	Main objective	Main distinguisher	Does it use PMIPv6 signaling?	Tunneling	Protocol	Other comments
1	Seong-Mun Kim et. al [41]	2014 Jan	Integration PMIPv6 to Openflow and avoid tunneling	Control and data planes separation and eliminate IP tunneling	Yes	No	Openflow	Name: OPMIPv6
2	Kyoung-Hee Lee [42]	2014 Jan	Keeping PMIPv6 signaling while moving it into the SDN era	Routeflow as a controller	Yes	No	RouteFlow and Openflow	Control plane elements are VMs
3	Syed M. Raza et al. [44][45]	2014 Jan	Integration PMIPv6 to Openflow to reduce handover latency	Keeping full backward compatibility on PMIPv6	Yes	Yes	Openflow	Name: OF-PMIPv6
4	Hao Yu [46]	2014 May	Solution for inter-domain heterogeneous mobility with SDN and PMIPv6	Heterogenous radio access networks are examined, including EPC	Yes	Yes	It is not mentioned directly.	Vertical handovers
5	You Wang et al. [47]	2014 Aug	PMIPv6-like Mobility Management with Openflow	Binding Cache placement algorithm	No	No	Openflow	Performance compared to PMIPv6 and ILNP[77]
6	Yuta Watanabe, et al.[48]	2015 Jan	Route optimization	Offload LMA by suggesting a route to CN without passing LMA	Yes	No	Openflow	Total throughput increased
7	Sakshi Chourasia et al. .[50]	2015 Apr	New EPC Architecture without MM protocols	PMIPv6 usage at 3GPP S5/S8 interface is also covered	No	No	Openflow	GTP also considered
8	Seong-Mun Kim et al. [51]	2015 Arp	Extended evaluation for Paper #1	Control and data planes separation and eliminate IP tunneling	Yes	No	Openflow	Performance gain compared to ordinary PMIPv6
9	Abbas Bradai et al. [52]	2015 Jun	Dynamic anchor point	Load-balancing considered	No	Yes	Openflow	HMIPv6 is used too
10	Weng-Kang Jia [53]	2015 Nov	PMIPv6 function integration to EPC with SDN	Inter-domain PMIPv6 - based handover with SDN help	partially	partially	Openflow	Performance gain in registration to non-SDN PMIPv6
11	A_Aissioui et al. [54]	2015 Dec	Follow-me-cloud based on PMIPv6	Cloud computing considered	Yes	No	Openflow	Server and mobile user move too
12	Pill-Won Park et al. [55]	2016 Feb	Support mobility in SDN architecture with OpenFlow signaling and PMIPv6 concept	No PMIPv6 signaling messages	No	No	OpenFlow	Name: OMM
13	Kuljaree Tantayakul et al. [56]	2016 Mar	Eliminates PMIPv6 for mobility management in the SDN world	Existing Openflow messages are used	No	No	Openflow	Refers to Seong-Mun Kim et. al and Syed M. Raza et. al
14	Ce Chen .[57]	2016 Apr	Low-Latency Handover in SDN-Based Enterprise Networks	Using and evolving OF-PMIPv6 concepts for IPv4	No	No	Openflow	Name: M-SDN
15	Syed M. Raza et al. [58]	2016 Jun	Improve the PMIPv6 based on OpenFlow to serve inter-domain IP mobility	Inter-domain handover	Yes	Yes	OpenFlow	The architecture of Paper #3 evolved
16	Walaa F. Elsadek et al. [59]	2016 Aug	Mitigating inter-domain handover with an SDN Mobility Framework	Focus on offloading core network	No	No	Openflow	SIPTO/LIPA considered
17	Syed M. Raza et al. [60]	2017 Jan	On-demand way of working is concerned	Inter-domain handover with multiple prefix assignment for the MN	Yes	Yes	OpenFlow	Extension of Paper #15 ( ) IDS-PMIP
18	N. Omheni et al. [61]	2018 Jan	Partially distributed way of working	MIH is used	Yes	Yes	Openflow	The number of handovers is reduced
19	Syed M. Raza et al. [62]	2019 Oct	on-demand inter-domain SDN-PMIPv6	New protocol for controllers communication Prefix retrieval mechanism	Yes	Yes	Openflow C3 between controller	Extension of Paper #15 () and #17 () Name: OIS-PMIP

Cost-Effective Delay-Constrained Optical Fronthaul Design for 5G and Beyond

Abdulhalim Fayad and Tibor Cinkler

***Abstract*—With the rapid growth of the telecom sector heading towards 5G and 6G and the emergence of high-bandwidth and time-sensitive applications, mobile network operators (MNOs) are driven to plan their networks to meet these new requirements in a cost-effective manner. The cloud radio access network (CRAN) has been presented as a promising architecture that can decrease capital expenditures (Capex) and operating expenditures (Opex) and improve network performance. The fronthaul (FH) is a part of the network that links the remote radio head (RRH) to the baseband unit (BBU); these links need high-capacity and low latency connections necessitating cost-effective implementation. On the other hand, the transport delay and FH deployment costs increase if the BBU is not placed in an appropriate location.**

In this paper, we propose an integer linear program (ILP) that simultaneously optimizes BBU and FH deployment resulting in minimal capital expenditures (Capex). Simulations are run to compare the performance of star and tree topologies with the varying line of sight probabilities (LoS) and delay thresholds. We consider fiber-optic (FO) and free-space optics (FSO) technologies as FH for the CRAN. Finally, we provide an analysis of Opex and the total costs of ownership (TCO), i.e., a techno-economic analysis.

***Index Terms*—5G and Beyond, BBU, Fronthaul, delay, optimization, Capex, Opex, TCO.**

I. INTRODUCTION

IN conjunction with the advent of applications that have high bandwidth-demanding and strict latency requirements such as e-health, online video gaming, security applications, smart farming, and connected cars, mobile traffic will exceed 5000 EB/month by 2030 [1]. As a consequence, there will be an inevitable overload on telecommunications networks, which brings many challenges to MNOs. For that, the fifth-generation (5G) of mobile networks pledges to deliver higher data rates, ultra-low latencies, more reliability, and increased availability for a large number of users [2]. However, as mobile traffic grows, and the critical mission applications emerge rapidly, 5G will eventually run into technical difficulties enabling vast interconnection with highly diversified and demanding service and computing requirements. To cope with this issue, the attention of academia and industry lately turned towards the research of the sixth-generation (6G) of mobile communications [3]. 6G mobile communication networks are predicted to deliver extreme peak data rates, ultra-low latencies, network availability, and reliability about 99.99999%, an exceptionally high connection density of over 10^7 devices/km^2, and 6G spectrum efficiency will be more than 5x of the 5G [4]. We use 5G and Beyond to express 5G and 6G cellular communications technologies. To meet the 5G and beyond goals, mobile network operators (MNOs) ought to improve the performance of their networks. Many solutions have been presented to address this issue, for instance, using additional spectrum, deploying additional sites (Base stations or small cells) [5]–[7], and by using massive multi input multi output (MIMO) access technology [8], [9]. The most common approach to achieve high throughput is to densify cell coverage by deploying additional Base Stations (BSs) [10]. This increases capital expenditure (Capex) and operational expenditure (Opex) while the revenue is not high enough [5]. As a result, researchers have developed cost-effective strategies to transform standard BS design into a Cloud Radio Access Network (CRAN) that can handle a massive capacity of cell sites [11], [12]. For more comprehensive information about C-RAN architecture, the reader is referred to [11], and [13]. In CRAN architecture, as shown in Figure 1, the processing operations are fulfilled at the baseband unit (BBU), which is allocated in a central location. In contrast, the remote radio heads (RRHs) are positioned at the antenna side with a relatively restricted range of responsibilities. To transmit the baseband signals between the RRHs and the BBU, a low-latency and high capacity FH, is needed. Although CRAN architecture can reduce both Opex and Capex, the cost of the fronthaul remains a barrier. Many technologies have been proposed for 5G and Beyond fronthaul, such as microwave, fiber optics (FO), and free-space optics (FSO) [14]. Microwave technology is considered an excellent candidate to link BSs and the core network, but the increasing number of bandwidth-intensive applications necessitates the use of technologies like FSO, which provide substantially better throughput [15] [16].FO is hailed as a vital enabler for the fronthaul of 5G and Beyond, as it can offer a large capacity and is not affected by the weather or interference, but this technology has many drawbacks. The main disadvantage is the high cost, particularly where trenching is required, as well as the time delays. On the other hand, FSO becomes a viable option for fronthaul since 5G and Beyond require a high number of cell sites and the distance between them can be hundreds of meters rather than miles. FSO has several advantages, such as being cost-effective in terms of deployment cost, no electromagnetic interference, easy to install, and an unlicensed frequency range. Nevertheless, FSO has the limitation of requiring line of sight (LoS); as well as it does not work successfully in bad

Abdulhalim Fayad and Tibor Cinkler are with Department Of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Hungary. (e-mail: Fayad, Cinkler@tmit.bme.hu)

DOI: 10.36244/ICJ.2022.2.2

weather conditions [17], [18]. This work provides a framework based on integer linear programming (ILP) to simultaneously plan and optimize BBU and FH deployment costs, satisfying delay and capacity constraints for 5G and Beyond networks. The proposed framework can be applicable for green-field scenarios and brown-field scenarios where we can benefit from the existing infrastructure, resulting in optimal Capex. The considered technologies for FH are FO and FSO. Based on the optimal Capex, we analyze the Opex and total costs of ownership (TCO) of the network.

This paper is structured as follows. Section II reviews the related works and literature studies. Section III describes the studied problem. The problem is formulated as an ILP problem in Section IV. Section V analyzes TCO and Opex. Section VI shows a case study and numerical results. Finally, VII concludes the paper.

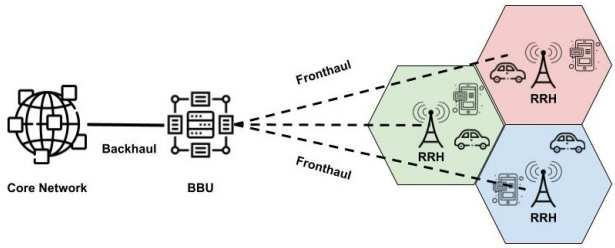


Fig. 1. Cloud Radio Access Network(C-RAN) Architecture

II. Related Works and Literature study

Planning and optimizing 5G and beyond networks brings many challenges for the MNOs, where there is a need to find the optimal placement of radio components and optimal deployment of the transport infrastructure. These challenges that should be considered are the optimal number of BBUs with optimal placement, optimal fronthaul design, and the costs of the network. To meet these challenges, in a brownfield scenario, Marotta et al. [19] propose an ILP model to evaluate the optimal deployment of CRAN fronthaul deployment for 5G using Optical fiber and microwave links. Tonini et al. in [20] present a C-RAN architecture with a hybrid fronthaul solution using FO and FSO, as well as two ILP models for optimization of the number of deployed remote radio heads (RRHs) and the cost of the FH deployment using (hybrid FO/FSO) in a greenfield and brownfield scenarios without considering BBU placement or delay issue. Klinkowski et al. in [21] examine the scalability of an ILP model for 5GCRAN deployment. They analyze two deployment options for RRHs, BBUs, and optical fronthaul in order to reduce deployment costs. Ranaweera et al. [22] suggest an integer linear program (ILP) model that optimizes small cell and fiber backhaul deployments in a greenfield scenario while meeting network capacity requirements. Dahrouj et al. [23] provide a low-cost hybrid RF/FSO backhaul solution in which base stations are linked by optical fibers or hybrid RF/FSO backhaul links. The authors address the problem of minimizing the cost of backhaul planning under reliability, connectivity, and data rate constraints. Simulations show that the suggested solution delivers a cost-effective backhaul deployment plan that is reliable, high-data-rate, and robust. Li et al. [24] provide an integer linear programming (ILP) model for optimizing FSO backhaul design while ensuring K-disjoint pathways between each node pair. Their findings demonstrate that FSO is a cost-effective option in large-scale applications, highlighting the trade-off between dependability and network costs. Jaffer et al. in [25] propose a hybrid FH architecture for 5G CRAN deployment that employs Passive Optical Network (PON) and Free Space Optics (FSO) in order to maximize flexibility and minimize FH network costs. They investigate the TCO of FH networks, taking into account both Capex and Opex; the results reveal that a hybrid PON-FSO fronthaul may save up to 42.89 $\%$ of TCO. Ranaweera et al. in [26] propose a generalized optimization framework to minimize the costs of 5G Fixed Wireless Access (FWA) and its optical x-haul network. The proposed ILP jointly optimizes the wireless and the optical x-haul segment of a 5G FWA. It can meet essential requirements of the FWA network, such as fixed user coverage and capacity. Regardless of the architecture of the radio access network , Ranaweera et al. [27] present a generic framework for creating a cost-optimal transport network and 5G and beyond mobile networks while addressing network and user demands. As a transport medium, the authors consider fiber-optic technology offered by passive optical networks. Mahloo et al. [28] provide a comprehensive cost evaluation methodology to compute the TCO of mobile backhaul networks considering microwave and fiber technologies. However, most of the existing studies do not consider the optimization of the FH and the BBU deployment of the 5G and Beyond under different delay thresholds, and various LoS probabilities. As well as, they do not analyze how can the delay threshold affects the costs of the network that can help the MNOs to plan their networks to be ready for upcoming time-sensitive services.

With this in mind, in this study, we propose an ILP that simultaneously optimizes the BBU and FH deployment cost for 5G and Beyond. The outcome is, minimizing the Capex of the network considering different delay values and different LoS probabilities for $tree$ and $star$ topologies considering FO and FSO technologies for the FH. We also provide analysis for Opex and TCO of the network, i.e., a techno-economic analysis.

III. Problem description:

In essence, the problem discussed in this study is finding optimal BBU placement and the optimal cost of FH deployment based on FO and FSO for 5G and Beyond. Figure 2 shows an example of the deployment scenario. We consider that only one operator serves this area, and there is no need for infrastructure sharing or leasing fiber. The problem can be divided into sub-problems as follows:

BBU placement: all RRHs are organized into groups accord-

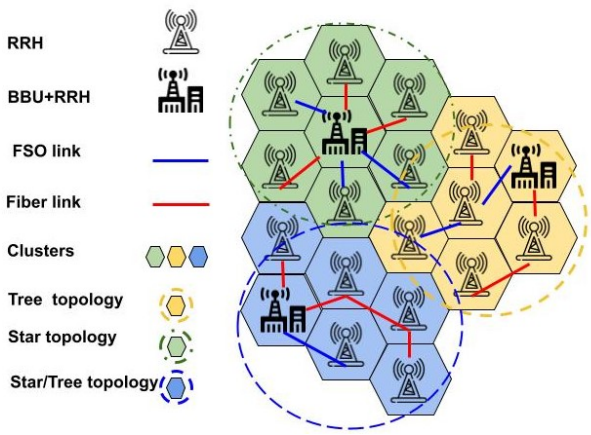


Fig. 2. Deployment scenario

ing to the shortest distance and minimum delay. Each group of these RRHs is assigned to one BBU; thus, it is considered a clustering problem. Finding the best location for the BBU is as follows: it must be somewhere close to the center of the group to limit the total length of the link between the RRHs and the BBUs inside the group. We consider a subset of the RRH sites as the set of possible BBU locations. This problem can be regarded as NP-complete one [29].

Fronthaul deployment: All RRHs should be connected to respective BBUs through the shortest pathways possible, and an optimal connection deployment is required. Two technologies are considered FO and FSO. Where when there is LoS, we use FSO link; else, we use FO link. The optimal deployment of FH must satisfy delay and capacity constraints. Additionally, we consider two different topologies to deploy the fronthaul. First is the $star$ topology, where there is a direct link from each RRH toward its serving BBU. Second is the $tree$ topology, where we take into account cascading links between RRHs, resulting in the same links, can carry more data of more than one RRH toward the serving BBU to minimize the deployment costs. The tree topology of RRHs and their serving BBUs with obeying the delay constraints can be modeled as Rooted Delay Constrained Steiner Tree [30], which is an NP-hard problem. The purpose of the optimization process was to reduce Capex, or the entire cost of installation of the network, which can be broken down into two parts: equipment expenses and transport network costs [31]. Therefore the optimization problem addresses the following open questions:

1) How to form groups of RRHs connected to the same BBU?
2) How many BBUs should be installed to serve all RRHs with minimum costs?
3) How can the optimum BBU placement and the minimum be determined while meeting all the network requirements (delay and capacity)?

IV. ILP FORMULATION:

This section elucidates our proposed optimization framework based on ILP [32], that minimizes the FH deployment cost while guaranteeing other deployment requirements, such as delay and network capacity. The proposed framework outputs the optimal locations of BBUs, the minimum number of BBUs, and optimal FO and FSO links to deploy the cost-effective FH for 5G and beyond networks. The key components of the proposed framework are depicted in Figure 3. The objective is to find the optimal total cost of fronthaul deployment and network equipment. The framework also includes a variety of constraints to meet network needs, such as latency, capacity, the maximum allowable distance between the RRH and the BBU, the number of RRHs, the maximum number of connections per one BBU, and financial constraints.

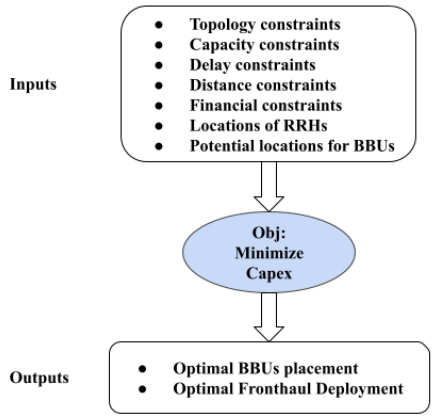


Fig. 3. Optimization framework

A. Network Parameters

The framework is built on parameters that the user may adjust to suit the deployment situation under consideration. Table I contains a list of these parameters. Based

B. Decision variables

1) Binary variable F_i
$$F_i=\begin{cases} 1 \text{ if a possible BBU } i \text{ is installed} \\ 0 \text{ otherwise} \end{cases}$$
2) Binary variable R_i
$$R_i=\begin{cases} 1 \text{ if RRH is chosen for site } i \\ 0 \text{ otherwise} \end{cases}$$
3) Binary variable λ_{ij}
$$\lambda_{ij}=\begin{cases} 1 \text{ if RRH } j \text{ is served by BBU } i \\ 0 \text{ otherwise} \end{cases}$$
4) Binary variable η_{ij}^{br}
$$\eta_{ij}^{br} = \begin{cases} 1 \text{ if link } (ij) \text{ is used to connect BBU } b \\ \text{ to RRH } r \\ 0 \text{ otherwise} \end{cases}$$

TABLE I
NETWORK PARAMETERS

Notation	Description
N	Set of RRHs locations.
M	Set of BBU locations.
$n_{\max}$	The maximum number of RRHs that can be connected to the same BBU.
C_m	BBU cost.
C_n	RRH cost.
C_{FOp}	Fiber optic cable purchasing cost per meter.
C_{FOi}	Fiber optic cable trenching cost per meter.
C_{FSOp}	Free space optic link purchasing cost.
C_{FSOi}	Free space optic link installation cost.
C_{ij}	The cost of link (ij) between point i and point j.
l_{ij}^{FO} [m]	The distance from point i to point j for FO link.
l_{ij}^{FSO}[m]	The distance from point i to point j for FSO link.
$l_{\max}$[m]	Maximum allowed transmission distance between each pair (BBU-RRH) based on the value of $\tau_{\max}$.
$T_{ij}^{FO}[\mu s]$	Delay over the link from point i to j using FO link.
$T_{ij}^{FSO}[\mu s]$	Delay over the link (ij) from point i to point j using FSO link.
$\tau_{\max}[\mu s]$	Maximum allowed fronthaul propagation between the RRH and its BBU.
θ_{ij}^{FO}[Gb/s]	Capacity of fiber optic link (ij).
θ_{ij}^{FSO}[Gb/s]	Capacity of FSO link (ij).
θ_{ij}[Gb/s]	Capacity of link(ij) between point i and point j using FO or FSO.
θ_{ij}^{A}[Gb/s]	Available capacity for link (ij).
θ_{ij}^{R}[Gb/s]	The required capacity for link (ij).
θ_r[Gb/s]	RRH capacity.
θ_b[Gb/s]	BBU capacity.
ϵ_{ij}	l, if there is line of sight from RRH i to RRH j. 0, otherwise.
x_{ij}	l, f there is already link deployed from RRH i to RRH j. 0, otherwise.
η_{ij}^{brFO}	1, if we use FO for link (ij) to connect RRH r to BBU b. 0, otherwise.
η_{ij}^{brFSO}	1, if we use FSO for link (ij) to connect RRH r to BBU b. 0, otherwise.

5) Binary variable η_{ij}^{FO}

$$\eta_{ij}^{FO} = \begin{cases} 1 \text{ if we use fiber to link RRH } i \text{ to RRH } j \\ 0 \text{ otherwise} \end{cases}$$

6) Binary variable η_{ij}^{FSO}

$$\eta_{ij}^{FSO} = \begin{cases} 1 \text{ if we use FSO to link RRH } i \text{ to } j \\ 0 \text{ otherwise} \end{cases}$$

C. Objective function

$$\min \underbrace{C_m \sum_{i=1}^{M} F_i}_{\text{BBU cost}} + \underbrace{C_n \sum_{i=1}^{N} R_i}_{\text{RRHs cost}} + \underbrace{\sum_{j=1}^{N_R}\sum_{i=1}^{N_B}(\eta_{ij}^{FO} l_{ij}^{FO}(C_{FOi}+C_{FOp}) + \eta_{ij}^{FSO}(C_{FSOi}+C_{FSOp}))}_{\text{Fronthaul deployment cost}}$$

D. Constraints

1) Topology constraints

a) Each RRH should be served by only one BBU

$$\sum_{i\in M} \lambda_{ij} = 1 \quad \forall j \in N \tag{1}$$

b) The BBU i that serves the RRH j must be selected

$$\lambda_{ij} \le F_i \quad \forall i \in M, j \in N \tag{2}$$

c) For tree topology, we assume that each connection may transport data from more than one RRH to the BBU, and that each link is associated with just one BBU

$$\sum_{b=1}^{M} \eta_{ij}^{br} = 1 \quad \forall i,j,r \in N \tag{3}$$

d) To guarantee that the flow from RRH r to BBU b is equal for each pair BBU-RRH (b, r), and to ensure the incoming flow equals to the outgoing flow at each intermediate node along the path, we appoint

$$\sum_{i\in N} \eta_{ij}^{br} - \sum_{i\in N} \eta_{ji}^{br} = \alpha \quad \forall r \in N, b \in M \tag{4}$$

$$\alpha = \begin{cases} 1 & \text{if} \quad j = b \\ -1 & \text{if} \quad j = r \\ 0, & \text{if} \quad j \ne b, j \ne r \end{cases} \tag{5}$$

e) Determining the maximum number of RRHs that can be served by one BBU:

$$\sum_{i\in M} \lambda_{ij} R_i \le n_{\max} F_i \quad \forall j \in N \tag{6}$$

2) Capacity constraint

a) Available capacity over link(i, j)

$$\theta_{ij}^{A} = \theta_{ij}^{FO}\eta_{ij}^{FO} + \theta_{ij}^{FSO}\eta_{ij}^{FSO} \tag{7}$$

b) The requested capacity over link (ij) can be calculated as follows:

$$\theta_{ij}^{R} = \sum_{b\in M}\sum_{r\in N} \theta_r \eta_{ij}^{br} \quad \forall i,j \in N \tag{8}$$

c) The requested capacity should be less or equal than the available capacity over link (i, j)

$$\theta_{ij}^{R} \le \theta_{ij}^{A} \quad \forall i \in M, j \in N \tag{9}$$

d) The maximum requested capacity from a group of RRHs connected to the same BBU should be less or equal then BBU capacity

$$\sum_{i\in M} \lambda_{ij}\theta_{ij}^{R} \le \theta_b F_i \quad \forall j \in N \tag{10}$$

3) Delay constraints

a) The total delay between each (BBU-RRH) pair should be equal or lower than the allowed delay

$$\sum_{i\in N}\sum_{j\in M} (T_{ij}^{FO}\eta_{ij}^{FO} + T_{ij}^{FSO}\eta_{ij}^{FSO}) \le \tau_{\max} \tag{11}$$

The delay constraints can be expressed as distance constraints, as follows:

$$\sum_{i\in N}\sum_{j\in M}(l_{ij}^{FO}\eta_{ij}^{brFO}+l_{ij}^{FSO}\eta_{ij}^{brFSO}) \leq l_{\max}, \forall b \in M, r \in N \tag{12}$$

4) Financial constraints

a) Calculation of link (i, j) cost

$$C_{ij} = C_{ij}^{FO}\eta_{ij}^{FO} + C_{ij}^{FSO}\eta_{ij}^{FSO} \tag{13}$$

b) If there is already deployed link (i, j) then there is no cost:

$$\text{if} \quad x_{ij} = 1 \quad \text{then} \quad C_{ij} = 0 \tag{14}$$

E. The Fronthaul capacity and delay analysis:

In the 5G CRAN architecture, the BS is split into BBU and RRH, with a fronthaul link between them. This architecture brings several benefits in terms of costs (reduction of energy consumption, and operational and maintenance costs) [12]. The FH link has to overcome two main challenges as follows:

1) High capacity: The fronthaul link must carry a very high bit rate, in the order of units to tens of Gb/s. For instance, sector configured as 64×64 MIMO with 20 MHz bandwidth requires about 64 Gb/s for fronthaul, and RRH with three sectors it requires 192 Gb/s [33].

2) Low latency connection: The connection over the fronthaul must guarantee that the low latency requirement between the RRH and its BBU (less or equal $100\mu s$). Figure 4 illustrates the one-way delay over a one-hop link RRH-BBU. For simplicity, the total one-way delay from one RRH to one BBU includes BBU processing delay T_B, the switching delay T_{sw} (neglected delay), the one way propagation delay T_F, and the RRH processing delay T_R, as shown in Equation 15.

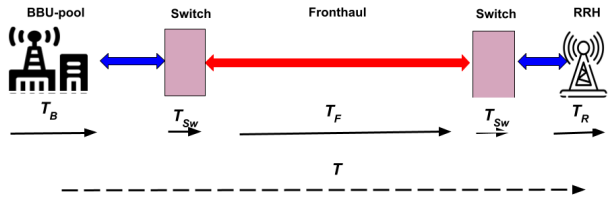


Fig. 4. Fronthaul delay analysis

$$T = T_B + 2T_{Sw} + T_F + T_R \tag{15}$$

In case of Multi-hop delay in the tree topology, the delay equation will be as follows:

$$T = T_B + m \cdot T_{Sw} + n \cdot T_F + n \cdot T_R \tag{16}$$

Where n is the number of RRHs, and m is the number of used switches (for one-hop link m=2, and n=1). All types of delay mentioned in Equation 15 have fixed values as they belong to the network devices, and the only variable delay is the propagation delay. In this study, we primarily examine the one-way propagation delay as the main constrain in order to prepare our network for ultra-low latency and time-sensitive applications [34], as it has a direct influence on the distance from BBU to RRH. The higher the delay, the longer the distance, and vice-versa. The propagation delay value should be equal to or less than $50\mu s$ [35]. The one-way propagation delay is affected by the physical medium used to implement the link, as the speed of light differs from one physical medium to another where the speed of light in FO equals 2.10^8 m/s. In contrast, in FSO (Air) equals $\sim 3.10^8$ m/s. The delay from point i to point j can be given as follows:

$$T_{ij}[\mu s] = \frac{d_{ij}[\text{m}]}{v[\text{m/s}]} \cdot 10^6 \tag{17}$$

Where *T* is the one-way propagation delay between *i* and *j*, *d* is the distance from i to j, and *v* is speed of light in fiber or air. Based on Equation 17 we can calculate the maximum allowed distance from *i* to *j* depending on the maximum allowed delay as follows:

$$d_{ij}[m] = \frac{T_{ij}[\mu s] \cdot v[\text{m/s}]}{10^6} \tag{18}$$

Fiber links are often installed in cable ducts designed to be readily maintained (e.g., along streets), whereas FSO links are established based on line-of-sight propagation. The walking path distance is used for fiber links, while for FSO links, a straight line is used to compute the link length. In this study, we consider the propagation delay threshold over the fronthaul from 1μs to 15μs. As a result, the maximum allowed distance for fiber link is 3000 m, and for FSO link is 4500 m. FSO connections are often utilized for transmission distances ranging from 300 m to 5 km. Moreover, they may also be installed for greater distances ranging from 8-11 km, depending on the speed and required availability [36]. From our point of view, holey fiber can be considered as an excellent alternative for single-mode fiber to tackle the delay issue where holey fiber has a distribution of air holes in the cladding that runs along the length of the fiber [37].

V. TOTAL COSTS OF OWNERSHIP (TCO)

This section presents a TCO model covering both Capex and Opex aspects. Figure 5 presents a cost classification according to the proposed cost model. Wherein [28] they provide an excellent cost modeling of backhaul for mobile networks, their proposed model accounts for both fiber optics and microwave. In addition to these results, we consider the FSO as well. Furthermore, we consider a hybrid FO/FSO solution for 5G and beyond fronthaul. TCO can be calculated as follows:

$$TCO = Capex + N_y \cdot Opex \tag{19}$$

Where N_y is the number of operation years.

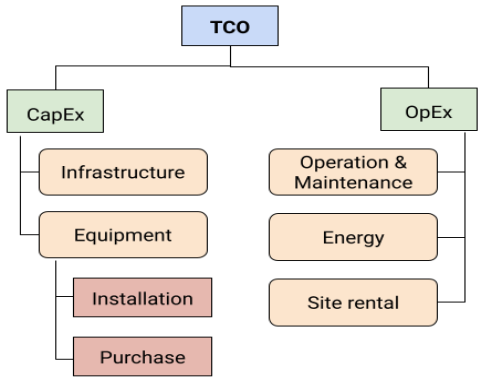


Fig. 5. Total Costs of Ownership (TCO) summary

A. Opex analysis:

Opex stands for operational expenditure, which refers to the ongoing costs of daily operating of the network, which consists of three different costs, energy consumption costs (C_E), operation and maintenance costs ($C_{O\&M}$), and site rental (C_{Sr}) [38]. The Opex calculation is given as follows:

$$Opex = C_E + C_{O\&M} + C_{Sr} \tag{20}$$

1) Energy consumption model: The energy consumption cost was calculated by summing the consumption costs of all electrical equipment in the various locations at the BBU, RRH, FSO links, and fiber links. The following equation defines the energy consumption model:

$$C_E = \sum_{i\in M}(C_{EB} + C_{Ecool}) + \sum_{i\in N}(C_{ER}) + \sum_{i\in nf}(C_{EFO}) + \sum_{i\in nFSO} C_{EFSO} \tag{21}$$

Where, C_{EB}, C_{Ecool}, C_{ER}, C_{EFO}, and C_{EFSO} represents the energy consumption in the BBU, cooling, the RRH, the FO link, and the FSO link respectively. M, N, nFO, and $nFSO$ are the number of BBUs, RRHs, FO links and FSO links. Where the energy consumption for any element over the years can be calculated as follows:

$$C_E = N_y \cdot E_c \cdot p_y \tag{22}$$

Where, N_y, E_c, and p_y, represent, number of operation years, energy cost per kW, and yearly energy consumption. Furthermore, P_y can be found based on [39] as follows:

$$P_y[Wh] = \frac{P_{eq} \cdot 24 \cdot 365}{1000} \tag{23}$$

where,P_{eq} denotes the energy consumed per the element per hour.

2) Operation and maintenance costs: Keeping the network up and operating requires a regular maintenance schedule. This includes equipment monitoring and testing, software updates (including license renewals as needed), and the replacement of supporting components such as batteries [28]:

$$C_{O\&M} = \sum_{i\in N}(C_{MB}) + \sum_{i\in M}(C_{MR}) + \sum_{i\in nf}(C_{MF}) + \sum_{i\in nFSO} C_{MFSO} \tag{24}$$

Where, C_{MB}, C_{MR}, C_{MF}, and C_{MFSO} represent the operation and maintenance costs for the BBU, the RRH, the fiber link, and the free space optic link, respectively. Based on [28] the annual operation and maintenance costs are equal to 10% of CapEx.

3) Cell site rental cost: means the yearly fees paid by mobile network operators to rent space for their equipment [28], which can be simply calculated as follows:

$$C_{Sr} = N \cdot Sr_y \tag{25}$$

where, N donates the number of cell sites (RRHs). While Sr_y, and yearly costs for renting one cell site.

VI. Case Study and Numerical Results

This section presents the numerical results when scattering 18 RRHs uniformly in a hexagonal area with a radius of 1.3 km as an urban geographical scenario. We used the commercially available ILOG CPLEX solver [40] on a computer with Intel i5 processors and 8 GB of RAM to solve our optimization framework. We assume that the RRH coverage radius equals 300 m [35]. In this paper, We consider that all RRHs locations can be potential locations for BBUs, where all cell sites are rooftop type. We also assume that each RRH has three sectors configured as 2×2 MIMO with 20 MHz bandwidth and a capacity of 7.5 Gb/s, and the maximum number of RRHs that one BBU can serve is 10 RRHs. We compare the optimal deployment costs for various thresholds of one-way propagation delay from 1 μs to 15 μs and for different line of sight (LoS) probabilities (0%, 50%, 100%). 0% means that FO is always used to establish connections from RRHs to BBUs. 50% means that half of the links in the studied area have LoS toward BBUs (chosen randomly). 100% LoS allows FSO to be used as FH. We evaluate our results considering $tree$ and $star$ topologies. We consider two technologies, i.e., FO with capacity of to 1 Tb/s and FSO links with 100 Gb/s capacity. Table II contains the input parameters for this work.

1) Number of used BBUs versus delay: Figure 6 clarifies the relation between the fronthaul propagation delay (T_f) and the minimum number of BBUs that are needed to serve the deployed number of RRHs. Where the higher the allowed delay, the lower the number of needed BBUs, and vice versa. For 0% LoS probability, the minimum number of BBUs is two, constant from 5 μs and higher. For 50% LoS probability, the minimum number of BBUs is three, stable from 9 μs and higher. For 100% LoS probability, the minimum number of BBUs is two, stable from 7 μs and higher. We can summarize

TABLE II
CAPEX AND OPEX OF THE CASE STUDY [12], [39], [41], [42]

Parameter	Cost[€]
C_m	75000
C_{FOp}	0.08
C_{FOi}	45
C_{FSOp}	10000
C_{FSOi}	200
C_n (RRH with 3 Antennas)	30000
C_{Sr}	8000
E_c	0.1367
$C_{O\&M}$	10% of Capex
Component	**Energy consumption [Wh]**
C_{EB}	200
C_{Ecool}	500
C_{ER}	100
C_{EFSO}	100
C_{EFO}	10

that the higher the LoS probabilities, the higher the number of used FSO links, which reduces the number of needed BBUs compared to lower LoS probabilities where we have to use FO links. Since FSO links can serve almost 1.5 times longer distances than FO links for the same delay values, one BBU can serve a larger area.

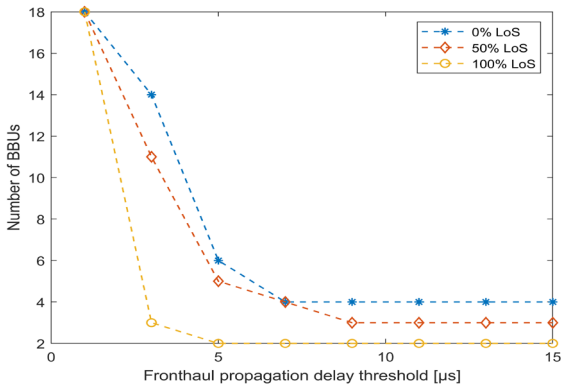


Fig. 6. Number of BBUs vs. allowed propagation delay threshold

2) Capex versus delay: Figure 7 illustrates the relation between the delay and Capex for tree and star topologies. For 0% LoS probability, Capex can decrease significantly with an increasing allowed delay. For a delay value of 15 μs, tree topology can reduce Capex by 20% compared to the star topology. Tree topology can be more cost-effective than star topology but less reliable when a fiber cut accident happens, leading to all linked RRHs being out of service. For 50%, and 100% LoS probabilities, both tree and star topologies need the exact value of Capex.

3) Opex versus delay: Figure 8 shows the Opex changes based on delay values. We notice that there is an inverse proportion between the value of the Opex and the value of the delay. FSO links consume more energy than fiber links,

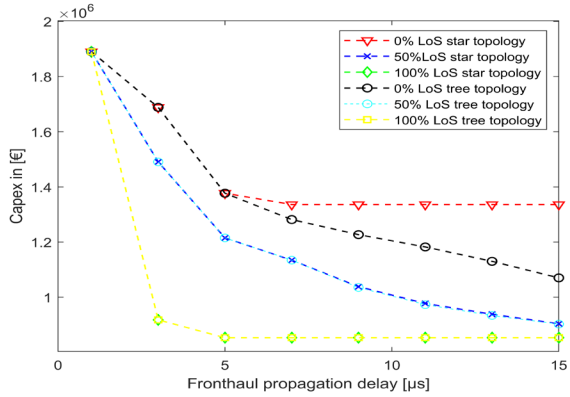


Fig. 7. Capex vs. allowed propagation delay threshold

resulting in FO being more energy-efficient than FSO [42]. Using FSO links can reduce the operation and maintenance costs due to the high costs caused by civil works in terms of fiber cuts. Also, using FSO can reduce the number of used BBUs. Furthermore, FSO can serve longer distances than FO for the same delay values. Figure 9 shows the changes of Opex over ten years, where we can conclude that Opex will become more dominant than Capex after five years of operation, while for one year of operation. Opex approximately equals 21% of Capex.

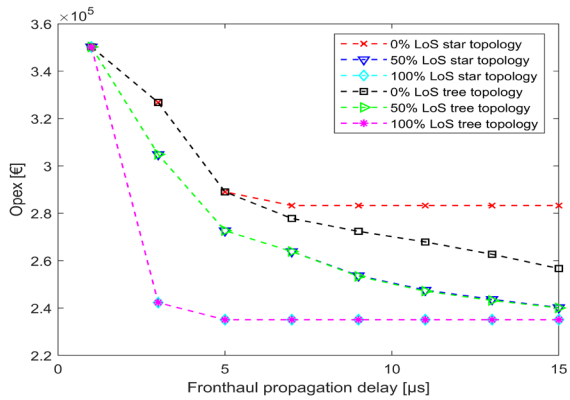


Fig. 8. Opex vs. allowed propagation delay threshold

4) Total cost of ownership (TCO) vs. propagation delay thresholds: The total cost of ownership for the case study for different delay values is shown in Figure 10. It can be observed that the minimal cost can be obtained in the case of 0% LoS probability when the FH is only FO links for either tree or star topologies. On the other hand, in the case of 50% LoS probability when the fronthaul is hybrid FO/FSO links for either tree or star topologies, we observe that we need approximately the exact cost for tree and star topologies. Once again, for 100% LoS probability, when we can use FSO links for the fronthaul, it is clear that tree topology is more cost-effective than star topology, especially for higher delay

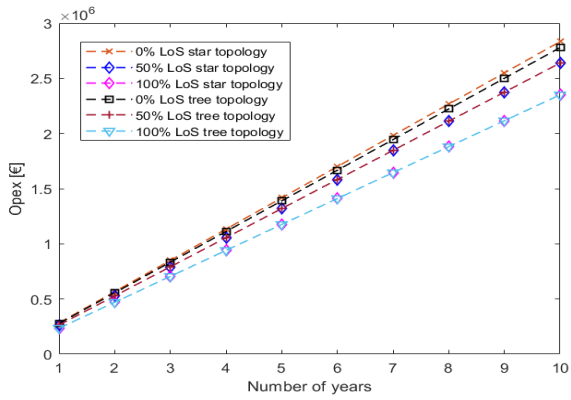


Fig. 9. Opex over 10 years

values. Finally, from the linear relation between Figure 8 and Figure 10, we find that Opex constitutes 15.6% of TCO.

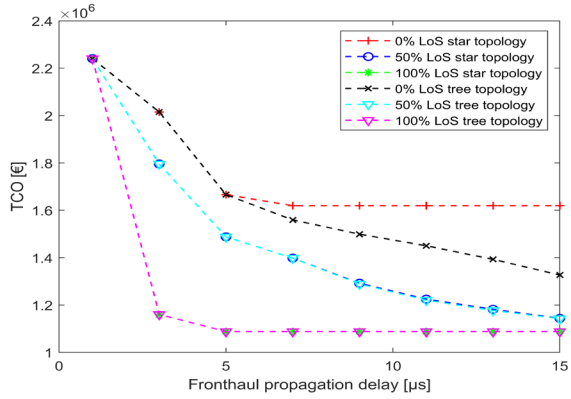


Fig. 10. TCO Vs. allowed propagation delay threshold

VII. Conclusion

This paper proposes an ILP based optimization framework for greenfield and brownfield deployment scenarios that jointly optimize the BBU placement and optical fronthaul (FH) deployment for 5G and Beyond networks under delay constraints. We considered fiber optic (FO) and free space optic (FSO) technologies as primary solutions to meet the FH challenges in terms of delay and capacity requirements. The outputs of the proposed framework are to find the optimal number and placement of BBUs and the optimal deployment of the FH, resulting in minimum Capex. We compare our results for *star* and *tree* topologies for various line of sight (LoS) probabilities. We can conclude from the above framework that FSO can be more cost-efficient than FO, especially for dense deployment scenarios. However, the signal quality is worst due to it is more sensitive to the weather conditions than FO. In the case of using FO, we should use the tree topology to link RRHs clusters to their BBUs as it is more cost-efficient than the star topology. However, it has somewhat lower availability, as if there is a fiber cut between the first RRH of the tree and the BBU, all of the tree will be out of service. Based on the optimal Capex, we provide Opex and TCO analysis. Opex constitutes about 15.6% of TCO. In future work, we plan to develop the ILP framework to optimize the total cost of ownership with additional constraints such as energy efficiency and path protection. This results in cost-effective, ultra-reliable low latency communications (URLLC) for time-sensitive and critical applications. We will also consider other technologies such as massive MIMO and millimeter waves (mmwaves).

Acknowledgement

This work was supported by the CHIST-ERA grant SAMBAS (CHIST-ERA-20-SICT-003) funded by FWO, ANR, NKFIH and UKRI.

References

[1] ITU-R M.2370-0, IMT traffic estimates for the years 2020 to 2030, Jul.2015.

[2] M. Agiwal, A. Roy and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," in IEEE Communications Surveys Tutorials, vol. 18, no. 3, pp. 1617-1655, third quarter 2016, DOI: 10.1109/COMST.2016.2532458.

[3] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," in IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021, DOI: 10.1109/OJCOMS.2021.3057679.

[4] C. D. Alwis et al., "Survey on 6G Frontiers: Trends, Applications, Requirements, Technologies and Future Research," in IEEE Open Journal of the Communications Society, vol. 2, pp. 836-886, 2021, DOI: 10.1109/OJCOMS.2021.3071496.

[5] I. A. Alimi, A. L. Teixeira and P. P. Monteiro, "Toward an Efficient C-RAN Optical Fronthaul for the Future Networks: A Tutorial on Technologies, Requirements, Challenges, and Solutions," in IEEE Communications Surveys Tutorials, vol. 20, no. 1, pp. 708-769, Firstquarter 2018, DOI: 10.1109/COMST.2017.2773462.

[6] J. Kani, J. Terada, K. Suzuki and A. Otaka, "Solutions for Future Mobile Fronthaul and Access-Network Convergence," in Journal of Lightwave Technology, vol. 35, no. 3, pp. 527-534, 1 Feb.1, 2017, DOI: 10.1109/JLT.2016.2608389.

[7] Pizzinat, Anna, et al. "Things you should know about fronthaul." Journal of Lightwave Technology 33.5 (2015): 1077-1083.

[8] J. Hoydis, S. ten Brink and M. Debbah, "Massive MIMO: How many antennas do we need?," 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2011, pp. 545-550, DOI: 10.1109/Allerton.2011.6120214.

[9] Gábor Fodor, László Pap and Miklós Telek, "Recent Advances in Acquiring Channel State Information in Cellular MIMO Systems", Infocommunications Journal, Vol. XI, No 3, September 2019, pp. 2-12. DOI: 10.36244/ICJ.2019.3.2

[10] N. Bhushan et al., "Network densification: the dominant theme for wireless evolution into 5G," in IEEE Communications Magazine, vol. 52, no. 2, pp. 82-89, February 2014, DOI: 10.1109/MCOM.2014.6736747.

[11] Hossain, Md Farhad, et al. "Recent research in cloud radio access network (c-ran) for 5g cellular systems-a survey." Journal of Network and Computer Applications 139 (2019): 31-48, DOI: 10.1016/j.jnca.2019.04.019.

[12] M. Masoudi, S.S. Lisi, C. Cavdar, Cost-efective migration toward virtualized C-RAN with scalable fronthaul design. IEEE Syst. J. 1–11 (2020). DOI: 10.1109/JSYST.2020.2982428.

[13] Ejaz, Waleed, et al. "A comprehensive survey on resource allocation for CRAN in 5G and beyond networks." Journal of Network and Computer Applications 160 (2020): 102638, DOI: 10.1016/j.jnca.2020.102638.

[14] Morais, Douglas H. "5G and Beyond Wireless Transport Technologies.", DOI: 10.1007/978-3-030-74080-1.

[15] Raddo, T.R., Rommel, S., Cimoli, B. et al. Transition technologies towards 6G networks. J Wireless Com Network 2021, 100 (2021). DOI: 10.1186/s13638-021-01973-9

[16] M. A. Khalighi and M. Uysal, "Survey on Free Space Optical Communication: A Communication Theory Perspective," in IEEE Communications Surveys Tutorials, vol. 16, no. 4, pp. 2231-2258, Fourthquarter 2014, DOI: 10.1109/COMST.2014.2329501.

[17] Sharma, Teena, Abdellah Chehri, and Paul Fortier. "Review of optical and wireless backhaul networks and emerging trends of next generation 5G and 6G technologies." Transactions on Emerging Telecommunications Technologies 32.3 (2021): e4155, DOI: 10.1002/ett.4155.

[18] Aditi Malik, Preeti Singh, "Free Space Optics: Current Applications and Future Challenges", International Journal of Optics, vol. 2015, Article ID 945483, 7 pages, 2015. DOI: 10.1155/2015/945483

[19] Marotta, A., Correia, L.M. Cost-effective joint optimisation of BBU placement and fronthaul deployment in brown-field scenarios. J Wireless Com Network 2020, 242 (2020). DOI: 10.1186/s13638-020-01844-9

[20] Tonini, Federico, et al. "Cost-optimal deployment of a C-RAN with hybrid fiber/FSO fronthaul." Journal of Optical Communications and Networking 11.7 (2019): 397-408, DOI: 10.1364/JOCN.11.000397.

[21] M. Klinkowski, "Planning of 5G C-RAN with Optical Fronthaul: A Scalability Analysis of an ILP Model," 2018 20th International Conference on Transparent Optical Networks (ICTON), 2018, pp. 1-4, DOI: 10.1109/ICTON.2018.8473987.

[22] C. Ranaweera, C. Lim, A. Nirmalathas, C. Jayasundara and E. Wong, "Cost-Optimal Placement and Backhauling of Small-Cell Networks," in Journal of Lightwave Technology, vol. 33, no. 18, pp. 3850-3857, 15 Sept.15, 2015, DOI: 10.1109/JLT.2015.2443066.

[23] H. Dahrouj, A. Douik, F. Rayal, T. Y. Al-Naffouri and M.-S. Alouini, "Cost-effective hybrid RF/FSO backhaul solution for next generation wireless systems," in IEEE Wireless Communications, vol. 22, no. 5, pp. 98-104, October 2015, DOI: 10.1109/MWC.2015.7306543.

[24] Y. Li, N. Pappas, V. Angelakis, M. Pióro and D. Yuan, "Optimization of Free Space Optical Wireless Network for Cellular Backhauling," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 9, pp. 1841-1854, Sept. 2015, DOI: 10.1109/JSAC.2015.2432518.

[25] Jaffer, Syed Saeed, et al. "A low cost PON-FSO based fronthaul solution for 5G CRAN architecture." Optical Fiber Technology 63 (2021): 102500.ibliography, DOI: 10.1016/j.yofte.2021.102500.

[26] C. Ranaweera et al., "Optical Transport Network Design for 5G Fixed Wireless Access," in Journal of Lightwave Technology, vol. 37, no. 16, pp. 3893-3901, 15 Aug.15, 2019, DOI: 10.1109/JLT.2019.2921378.

[27] Ranaweera, Chathurika, et al. "Rethinking of Optical Transport Network Design for 5G/6G Mobile Communication.

[28] M. Mahloo, P. Monti, J. Chen and L. Wosinska, "Cost modeling of backhaul for mobile networks," 2014 IEEE International Conference on Communications Workshops (ICC), 2014, pp. 397-402, DOI: 10.1109/ICCW.2014.6881230.

[29] Lenstra, Jan K., and AHG Rinnooy Kan. "Computational complexity of discrete optimization problems." Annals of discrete mathematics. Vol. 4. Elsevier, 1979. 121-140, DOI: 10.1016/S0167-5060(08)70821-5.

[30] Leitner, Markus, Mario Ruthmair, and Günther R. Raidl. "Stabilized column generation for the rooted delay-constrained Steiner tree problem." Agra, Agostinho and Doostmohammadi, Mahdi (2011) A Polyhe- dral Study of Mixed 0-1 Set. In: Proceedings of the 7th ALIO/EURO Workshop. ALIO-EURO 2011, Porto, pp. 57-59. 2011.

[31] Mitcsenkov, A., Paksy, G. Cinkler, T. Geography- and infrastructure-aware topology design methodology for broadband access networks (FTTx). Photon Netw Commun 21, 253–266 (2011). DOI: 10.1007/s11107-010-0297-4

[32] Schrijver, Alexander. Theory of linear and integer programming. John Wiley Sons, 1998.

[33] ITU-T G series Supplement 66, Supplement 66 to ITU-T G series Recommendation: ,5G Wireless fronthaul requirements in a PON context, 2019.

[34] P. Schulz et al., "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," in IEEE Communications Magazine, vol. 55, no. 2, pp. 70-78, February 2017, DOI: 10.1109/MCOM.2017.1600435CM.

[35] G. O. Pérez, J. A. Hernández and D. L. López, "Delay analysis of fronthaul traffic in 5G transport networks," 2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB), 2017, pp. 1-5, DOI: 10.1109/ICUWB.2017.8250956.

[36] Alkholidi, Abdulsalam Ghalib, and Khaleel Saeed Altowij. "Free space optical communications – Theory and practices." Contemporary Issues in Wireless Communications (2014): 159-212.

[37] I. Kim, Analysis and applications of microstructure and holey optical fibers, Ph.D. dissertation, Virginia Polytechnic Institute and StateUniversity, 2003.

[38] L. Chiaraviglio et al., "Bringing 5G into Rural and Low-Income Areas: Is It Feasible?," in IEEE Communications Standards Magazine, vol. 1, no. 3, pp. 50-57, SEPTEMBER 2017, DOI: 10.1109/MCOMSTD.2017.1700023.

[39] Sousa, Ivo, Nuno Sousa, Maria P. Queluz, and António Rodrigues. 2020. "Fronthaul Design for Wireless Networks" Applied Sciences 10, no. 14: 4754. DOI: 10.3390/app10144754

[40] IBM ILOG CPLEX Optimization Studio V12.6.3.

[41] G. V. Arévalo, M. Tipán and R. Gaudino, "Techno-Economics for Optimal Deployment of Optical Fronthauling for 5G in Large Urban Areas," 2018 20th International Conference on Transparent Optical Networks (ICTON), 2018, pp. 1-4, DOI: 10.1109/ICTON.2018.8473801.

[42] F. Yaghoubi et al., "A Techno-Economic Framework for 5G Transport Networks," in IEEE Wireless Communications, vol. 25, no. 5, pp. 56-63, October 2018, DOI: 10.1109/MWC.2018.1700233.

Abdulhalim Fayad received a B.Sc.degree in Electronics and Communications Engineering at Damascus University, Damascus, Syria, in 2014. He received his M.Sc. degree in Advanced Communication Engineering from Damascus University in 2019. He is currently pursuing his Ph.D.studies at the HSN Lab Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics (BME), Hungary. His main research interests are the Optical Access Networks, Passive Optical Networks, optimization of optical access networks for 5G and Beyond, designing cost-efficient fronthaul/backhaul for 5G, and resource allocation for optical networks.

Tibor Cinkler has received MSc (1994) and Ph.D. (1999) degrees from the Budapest University of technology and Economics (BME), Hungary, where he is currently a full professor at the Department of Telecommunications and Media Informatics (TMIT). He habilitated in 2013 and received the DSc degree from the Hungarian Academy of Sciences the same year. He has been visiting professor at the Department of Computer Communications at the Gdansk University of Technology, Poland, in the academic year of 2018/2019. His research interests focus on the optimization of routing, traffic engineering, design, configuration, dimensioning, and resilience of IP, Ethernet, MPLS, OTN, 4G/5G, and particularly of heterogeneous GMPLS-controlled WDM-based multi-layer networks. He is the author of over 300 refereed scientific publications, including four patents, with over 2300 citations. He has been involved in numerous related European and Hungarian scientific projects and many industrial co-operations including ACTS METON and DEMON; COST 266, 291, 293, 15127; IP NOBEL I and II and MUSE; NoE e-Photon/ONe, e-Photon/ONe, and BONE; CELTIC PROMISE and TIGER2; NKFP, GVOP, ETIK, as well as Ericsson, Magyar Telekom, Vodafone industrial projects. He has been a member of ONDM, DRCN, BroadNets, AccessNets, IEEE ICC and Globecom, EUNICE, CHINACOM, Networks, WynSys, ICTON, RNDM, HPSR, etc. Scientific Programme Committees. He has been guest editor of a FeatureTopic of the IEEE ComMag and reviewer for many journals. He has organized DRCN 2001 and 2013, ONDM 2003 and 2017, as well as Networks 2008, HPSR 2015, RNDM 2011, ICUMT 2011 conferences in Budapest. He teaches various courses on networking and optimization at the niversity, as well as for companies, and also gives tutorials at conferences and summer and winter schools. He received numerous awards, including Dimitris Chorafas Prize for Engineering, ICC best paper award, numerous HTE awards (HTE is the Hungarian IEEE sister society), including Tivadar Puskas, Virag-Pollak prize 4 times, and the HTE 60-year anniversary medal, Bolyai Medal, etc. He has been an ERC panel member. He chairs the IFIP WG 6.10 on Optical Networking.

Analytical Review and Study on Various Vertical Handover Management Technologies in 5G Heterogeneous Network

Kotaru Kiran and D. Rajeswara Rao

***Abstract*—In recent mobile networks, due to the huge number of subscribers, the traffic may occur rapidly; therefore, it is complex to guarantee the accurate operation of the network. On the other hand, the Fifth generation (5G) network plays a vital role in the handover mechanism. Handover management is a prominent issue in 5G heterogeneous networks. Therefore, the Handover approach relocates the connection between the user equipment and the consequent terminal from one network to another. Furthermore, the handover approaches manage each active connection for the user equipment. This survey offers an extensive analysis of 50 research papers based on existing handover approaches in the 5G heterogeneous network. Finally, existing methods considering conventional vertical handover management strategies are elaborated to improve devising effective vertical handover management strategies. Moreover, the possible future research directions in attaining efficient vertical handover management in a 5G heterogeneous network are elaborated.**

***Index Terms*—Handover, Fifth Generation, Software-defined network, Ping-pong handover, Handover success probability.**

I. Introduction

The next-generation communication structure provides Internet connectivity by several wireless approaches. The usage of various wireless techniques is growing rapidly for the communication system. The 4G Long term evolution (4G-LTE) wireless technologies have several factors, like the accessibility of devices, namely smart phones, notebooks, laptops, and so on, termed Long-term evolution and Wireless Local Area Network (WLAN) various networks simultaneously [1]. 5G is the imminent mobile cellular network technology to enhance the quality of service (QoS), low latency and elevated data rate. However, the 5G network has 10 to 100 times greater base station potential than the current 4G Long-term evolution networks. Generally, 5G network operates on up to the 3 frequency bands such as high, low as well as medium [2]. In the internet mobility management protocols, secure handover of the mobile nodes is an important safety issue. The management of handover reliably and efficiently is a significant challenge for the handover management. The re-authentication process is an important factor for the handover delay, which guarantees a secure network. However, the handover delays affect the cell sizes; therefore, the QoS may be reduced [3, 4].

Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
Kotaru.kiran@gmail.com

Several handover approaches are recently introduced for various circumstances, but most approaches are unsuitable for inter-domain handover. They failed to concurrently convince the requirements of key agreement, mutual authentication, and several other factors [5]. Thus, soft computing techniques, such as chaos theory [6] [7] [8], neural network [9], genetic programming, and fuzzy controller [8], are extensively used in telecommunication systems [10] [8]. The soft computing methods applied in the 5G systems provide more capacity in traffic controlling and many other systems related to decision making [11]. It is widely used in industry and academia because of advantages, like enhanced energy efficiency throughput, offload cellular traffic, and robustness [12]. Furthermore, the elliptic curve cryptography is also employed for a handover mechanism, enhancing security and reducing latency [13] [14]. The complexity and reduces authentication delay can get reduced by lightweight physical layer authentication approach [15] [14]. Effectual handover management is important in cellular networks extended with small a cell that enables multiple coverage and therefore maximized capacity in certain geographical areas. This work suggests the possible solutions for the handover management issues in 5G heterogeneous network.

The paper is arranged as below: Section 2 describes conventional approaches in the 5G heterogeneous network and the section 3 demonstrates the research gaps and issues of handover mechanisms. Section 4 represents the analysis of handover techniques in terms of utilized software, toolset used, performance metrics, Year of publication, and at last, the conclusion of the paper presents in section 5.

II. Literature Survey

This research considers the various approaches developed for the handover mechanism in 5G heterogeneous networks described in this section. The categorization of handover approaches in 5G heterogeneous networks is shown in Figure 1. Various approaches, such as Radio access-based approaches, self optimization-based approaches, software-defined network-based approaches, authentication-based approaches, eNodeB-based approaches, neural network-based approaches, and blockchain-based approaches, are developed for the handover approaches in 5G heterogeneous networks. The challenges associated with these methods are assessed for motivating the researchers to develop the new handover mechanism for 5G heterogeneous networks.

DOI: 10.36244/ICJ.2022.2.3

A. Classification of handover techniques

The homogeneous network is the one in which all the nodes employ the same operation. While in the heterogeneous network, the nodes perform both the function and utility. Besides, the heterogeneous networks based handover offers a better quality of services to the user with high availability of connections. The research works adapted several approaches utilized for handover mechanism in 5G heterogeneous networks are explained below, in Figure 1.

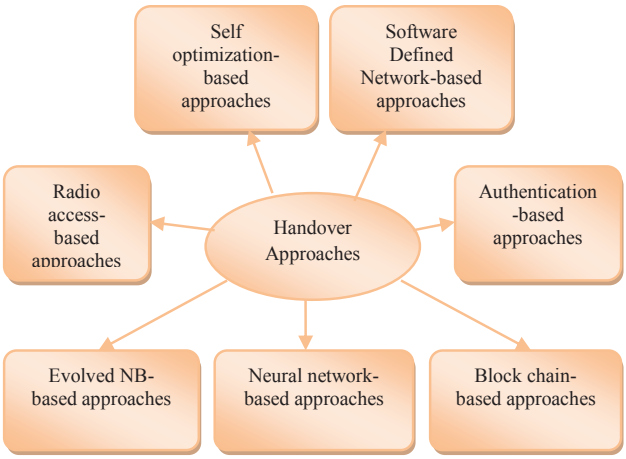


Fig. 1. Classification of Handover approaches

a) Radio access-based techniques

This section describes the Radio access-based approaches applied to manage the handover mechanisms in 5G heterogeneous networks. Zhang et al.[16] developed the cloud radio access network for the handover method in the 5G heterogeneous networks. The coordinated multipoint clustering technique based on affinity propagation was applied to diminish cell edge users' interference. After that, the low complexity handover management method was employed, and the signaling process was analyzed in the heterogeneous network. The coordinated multipoint was employed with the heterogeneous network for enhancing the cell edge user spectral efficiency and system coverage. Combining the handover and clustering methods can rapidly improve the heterogenous network's ability and preserve service quality. However, this method was not developed in the self-organized heterogeneous cloud small cell network for controlling the interference mitigation and handover management.

Stamou et al. [17] presented a context-aware handover approach. This advanced method integrated context-aware theory and multiple attribute decision-making concerning the radio access technique. The preventable handover of user equipment among the radio access technologies was decreased by this introduced method. Here, the developed technique was formed by a context-aware analytical hierarchy process for obtaining weight. Consequently, the Context-aware technique for Order Preference by Similarity to an Ideal Solution (CTOPSIS) method was also employed. The long-term cell association in terms of path loss was included for avoiding computational cost. In addition, the location error method on the global positioning system was also introduced for computing the latency. The developed method was not included the virtualized and programmable design for better efficiency.

Maksymyuk et al. [18] developed the converged access network for handover mechanism in the 5G heterogeneous network. Here, the wireless access segments and the optical backhaul were included with the radio access network. Furthermore, the developed method has a good granularity bandwidth allocation. Moreover, this method permits the channel of cloud radio access network to adapt the radio signals among the remote radio head and baseband processing unit in similar resource blocks. Additionally, the multicast data transmission was also developed to the complicated eNodeB through the mutual task of resource elements for multiple cells. This data transmission was developed in the handover process, which effectively reduces the backhaul traffic. The drawback is that it creates congestion in the network.

Kaloxylos et al. [19] introduced the multi-criteria handover method in the 5G heterogeneous network. This method was used to obtain the essential contextual information and choose the best suitable radio access network. The user equipment gathered the local instance of access network discovery and selection function in this approach. Furthermore, this method involves the fuzzy logic controllers to combine the various inputs, like a load of candidate base station and user mobility. After that, the session launch or per-flow handover was executed based on the third-generation partnership project. Finally, the classification of applications regarding latency and sensitivity was extracted through the user equipment connection manager. They achieved better throughput and delay. The method failed to include a reinforcement learning method for identifying the appropriate thresholds in the fuzzification procedure.

Polese et al. [20] developed the dual connectivity heterogeneous network handover mechanism. The dual connectivity approach allows the mobile user equipment procedure for maintaining the physical layer connections. The uplink control signaling method was joined along with the local coordinator, which allows the path switching in the event of failures. In addition, the utilization of a local coordinator controls the traffic among the cells. The network handover process was utilized for enhancing mobility management in the millimeter-wave network. Finally, the switch decision timing is improved using the busy time to trigger adaptation. The method was failed to use the concurrent millimeter-wave channel measurements for the precise analytical method.

Barmpounakis et al. [21] presented the radio access network method for the handover process in a 5G heterogeneous network. The fuzzy logic method was utilized in the context-aware selection technique for selecting the best suitable radio access technology. Moreover, the network extensions were developed for allowing access to network discovery. After that, the selection function method was employed to provide information about network status. The various groups of parameters, such as mobility, bandwidth, and power consumption, were computed through the network and user equipment. Consequently, the fuzzy logic system was employed for managing the multi-criteria issues. However, the developed method was not included the adaptive sampling rate for optimizing the battery power of user equipment.

Addali et al. [22] developed the Utility-based Mobility Load Balancing (UMLB) technique. This handover method was named the Load Balancing Efficiency Factor (LBEF). The mobility load balancing is employed using the user utility and the operator. Moreover, a centralized controller was also included to

balance the small cells. Consequently, for improving the performance of the network, key performance indicators were applied. Finally, an adaptive threshold method was also developed to identify the overload state of small cells. This developed UMLB method decreases the standard deviation by the higher average user equipment data rate. The method failed to observe the impacts of the mobility patterns and user equipment distributions to improve the UMLB method's effectiveness.

A generalized Random-Access Channel Handover (RACH) technique was developed by Choi and Shin [23] for handover in a 5G heterogeneous network. This developed approach obtained the perfect mobility with the absence of a synchronized network. This developed RACH technique combined the Make Before Break (MBB) handover and RACH less handover. The seamless mobility was obtained through the developed RACH technique by relocating from the serving cell to the user equipment. Subsequently, the developed method contains several key features, which correspondingly work along with the long-term evaluation handover, and it also does not influence any other delay factors. Although, this developed method was failed to involve packet duplication for optimizing the path switching.

The rateless properties of the channel conditions are detailed in Mehran et al. [24]. The rateless coding is used for protecting the packets at the physical layer through the noisy channel. The drawback of the rateless channel coding is that it is not applicable for the higher layers because decoding certain information will be lost.

Bogale et al. [25] reviewed the millimeter-wave technologies solve several issues in the 5G heterogeneous networks, enhancing spectral and energy efficiency. It is widely used in high-speed wireless networks and different communication purposes. Besides, large-scale antennas increase the performance of the system. The challenges in the 5G heterogeneous networks are spectral efficiency and energy efficiency to reduce operating costs.

b) Self-optimization-based approaches

The self-optimization approaches employed in the handover mechanism for the 5G heterogeneous networks are detailed in this subsection. Alhammadi et al. [12] introduced the weighted fuzzy optimization technique for handover. It was developed for optimizing the handover control parameters. The developed method used 3 features, namely velocity of user equipment, target and serving base station traffic load, and signal-to-interference-plus-noise ratio. Additionally, the self-optimized handover control parameters and were altered based on these characteristics to enhance handover performance. Although, the developed weighted fuzzy optimization method reduces the handover ping-pong, handover failure, and radio link failure. It also enhances the performance of the system along with various mobile speeds. On the other hand, the developed method was not involved in the cell pair-specific handover offsets for enhancing handover performance. The handover self-optimization technique was introduced for the handover mechanism, but the development was failed to solve the inter-cell interference problem.

Boujelben et al. [26] employed the handover self-optimization technique in the 5G heterogeneous network. This developed approach was utilized for decreasing the consumption of energy. Here, the handover target cell considered the neighbor cell load, user speed, and received signal power for identifying high-loaded cells. However, the developed method has three significant handover decisions: received cell load, user speed, and signal level. Moreover, the universal mobile telecommunication terrestrial radio access network was considered an access layer, and the packet core was considered a control layer. This advanced method was decreased the energy consumption significantly. The inter-cell interference is the major challenge associated with the developed technique.

c) Software-defined network-based approaches

The software-defined network-based handover approaches in the handover mechanism for the 5G heterogeneous network are illustrated in this section. Bilen et al. [27] presented the software-defined network. In this method, the Markov chain formulation was employed for computing the neighbor eNodeB transition probabilities. This method was also applied for selecting optimal eNodeB and also allocated the mobile nodes. Here, the handover failure ratios were also computed based on the user number. Likewise, the observed delays were also estimated by using the densification ratio parameter. Finally, the dual-track estimation approach and allocation strategy were utilized for separating the data channels and control channels. The method was failed to compute the effects of several parameters.

Tartarini et al. [28] developed the combination of software-defined handover decision engine and software-defined wireless networking method for improving the handover in 5G heterogeneous network. Here, the wireless controller was utilized in the baseband pool for receiving the handover information. The controllers published the communication information, and the handover decision was processed for every user optimally. Furthermore, the candidate network selection approach was formulated for solving binary integer linear programming optimization problems. In addition, the adaptive timing approach was also performed to minimize the handover failures, and the unwanted handover was eliminated through user equipment mobility patterns. The developed method has not enhanced the performance by implementing semi-optimal solutions. However, the developed approach failed to improve network performance when higher diversity of network types.

Rizkallah and Akkari [29] introduced a software-defined network for the vertical handover method in 5G heterogeneous networks. By means of exploiting the software-defined network, the data plane, as well as the control plane, were set apart. In addition, the handover signaling message was reduced by using the software-defined network. After that, the software-defined controller was applied for collecting network information. Based on the software-defined controller, the best handover decision was taken, and every network's quality of service was improved.

Alfoudi et al. [3] developed the seamless mobility management method. Initially, a key-value distributed hash table was employed for catching user mobility. It was employed in distributed software-defined controller for addressing the seamlessness and scalability in which the mobile nodes were joined between the correlated software-defined controller. The developed approach was permitted to choose the appropriate access point and the assistance of controllers and mobile nodes. Because of the appropriate access point, the network's performance was enhanced, and the users were satisfied with the required network condition. However, the developed approach was failed to reduce the cell size and enhance security.

In order to conduct the handover procedures in 5G HetNets, multiple-attribute decision-making approaches have also attracted a great deal of interest. More specifically, Liya et al. [30] have shown that eNodeB and many mobile nodes were utilized to maximize the handover count. This advanced method widely reduced the communication time and handover training duration during a handover procedure. Furthermore, for monitoring and performing normal operations, the software-defined controller was applied. On the other hand, the entire operations were managed by a handover controller, and the data plan devices were stated through tables of open flow at the instance of the handover procedure. However, this handover method was not enhanced the execution time.

Hu et al. [31] introduced the intelligent vertical handover process in the 5G heterogeneous network. Here, Media Independent Handover (MIH) and a software-defined network were integrated to ensure a handover, which presents between the two potential networks. After that, the characteristics of the software-defined network were applied to allocate the best networks to obtain the final declaration. This structure correctly chooses an optimal network effectively, decreasing the huge traffic present during the handover procedure. On the other hand, this technique was failed to utilize high-energy transmission approaches for securing backhaul networks. The developed technique was not included the radio resource allocation features for an efficient handover process.

Gharsallah et al. [32] developed the software to define the network-based handover. Here, the network function visualization and the software-defined network were applied for the handover process. Especially, the Software-Defined Handover Management Engine (SDHME) was designed for managing the handover control method. Furthermore, the developed SDHME technique was defined in the application plane of software-defined network structure. In addition, the developed approach was determined the better handover for mobile nodes to increase the quality of service. They achieved a reduced handover failure ratio and delay. The resource allocation is not evaluated for the better optimization of the resources.

Valiveti and Rao [33] presented the exemplary handover method in 5G heterogeneous networks during Device to Device (D2D) communication. The vertical handover process was performed in Content Delivery Network (CDN) using fuzzy logic. Furthermore, the Cramer-Shoup Key Encapsulation Mechanism (CS KEM) was utilized to provide security to users. In addition, the quality of service parameters, such as coverage area, security, and handover mechanisms, was enhanced. Here, the fuzzy rule was employed for handover management, and also L7 switch was utilized for reducing the load. The parameters, like signaling overhead, the number of handovers, signaling cost, end-to-end delay, bandwidth, and throughput, were considered for better D2D communication. However, the developed method failed to improve the network capacity and the energy efficiency.

Sadik et al. [34] introduced the software-defined handover. Here, the software-defined network and the fuzzy logic system were integrated for helping D2D communication. After that, user equipment decides the final handover decision to choose the better network in terms of predicted quality of service. Finally, suitable power control and frequency reuse was employed to increase network capacity and reduce interference. Additionally, the software-defined network was utilized to enhance the handover decision and reduce the decision phase delay and network detection. They achieved delay tolerance in service, whereas single point failure and security is the challenge of the SDN handover approach.

Yaseen and Al-Raweshidy [35] developed a handover mechanism in a 5G heterogeneous network. Here, the tag was generated for identifying the mobile nodes with media access control. Based on the identity of media access control, the mobile node tag was created for controlling the packets inside the mobile network. In the software-defined network, the application layer was performed for managing the interactions of controlling data. Similarly, the control plane was performed for controlling the forwarding decisions and rules of network data. They achieved low packet loss and delay for seamless communication. They failed to consider the registration delay of the network.

The capability-based privacy protection handover authentication method was introduced by Cao et al. [36] in 5G heterogeneous networks. The developed approach was the combination of software-defined network and user capability. Moreover, this technique was employed for obtaining the key agreement and mutual authentication among the base station and user equipment with decreased authentication handover cost. Furthermore, the method effectively protects various mobility scenarios, like intra radio access technology and inter-road access networks. They achieved reduced communication cost and computational cost. Besides, the developed method is efficient and secure handover compared to the existing systems. However, it doesn't withstand unknown attacks.

Kukliński et al. [37] presented the software-defined network for handover management in a 5G heterogeneous network. The centralized or semi-centralized plane was executed to obtain the handover and scalability issues. Subsequently, the centralized technique was employed for reducing the number of messages in the handover process. Finally, handover parameters based on various input parameters and data control were jointly optimized for transmission resources. They evaluated the procedure execution time and the handover execution time. The execution time of handover is higher for the inter-switch handover, which is the drawback of the developed methodology.

Duan and Wang [38] developed the software-defined network for handover mechanism in the heterogeneous network. The developed technique was utilized for enabling privacy protection and authentication handover. The privacy protection was enabled between the correlated access points. Furthermore, the developed software-defined network provides the reconfigurable network management platform and reduces the authentication handover latency. They evaluated the authentication delay and utilization rates. The single point failure and security is the challenge of the SDN handover approach.

d) Authentication-based approaches

This section portrays authentication-based methods employed from the various existing handover mechanism for 5G heterogeneous research. Om and Kumar [1] applied the Universal Subscriber Identity Module (USIM) and Elliptic Curve Cryptography (ECC) for the handover mechanism in the 5G heterogeneous network. A uniform handover authentication approach was also applied along with pairing approaches, and it was used in mobility improvements, like WLAN. The developed method provides secure communication between the

visited access point and user equipment in handover authentication. Moreover, the developed method was also offered enhanced computational and communication costs. A handover authentication method was adopted between the access points and mobile users to obtain mutual authentication among the user and target network. Finally, the session key was also computed among the access point and mobile user to attain secure communication. They achieved reduced storage and computation cost. Besides, it provides security against various attacks. While handovering for both the formal and informal security attacks they achieved better communication and storage costs with better authentication. The evaluation is performed on the basis of normal order is the drawback of the developed method.

Cao et al. [39] introduced the secure and efficient re-authentication and the group-based handover authentication procedure for 5G heterogeneous networks. The developed approach includes four stages for efficient handover. Moreover, this approach was used for obtaining robust security protection more effectively. At last, a detached session key was integrated into the network and the machine-type communication devices to attain the subsequent communications. They achieved better security with ideal efficiency. The developed method provides authentication against only some of the unknown attacks.

Ozhelvaci and Ma [4] presented the vertical handover approach for secure communication. The advanced authentication method can obtain strong, mutual, and quick authentication. Here, the certificate from a certification authority was obtained through the authentication server and user equipment. Moreover, a validation was present in the certificate of the user equipment to begin the communication among the server. Finally, the software-defined network was also applied to compact the rising requirements of wireless mobile networks, like mobile fog computing, e-health services, UAV systems, and vehicular communications. They achieved mutual authentication with security and integrity, but the method resists only the passive attacks.

Ma and Hu [14] developed the cross-layer collaborative handover. The method was designed using the upper layer and physical layer of cryptographic approaches for obtaining more reliable and rapid authentication. However, cryptographic approaches can attain high-security objectives, like data integrity, non-reputation, and data confidentiality. Initially, the handover authentication was performed in the physical layer, and then the Kolmogorov-Smirnov theory was applied through developing physical layer behaviors of the wireless channels. After the initial authentication process, the key agreement mechanism and extensible authentication protocol were also developed for obtaining more dependable security. They achieved efficient computational resources with more reliable security and reduced handover delay. The method failed to implement the distribution parameters.

Lee and You [40] employed the security method for the handover approaches. The ticket-based secure handover approach was also employed for the Fast PMIPv6 protocol (F-PMIPv6) protocol. Then, ticket encircling authentication protocol was applied in the mobile nodes for quick handover authentication. Additionally, SPFP has supported the mobile node ambiguity for conserving location privacy. The developed approach was also decreased the quantity of the authentication correlated message connections with the authentication server. The developed SPFP approach was mainly utilized for various security needs. The authentication latency, handover failure probability, handover latency, and buffered packets were also computed for finding the efficiency of the developed SPFP approach. However, the developed method was failed to decrease handover latency.

Fan et al. [41] employed a secure region-based handover technique in the 5G heterogeneous network. The developed region-based fast authentication protocol was applied to reduce communication and computation costs with no core network elements. Furthermore, the protocol assures the uniqueness of anonymity among communication footprints. After that, revocation of user membership was performed through the gathered one-way hash, and it was eliminated the computational cost in the 5G heterogeneous system. This developed approach was decreased the handover latency effectively through region-based secure handover. This advanced method was also satisfied the security conditions of each user. On the other hand, the developed approach did not include performance analytics and key management for providing security.

e) Evolved nodeB-based approaches

The eNodeB-based approaches in the handover mechanism for the 5G heterogeneous networks are detailed in this section. Zhang et al. [42] developed the Control/User Plane Split (CUPS) for executing the handover between the two neighbors. In the overlapping sector, the eNodeB was applied to reduce the failure of handover between two neighbours. In addition, the user equipment was preserved dual connectivity among the two neighbours. Subsequently, the user equipment was managed by the real connection. This method has improved the dependability of communication systems and also attains the better performance of handover. Besides, the handover outage is reduced. The drawback of the method is, the handover decision signaling, and control information is not considered for the analysis.

Huang et al. [43] employed the dynamic Femtocell gNB (F-gNB) on/off system for the handover process in a 5G heterogeneous network. This technique was effectively improved the energy efficiency of the network through calculation of traffic load. The developed optimization technique was detached into two levels. Furthermore, the developed CALB approach was protected the minimum SIR requirements of user equipment and load balancing. In addition, the introduced DFOO method includes the base station operation as per the forecasting time. At last, the dual connectivity-based perfect handover technique was introduced to protect the transmission of quality of service and user equipment. They analyzed the load prediction SINR requirements, and hence improved the QoS of the network. The simulation time increases with the increase in learning time is the drawback of the method.

Bilen et al. [44] developed the optimal eNodeB selection approach. The gain function was computed with dynamic weights for selecting the candidate cells. Here, the Kriging Interpolator and Semivariogram analysis were evaluated by the autoregressive method in spatial estimation to choose the optimal eNodeB. The Kriging Interpolations statistical and stochastic behaviors offer the best modeling performance. The unidentified indicator value of mobile user equipment was also computed through the definite values of neighbor user equipment. Every operation was executed using the developed enodeB estimation entity, which correlates with every network node. In addition, these

evaluations were also employed in both control and data channels separately. They reduced the unnecessary, frequent, and ping-pong handover risk, but it doesn't change the throughput.

Islam et al. [45] devised an Integrated Access Backhaul (IAB), allowing to extend coverage by providing wireless backhaul. Besides, the partition between the nodes may change concerning the demand of the network, which is not possible in the fixed access network. Hence, they have a substantial impact on providing seamless connection and avoiding frequent handovers in HetNets. It also reduces the fiber deployment and reduced interference because of narrow beamwidth. The method failed to solve the mixed-integer linear programming (MILP) problem.

f) Neural network-based approaches

The research that used neural network-based techniques is illustrated in this section. Semenova et al. [8] applied a neuro-fuzzy controller for improving the handover mechanism in 5G heterogeneous networks. This controller was developed with three linguistic variable inputs concerning the network's signal strength. In addition, the adaptive fuzzy interface structure was utilized to decrease the handover breakdown rate in the 5G heterogeneous network, and hence the QoS was also enhanced. At last, the adaptive network fuzzy interference systems application was also utilized to enhance the attachment point selection process and avoid useless handovers. On the other hand, the developed method was not considered the other effective parameter for improving the accuracy.

Maksymyuk and Shubyn [9] employed the Recurrent Neural Network (RNN). Here, the neural network was executed using user mobility knowledge to offer maximum efficiency. Consequently, the gated recurrent unit-based neural network was also applied to enhance the performance of the developed system. In addition, the mobility load balance was also executed by locating a cell individual offset parameter. Additionally, the gated recurrent unit-based neural network was adapted for finding the movement of the subscriber. They predicted the traffic in the NN network and achieved an accuracy lower than 90%. Besides, the network parameters are not evaluated to check the efficiency, and benchmark data are also not used.

Morghare and Mishra [46] presented the neural network-based handover approach. Here, the Particle Swarm Optimization (PSO) approach was employed with the neural network scheme. The developed scheme was applied for fast delivery handover route and network selection for increasing the system efficiency. However, a huge amount of secondary users was also considered in a network to enhance the system's efficiency. Additionally, the developed neural network method was adopted for solving an optimization problem and network selection problems. At last, the network selection for the handover route and free route was identified for transferring data. They achieved a better fitness value with a reduced number of iterations while considering the interference and the population size. The method failed to consider the security and the malicious attacks of the network while handovering the mobile terminal.

g) Blockchain-based approaches

This section depicts the blockchain-based techniques gathered from the various existing handover methods in 5G heterogeneous network research works. Ma and Lee [47] developed the blockchain scheme for the 5G heterogeneous network handover process. In this method, Parallel Block-chain Key Derivation Function (PB-KDF) was applied, which controls the principles of Bitcoin blockchain for supporting the key derivation process structurally. Furthermore, the PB-KDF helps to enhance the handover performance. After that, the blockchain outside the crypto-currency realm was utilized to improve the security of the system. In this case, the mining process uses the handover key for supporting full backward and full-forward partition. Therefore, the developed PB-KDF technique enhances the performance of the handover and the security of the handover. They achieved better security in the intracellular handover phase but failed to consider the computation complexity because the key management produces the computation overhead and is not included in the processing phase.

Yazdinejad et al. [48] employed the blockchain-enabled authentication handover method in the 5G heterogeneous network. In addition, heterogeneous network management and the software-defined network were employed for maximizing programmability. In this method, the encryption resources were also utilized for preserving the security privacy of the user. Furthermore, the introduced technique was utilized to reduce the preventable re-authentication in recurrent handover among heterogeneous networks. The software-defined network was also adopted to protect the user's privacy and supply intelligent control among the heterogeneous cells. The optimized energy consumption and scalability are achieved, but the security, data leakage, and delay prevail while handovering, which is considered as the drawback of the system.

Zhang et al. [5] introduced a universal and robust handover approach. This developed approach was used to manage the key agreement. Furthermore, this robust, seamless method was used to obtain a universal handover authentication management. The randomness secrecy, key escrow freeness, and several factors were attained through the developed method. On the other hand, computation efficiency and communication efficiency were also enhanced by this developed scheme. However, the developed method was failed to provide mobile user-friendly and secure design handover authentication methods in 5G heterogeneous networks.

h) Other handover mechanism approaches

In this section, the other handover approaches applied for the 5G heterogeneous networks are explained below. Arshad et al. [2] developed the aware skipping technique for the 5G heterogeneous network handover mechanism. Here, the two-tier network was preoccupied with the poison cluster process, whereas the single-tier cellular network was preoccupied with a poison point process. Moreover, the introduced technique utilizes the cell size and user position to make the handover assessment, and thus the avoidable handovers were evaded easily. In the two network states, handover on user throughput was computed, which portrays the efficiency and gain of the developed approach. The applied method was failed to consider the tractable analysis for finding accurate efficiency.

Mahardhika et al. [49] applied the multi-criteria metrics for the vertical handover decision in a 5G heterogeneous network. This advanced method contains three network interfaces and the metrics, like network occupancy, traffic class, mobile

speed, and received signal strength, were considered for vertical handover. The equal priority approach was applied to avoid the unnecessary usage of network resources and balance the traffic load for every complex network, thus enhancing accuracy. After that, the mobile priority approach was employed for balancing the network function. Finally, the network priority approach was utilized for reducing network blocking probability. However, the developed technique failed to integrate other handover methods, like fuzzy logic, for improving the network's performance.

Chandavarkar, and Guddeti [50] introduced the simplified and enhanced multiple attributes alternate ranking approach. This method has effectively reduced the feature normalization and weight estimation dependency, thus enhancing the network selection reliability. After that, the network score and rank were computed to avoid rank reversal issues. Therefore, the developed multiple attributes ranking technique was applied for network selection and solving rank reversal issues. The developed method was not utilized the network simulator for effective performance analysis.

Ouali et al. [51] employed the D2D handover management approach in 5G heterogeneous network. The developed D2D approach permits the mobile terminals to communicate directly with no base station connections as well as, this developed approach has fulfilled the quality of service requirements. Furthermore, it was applied for computing the performance of handover with the practical Reference Point Group Mobility (RPGM) structure. Finally, the spectral expansion technique was employed for attaining the precise solution. Although the developed approach failed to consider D2D relay concerns to enhance the simulation.

Ahmed et al. [52] developed a handover approach for secure communication. This method was mainly applied to offer connectivity to mobile users everywhere. Here, the handover decision approaches were classified into five types based on the parameters. The received signal strength-based handover decision method was applied for providing low cost and high bandwidth. Furthermore, the users handle the different traffics, such as multimedia, data, and voice streams.

He et al. [53] introduced the adaptive handover trigger approach. Initially, the clustering and mobility pattern detection of the mobile node. After that, using the outcome of clustering, the multiple hidden Markov models were trained to determine the grid sizes. At last, the correspondence among every cluster center and test route was identified using the hidden Markov models. Consequently, the Adaptive Received signal strength Prediction (ARP) approach was performed for predicting the strength. The technique was failed to enhance the predication efficiency for the developed adaptive handover

Liu et al. [54] employed a fuzzy-based handover approach. This developed fuzzy clustering approach integrates multiple attributes decision approaches and fuzzy logic to ensure a handover mechanism. At first, the optimal neighboring base station at the handover target was selected. The handover target can compact efficiently with multiple conflicts characteristic. After that, the decision engine closeness coefficient was utilized for determining the triggering timing. Here, the triggering method was permitted the base station for deciding the triggering time. The subtractive approach was executed to detect the membership functions in fuzzy systems. Using this, the maintenance cost and the optimization cost is reduced and the number of handovers is also reduced. The security issues are not considered and network performances are not analyzed.

Danyang et al. [55] presented the effective handover technique and a comprehensive load index in the 5G heterogeneous network. Here, handover parameters were utilized for selecting the network, and it was divided into several modules. Moreover, the triangle module fusion operator was employed for total network load and optimized the switching method. Additionally, this technique was deployed for controlling higher-level user satisfaction as well as for reducing handover frequency. Finally, the block calculation technique was applied in this method to enhance the execution time of the system. The developed handover approach was failed to offer better QoS because of the heavy load.

Fang et al. [56] introduced the Long-term evolution-based handover approach. In this technique, U-plane and C-plane were differentiated to resolve crucial handover problems. Here, the gray system theory was employed to develop handover trigger decisions and predict several values to take the decision. In addition, different signaling was executed through a physical downlink shared channel, and it was planned through the channel. The other signaling element was utilized to generate and control the communication to ensure reliability. This method was failed to include Doppler Effect to manage handover problems.

Chopra et al. [57] developed a technique for handover in a 5G heterogeneous network. The thermal pattern approach was utilized to compute the precise position. Additionally, secrecy capacity analysis was executed on mobile user equipment with the consideration of various channel losses. Here, the thermal pattern-based tracing approach was utilized for effectively identifying low-security regions.

Lahby et al. [58] introduced the k-partite graph-based handover approach in the 5G heterogeneous network. Initially, k-partite graph theory was employed for representing the vertical handover problems. Here, the deployment of access points and user dynamics was considered through a mobile network operator. After that, the robust analytic hierarchy process was performed for calculating the weight of every edge. Finally, Dijkstra's technique was executed for identifying the better paths based on QoS. The packet loss and the delay are reduced with an increase in throughput value is achieved, but the computational cost and complexity are not analyzed. Besides, secure handover is also not analyzed.

III. Analysis and Discussion

This section illustrates the analysis and discussion of vertical handover approaches in 5G heterogeneous networks using various research papers based on the utilized dataset, categorization of techniques, performance evaluation metrics, and publication year.

A. Analysis based on approaches

This section represents the review using several vertical handover approaches in 5G heterogeneous networks. Several approaches devised for the vertical handover are shown in Figure 2. From Figure 2, it is recognized that 24% of the researchers utilized software-defined network-based approaches and 19% of the researchers are created radio access-based ap-

proaches. Moreover, an authentication- based approach was used by 13% of researchers, and 21% of the research works are based on other approaches. Likewise, 7% of the researchers utilized the neural network-based, 6% of the research works are based on blockchain-based, and eNodeB-based approaches, and the remaining 4% of the researchers developed self-optimization-based approaches. Hence, from this analysis, software-defined network-based approaches are extensively utilized for vertical handover in 5G heterogeneous networks.

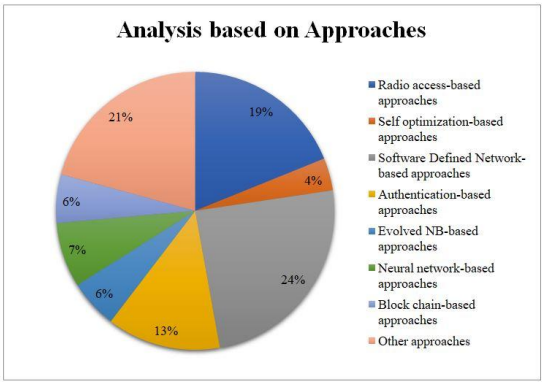


Figure 2. Analysis based on approaches

B. Analysis in terms of toolset

The different tools used by the traditional methods for the evaluation of the developed method are detailed in this section. The analysis in terms of several tools applied is displayed in Figure 3. The software toolsets employed in the research papers are Monte Carlo simulations, MATLAB, Java Pairing Based Cryptography (JPBC), Pro Verif tool, High-Level Protocol Specification Language, AVISPA, BAN logic, OMNeT++ simulator, Scyther, INET 3.6.4 framework, NS-3 simulator, system-level simulation, Java Cryptography Extension (JCE), and Mininet emulator. Based on Figure 3, it is realized that MATLAB is the widely used software toolset for evaluating the vertical handover approaches.

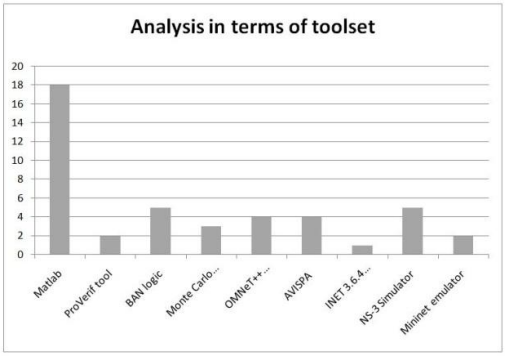


Figure 3. Analysis based on the toolset

C. Analysis based on publication year and the source

This section displays the analysis based on published years. The analysis using the published Year is represented in Table I. Out of the 50 papers surveyed, more number of research papers was published in the Year 2019. The analysis based on the publications is summarized in Table II. In IEEE journal more number of papers was published regarding the handover mechanism.

TABLE I
ANALYSIS USING PUBLISHED YEAR

Year	Journal	Conference
1999	2	-
2012	1	-
2013	1	-
2014	1	3
2015	3	4
2016	3	1
2017	7	4
2018	5	2
2019	15	5
2020	1	1

TABLE II
ANALYSIS BASED ON PUBLICATIONS

Publications	Journal	Conference
IEEE	20	5
Elsevier	7	-
Springer	2	-
KeAi	2	-
Hindawi	1	-
IAENG	1	-
Publons	2	-
Wiley	1	-
TUCS	2	-
Others	1	14

D. Analysis in terms of employed datasets

The analysis concerning the used dataset is detailed in this section. The precision and the legitimacy of the algorithm are validated effectively based on selecting the suitable dataset. Hence the best dataset selection also influences the performance of the system. Figure 4 shows the analysis of various datasets. The commonly utilized datasets in the vertical handover approach in 5G heterogeneous networks are Oracle, multi-dimensional radio-cognitive databases, GitHub, base station parameters, horizontal plane of artificial intelligence, Trajectory set, Fuzzy Logic Inference Systems, linguistic variables. From Figure 4, it is understandable that the most continually utilized dataset is Fuzzy Logic Inference Systems.

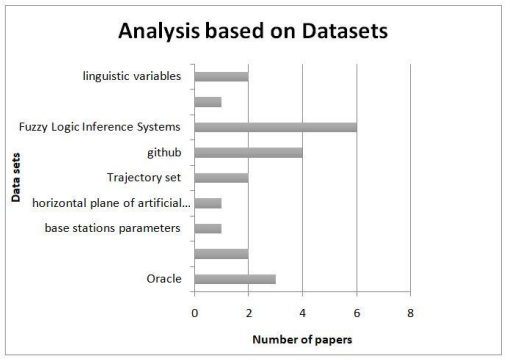


Figure 4. Analysis based on datasets

E. Analysis in terms of evaluation metrics

The performance metrics, such as ping-pong handover, spectrum efficiency, bandwidth, computation cost, communication cost, throughput, number of Handovers, handover failure, traffic overhead, packet loss, accuracy, signaling overhead, and handover success probability are considered for vertical handover approaches in 5G heterogeneous networks. From Table III, it is observed that the number of handovers, handover failure probability, handover success probability, and ping-pong handover are usually chosen performance metrics.

TABLE III
ANALYSIS USING EVALUATION METRICS

Performance metrics	Number of Research papers
Ping-pong handover	[16] [11] [12] [51] [44] [54] [56] [34] [23]
Spectrum efficiency	[16] [57]
Bandwidth	[27] [39] [36] [19] [20] [33]
Computation cost	[1] [36] [47] [5] [41]
Communication cost	[1] [39] [36] [5] [41]
Throughput	[2] [28] [43] [34] [58]
Number of Handovers	[27] [17] [29] [49] [26] [38] [52] [54] [31] [40] [19] [21] [33] [34] [35]
Handover failure probability	[2] [28] [11] [51] [52] [53] [44] [30] [40] [56] [32] [21] [34]
Traffic overhead	[3] [18]
Packet loss	[44] [40] [21] [58]
Accuracy	[9] [46] [40] [57] [58]
Signaling overhead	[29] [48]
Handover Success probability	[42] [29] [51] [52] [53] [30] [48] [40] [56] [20] [34]

F. Analysis based on the handover failure probability

The analysis in terms of handover failure probability is illustrated in this section. Furthermore, Table 4 portrays the review in terms of handover failure probability is specified by five ranges. From the below Table, it is recognized that the research papers [2] [56] obtained better handover failure probability within the range of 90% - 99%, and the research papers [51] [30] had less handover failure probability within the range 50% - 59%. When a handover failure happens the interruption time maximizes considerably to more than hundreds of milliseconds. Hence, to fullfil the requirement in as numerous cases as possible, we need to reduce the rate of handover Failure as close to zero as possible. As the network density increases, the frequency of handover increases that might outcome in higher handover failure rates.

TABLE IV
ANALYSIS USING HANDOVER FAILURE PROBABILITY RANGE

Handover failure probability range	Number of Research papers
50% - 59%	[51] [30]
60% - 69%	[52] [53] [34]
70% - 79%	[28] [44] [32]
80% - 89%	[11] [40] [21]
90% - 99%	[2] [56]

IV. CONCLUSION

In this study, a survey is done on various vertical handover methods in the 5G heterogeneous network. The reviews are collected from several research papers that categorized into Radio access-based approaches, self optimization-based approaches, Software Defined Network-based approaches, Authentication-based approaches, eNodeB-based approaches, Neural Network-based approaches, and Block chain-based approaches. The review of the traditional methods suggests future works for the vertical handover approaches in the 5G heterogeneous network by considering several research gaps and problems. The analysis of the survey is represented in terms of categorization techniques, utilized toolset, datasets used, and evaluation metrics. From this analysis, it is reviewed that a software-defined network-based approach is commonly used in research papers. Similarly, MATLAB is a frequently utilized software tool, and the fuzzy logic interference system database is a generally used toolset in the existing papers. Likewise, the number of handovers, handover failure probability, handover success probability, and ping-pong handover are the most generally used performance metric in most of the research papers.

Besides, the proposed review helps the researchers to identify the research gaps and devise the new technique to overcome the challenges faced by the existing systems. The performance of the heterogeneous network is enhanced by reducing the delay and energy consumption. The context-aware and QOS aware networks are flexible and are the most preferable for the vertical handover.

REFERENCES

[1] A. Kumar, H. Om, Design of a USIM and ECC based handover authentication scheme for 5G-WLAN heterogeneous networks", Digital Communications and Networks, July 2019. DOI: 10.1016/j.dcan.2019.07.003.

[2] R. Arshad, H. ElSawy, S.Sorour, T. Y. Al-Naffouri, M. S. Alouini, "Handover management in 5G and beyond: A topology aware skipping approach", IEEE Access, vol. 4, pp. 9073-81, December 2016. DOI: 10.1109/ACCESS.2016.2642538

[3] A. S. Alfoudi , S.S. Newaz, R. Ramlie, G. M. Lee, T. Baker, "Seamless mobility management in heterogeneous 5G networks: A coordination approach among distributed SDN controllers", In proceedings of IEEE 89th Vehicular Technology Conference (VTC2019-Spring), pp. 1-6, April 2019. DOI: 10.1109/VTCSpring.2019.8746712

[4] A. Ozhelvaci, M. Ma, "Secure and efficient vertical handover authentication for 5G HetNets", In proceedings of IEEE International Conference on Information Communication and Signal Processing (ICICSP), pp. 27-32, September 2018. DOI: 10.1109/ICICSP.2018.8549809

[5] Y. Zhang, R. Deng, E. Bertino, D. Zheng, "Robust and universal seamless handover authentication in 5G HetNets", IEEE Transactions on Dependable and Secure Computing, pp. 1545-5971, July 2019. DOI: 10.1109/TNSE.2019.2937481

[6] R. Fullér, "On fuzzy reasoning schemes", Turku Centre for Computer Science, 1999.

[7] R. Fuller, "Fuzzy logic and neural nets in intelligent systems", Information Systems Day.–Turku Centre for Computer Science, vol. 17, pp. 74-94, 1999. DOI: 10.1117/12.48402

[8] O. Semenova, A. Semenov, O. Voitsekhovska, "Neuro-Fuzzy Controller for Handover Operation in 5G Heterogeneous Networks", In proceedings of 3rd International Conference on Advanced Information and Communications Technologies (AICT), pp. 382-386, July 2017. DOI: 10.1109/AIACT.2019.8847898

[9] B. Shubyn, T. Maksymyuk, "Intelligent Handover Management in 5G Mobile Networks based on Recurrent Neural Networks", In proceedings of 3rd International Conference on Advanced Information and Communications Technologies (AICT), pp. 348-351, July 2019. DOI: 10.1109/AIACT.2019.8847734

[10] A. A. Atayero, M. K. Luka, "Applications of soft computing in mobile and wireless communications", International Journal of Computer Applications, vol. 45, no. 22, pp. 48-54, 2012. DOI: 10.1007/978-81-322-2407-5_16

[11] C. Suarez-Rodriguez, Y. He, E. Dutkiewicz, "Theoretical Analysis of REM-Based Handover Algorithm for Heterogeneous Networks", IEEE Access, vol. 7, pp. 96719-31, July 2019. DOI: 10.1109/ACCESS.2019.2929525

[12] A. Alhammadi, M. Roslee, M. Y. Alias, I. Shayea, S. Alriah, A. B. Abas A B, "Advanced handover self-optimization approach for 4G/5G HetNets using weighted fuzzy logic control", In proceedings of 15th International Conference on Telecommunications (ConTEL), pp. 1-6, July 2019. DOI: 10.1109/ConTEL.2019.8848507

[13] Y. E. El Idrissi, N. Zahid, M. Jedra, "An efficient authentication protocol for 5G heterogeneous networks", In proceedings of International Symposium on Ubiquitous Networking, pp. 496-50, May 2017. DOI: 10.1007/978-3-319-68179-5_43

[14] T. Ma, F. Hu, "A cross-layer collaborative handover authentication approach for 5G heterogeneous network", InJ. Phys. Conf. Series, vol. 1169, no. 1, February 2019. DOI: 10.1109/MWC.2015.7143331

[15] F. Pan, H. Wen, H. Song, T. Jie, L. Wang, "5G security architecture and lightweight security authentication", In Proceedings of IEEE/CIC International Conference on Communications in China-Workshops (CIC/ICCC), pp. 94-98, November 2015. DOI: 10.1109/ICCChinaW.2015.7961587

[16] H. Zhang, C. Jiang, J. Cheng, V. C. Leung, "Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks", IEEE Wireless Communications, vol. 22, no. 3, pp. 92-9, July 2015. DOI: 10.1109/MWC.2015.7143331.

[17] A. Stamou, N. Dimitriou, K. Kontovasilis, S. Papavassiliou, "Context-aware handover management for HetNets: Performance evaluation models and comparative assessment of alternative context acquisition strategies", Computer Networks, pp. 107272, May 2020. DOI: 10.1016/j.comnet.2020.107272

[18] T. Maksymyuk, O. Krasko, M. Kyryk, V. Romanchuk, R. Kolodiy, "Designing the new backhaul for 5G heterogeneous network based on converged optical infrastructure", Acta Electrotechnica et Informatica, vol. 17, no. 4, pp. 9-13, January 2017. DOI: 10.15546/aeei-2017-0028

[19] A. Kaloxylos, S. Barmpounakis, P. Spapis, N. Alonistioti, "An efficient RAT selection mechanism for 5G cellular networks", In proceedings of International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 942-947, August 2014. DOI: 10.1109/IWCMC.2014.6906482

[20] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks", IEEE Journal on Selected Areas in Communications, vol. 35, no. 9, pp. 2069-84, June 2017. DOI: 10.1109/JSAC.2017.2720338

[21] S. Barmpounakis, A. Kaloxylos, P. Spapis, N. Alonistioti, "COmpAsS: A Context-Aware, user-oriented RAT selection mechanism in heterogeneous wireless networks", InProc. Int. Conf. on Advanced Commun and Computation (INFOCOMP), july 2014. DOI: 10.1016/j.comnet.2016.12.008

[22] K. M. Addali, S. Y. Melhem, Y. Khamayseh, Z. Zhang, M. Kadoch, "Dynamic mobility load balancing for 5G small-cell networks based on utility functions", IEEE Access, vol. 7, pp. 126998-7011, September 2019.

[23] J. H. Choi, D. J. Shin, "Generalized RACH-Less Handover for Seamless Mobility in 5G and Beyond Mobile Networks", IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1264-7, May 2019. DOI: 10.1109/LWC.2019.2914435

[24] F. Mehran, K. Nikitopoulos, P. Xiao and Q. Chen, "Rateless wireless systems: Gains, approaches, and challenges," 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), 2015, pp. 751-755. DOI: 10.1109/ChinaSIP.2015.7230505

[25] T. E. Bogale and L. B. Le, "Massive MIMO and mmWave for 5G Wireless HetNet: Potential Benefits and Challenges," in IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 64-75, March 2016. DOI: 10.1109/MVT.2015.2496240

[26] M. Boujelben, S. B. Rejeb, S. Tabbane, "A novel green handover self-optimization algorithm for LTE-A/5G HetNets", In proceedings of International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 413-418, August 2015. DOI: 10.1109/IWCMC.2015.7289119

[27] T. Bilen, B. Canberk, K. R. Chowdhury, "Handover management in software-defined ultra-dense 5G networks", IEEE Network, vol. 31, no. 4, pp. 49-55, July 2017. DOI: 10.1109/MNET.2017.1600301.

[28] L. Tartarini, M. A. Marotta, E. Cerqueira, J. Rochol, C. B. Both, M. Gerla, P. Bellavista P, "Software-defined handover decision engine for heterogeneous cloud radio access networks", Computer Communications, vol. 115, pp. 21-34, January 2018. DOI: 10.1016/j.comcom.2017.10.018

[29] J. Rizkallah, N. Akkari, "SDN-based vertical handover decision scheme for 5G networks", In proceedings of IEEE Middle East and North Africa Communications Conference (MENACOMM), pp. 1-6, April 2018. DOI: 10.1109/MENACOMM.2018.8371040

[30] X. Liya, D. Anyuan, G. Mingzhu, S. Jiaoli, G. Guangyong, "A MADM-based Handover Management in Software-defined 5G Network", Engineering Letters, vol. 27, no. 4, December 2019. DOI: 10.1109/MWC.2015.7143331

[31] S. Hu, X. Wang, M. Z. Shakir, "A MIH and SDN-based framework for network selection in 5G HetNet: Backhaul requirement perspectives", In proceedings of IEEE international conference on communication workshop (ICCW), pp. 37-43, June 2015. DOI: 10.1109/ICCW.2015.7247072

[32] A. Gharsallah, F. Zarai, M. Neji, "SDN/NFV-based handover management approach for ultradense 5G mobile networks", International Journal of Communication Systems, vol. 32, no. 17, pp. e3831, November 2019. DOI: 10.1002/dac.3831

[33] H. B. Valiveti, P. T. Rao, "EHSD: an exemplary handover scheme during D2D communication based on decentralization of SDN", Wireless Personal Communications, vol. 94, no. 4, pp. 2393-416, June 2017. DOI: 10.1007/s11277-016-3490-7

[34] M. Sadik, N. Akkari, G. Aldabbagh, "SDN-based handover scheme for multi-tier LTE/Femto and D2D networks", Computer Networks, vol. 142, pp. 142-53, September 2018. DOI: 10.1016/j.comnet.2018.06.004

[35] F. A. Yaseen, H. S. Al-Raweshidy, "Smart Virtualization Packets Forwarding During Handover for Beyond 5G Networks", IEEE Access, vol. 7, pp. 65766-80, May 2019. DOI: 10.1109/ACCESS.2019.2915268

[36] J. Cao, M. Ma, Y. Fu, H. Li, Y. Zhang, "CPPHA: Capability-based privacy-protection handover authentication mechanism for SDN-based 5G HetNets", IEEE Transactions on Dependable and Secure Computing, May 2019. DOI: 10.1109/TDSC.2019.2916593

[37] S. Kukliński, Y. Li, K. T. Dinh, "Handover management in SDN-based mobile networks", In proceedings of IEEE Globecom Workshops (GC Wkshps), pp.194-200, December 2014.

[38] Duan X, Wang X, "Authentication handover and privacy protection in 5G hetnets using software-defined networking", IEEE Communications Magazine, vol. 53, no. 4, pp. 28-35, April 2015. DOI: 10.1109/MCOM.2015.7081072

[39] J. Cao, M. Ma, H. Li, Y. Fu, X. Liu, "EGHR: Efficient group-based handover authentication protocols for mMTC in 5G wireless networks", Journal of Network and Computer Applications, vol. 102, pp. 1-6, January 2018. DOI: 10.1016/j.jnca.2017.11.009

[40] I. You, J. H. Lee, "SPFP: Ticket-based secure handover for fast proxy mobile IPv6 in 5G networks", Computer Networks, vol. 129, pp. 363-72, December 2017. DOI: 10.1016/j.comnet.2017.05.009

[41] C. I. Fan, J. J. Huang, M. Z. Zhong, R. H. Hsu, W. T. Chen, J. Lee, "ReHand: secure region-based fast handover with user anonymity for small cell networks in 5G", pp. 1806.03406, June 2018. DOI: 10.1109/TIFS.2019.2931076

[42] Z. Zhang, Z. Junhui, S. Ni, Y. Gong, "A Seamless Handover Scheme With Assisted eNB for 5G C/U Plane Split Heterogeneous Network", IEEE Access, vol. 7, pp. 164256-64, November 2019. DOI: 10.1109/ACCESS.2019.2952737

[43] X. Huang, S.Tang, Q. Zheng, D. Zhang, Q. Chen,"Dynamic femtocell gNB on/off strategies and seamless dual connectivity in 5G heterogeneous cellular networks", IEEE Access, vol. 6, pp. 21359-68, January 2018. DOI: 10.1109/ACCESS.2018.2796126

[44] T. Bilen, T.Q. Duong, B. Canberk,"Optimal eNodeB estimation for 5G intra-macrocell handover management", In Proceedings of the 12th ACM Symposium on QoS and Security for Wireless and Mobile Networks, pp. 87-93, November 2016. DOI: 10.1145/2988272.2988284

[45] M.N. Islam, S. Subramanian and A. Sampath, "Integrated Access Backhaul in Millimeter Wave Networks", IEEE Wireless Communications and Networking Conference (WCNC), pp. 1-6, 2017. DOI: 10.1109/WCNC.2017.7925837

[46] O. P. Mishra, G. Morghare,"An Efficient approach Network Selection and Fast Delivery Handover Route 5G LTE Network", In proceedings of 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 857-862, April 2019. DOI: 10.1109/ICOEI.2019.8862791

[47] H. Lee, M. Ma,"Block chain-based mobility management for 5G", Future Generation Computer Systems, August 2019. DOI: 10.1016/j.future.2019.08.008

[48] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, K. K Choo,"Blockchain-enabled authentication handover with efficient privacy protection in SDN-based 5G networks", IEEE Transactions on Network Science and Engineering, August 2019. DOI: 10.1109/TNSE.2019.2937481

[49] G. Mahardhika, M. Ismail, R.Nordin,"Vertical handover decision algorithm using multi-criteria metrics in heterogeneous wireless network", Journal of Computer Networks and Communications, January 2015. DOI: 10.1155/2015/539750

[50] B.R. Chandavarkar, R.M Guddeti, "Simplified and improved multiple attributes alternate ranking method for vertical handover decision in heterogeneous wireless networks", Computer Communications, vol. 83, pp. 81-97, June 2016. DOI: 10.1016/j.comcom.2015.10.011

[51] K. Ouali, M. Kassar, T.M. Nguyen, K. Sethom, B. Kervella, "Modeling D2D handover management in 5G cellular networks" In proceedings of 13th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 196-201, June 2017. DOI: 10.1109/IWCMC.2017.7986285

[52] A. Ahmed, L.M. Boulahia, D. Gaiti,"Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification", IEEE Communications Surveys & Tutorials, vol.16, no. 2, pp. 776-811, August 2013. DOI: 10.1109/SURV.2013.082713.00141

[53] H. He, X. Li, Z. Feng, J. Hao, X. Wang, H. Zhang,"An adaptive handover trigger strategy for 5G C/U plane split heterogeneous network", In proceedings of IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), pp. 476-480, October 2017. DOI: 10.1109/MASS.2017.28

[54] Q. Liu, C.F. Kwong, S. Zhang, L. Li, J. Wang,"A fuzzy-clustering based approach for MADM handover in 5G ultra-dense networks", Wireless Networks, pp. 1-4, September 2019. DOI: 10.1007/s11276-019-02130-3

[55] L. Danyang, Z. Zhizhong, G. Yiyi,"Modular handover algorithm for 5G HetNets with comprehensive load index", The Journal of China Universities of Posts and Telecommunications, vol. 24, no. 2, pp. 57-65, April 2017. DOI: 10.1016/S1005-8885(17)60199-7

[56] H. Song, X. Fang, L. Yan,"Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway", IEEE Transactions on Vehicular Technology, vol. 63, no. 9, pp. 4633-46, April 2014. DOI: 10.1109/TVT.2014.2315231.

[57] G. Chopra, R. K. Jha, S. Jain,"TPA: prediction of spoofing attack using thermal pattern analysis in ultra dense network for high speed handover scenario", IEEE Access, vol. 6, pp. 66268-84, October 2018. DOI: 10.1109/ACCESS.2018.2875921

[58] M. Lahby, A. Essouiri, A. Sekkaki,"A novel modeling approach for vertical handover based on dynamic k-partite graph in heterogeneous networks", Digital Communications and Networks, vol. 5, no. 4, pp. 297-307, November 2019. DOI: 10.1016/j.dcan.2019.10.001

Kotaru Kiran having 17+ Years of Consulting, Program Management experience in AI/ML & Telecommunications (OSS/BSS). Kotaru Kiran have special emphasis on Experience in Optical Networking SONET / SDH / WDM / DWDM and management systems like EMS /NMS. Kotaru Kiran worked with various clients across the globe and has vast experience in understanding local people based on their geographical living style. Kotaru Kiran currently doing research on GPL – Garuda Programming Language – Sanskrit based AI framework. Conducted AI Roadshows, Hackathons, Developer Workshops across all European regions Kotaru Kiran received various honours and awards, 2018 - Applause for Team for outstanding contribution at client location, 2015 Champions of ILP Award, 2017 Multiple AIP Anchor Awards for outstanding contribution to the organization, 2016 Multiple AIP Anchor Awards for outstanding contribution to the organization, 1998 - 2000: Received BEST STUDENT AWARD from Aptech Computer Education, 2000 - 2002: Got SECOND PRIZE in presentation held at Osmania University Campus for OFC presentation, Appreciation from client for the commitment towards the project executing & meeting the deliverables in time, Nominated for the "Vibrant Heart" award for making team coordination and unity in Wipro Technologies.

Dr. Rajeswara Rao is a Motivating and talented professor driven to inspire students to pursue academic and personal excellence. He consistently strives to create a challenging and engaging learning Environment in which students becomes life-long learners. Rajeswara Rao got doctorate from JNTUH in 2011. Rajeswara Rao, area of research is Machine Learning, Soft Computing Techniques, Data Mining and Data Warehousing. Rajeswara Rao was honoured and awarded as Best Teacher Award for Academic Year 2016-2017, KLEF (Deemed to be University), One of the training faculty for Infosys from the selected JKC's in Andhra Pradesh, Mission10X Certificate in Teaching and Learning. Rajeswara Rao, conducted around 26 Faculty Development and Training Programmes(each program at least 5 Days) and around 50 workshops and 35 webinars. Rajeswara Rao is a Life Member in Computer Society of India (CSI), Indian Society of Technical Education and holds membership in ACM. Rajeswara Rao conducted 2 day seminars with grants supported by DST sponsored National Conference and KLEF (Deemed to be University) sponsored Enterprise Big Data Analytics.

Conducted emission simulation and measurement of interleaved DC-DC converters

Tamás Kőnig[1] and Lajos Nagy[2]

***Abstract*—Switching-Mode Power Supplies (SMPS) are often used to power on-board satellite payloads due to their good conversion efficiency. However, they emit radiated and conducted noise, which can disturb the operation of other payloads. The current ripple sum of the power supplies will appear on the power bus. There are many methods to reduce this summarized noise, one of them is to interleave the on-switching times of the converters. This ripple cancellation method can decrease the noise component on the switching frequency and on its upper harmonics. In this article we are going to demonstrate the effects of the distributed interleaving with a measurement platform consisting of two Pulse Width Modulation (PWM) controlled Buck controllers.**

***Index Terms*—switching-mode power supply, interleaving, conducted emission measurement, satellite power system**

I. INTRODUCTION

ON-board a satellite there are various payloads, which require different voltage levels to function. Switching-mode power supplies are typically used to convert the voltage of the power bus to the adequate level. In a satellite power system the inputs of the converters are parallel connected and because of this their input current ripples will be summarized on the bus. There are different regulations to limit the noise emission of the converters (e.g. European Cooperation for Space Standardization (ECSS) standard). Most of the time differential and common mode filters are used for noise mitigation. However, these filters often contain bulky components requiring a lot of space. In case of constant-frequency controlled power supplies the switching periods can be synchronized by an external clock. The phases of converter oscillator signals can be set to different values. With this phase division the on-switching times can be interleaved. In the case of two interleaved converters their input current ripples will be 180 degree out of phase, so they will cancel each other on the power bus [1]-[4]. The frequency of the summarized input current will be twice of the original, which can be filtered out with a smaller capacitor [3]. In the following chapters we will present the effects of the distributed interleaving with a setup consisting of two Buck converters. The setup was simulated in LTspice XVII (v.17.0.34.0) [5] simulation environment and later it was built and measurements were conducted.

[1,2]Department of Broadband Infocommunications and Electromagnetic Theory, Budapest University of Technology and Economics, Hungary
[1](e-mail: konigtamas@edu.bme.hu)
[2](e-mail: nagy.lajos@vik.bme.hu)

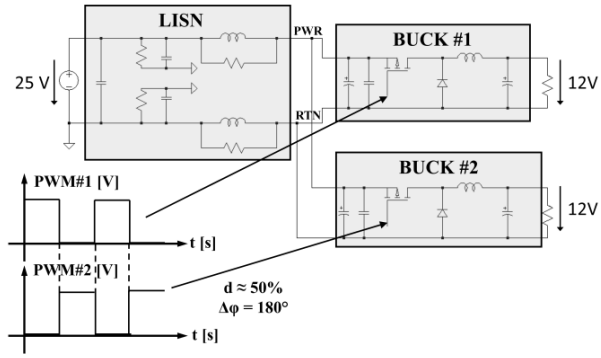


Fig. 1 LTspice Simulation block diagram

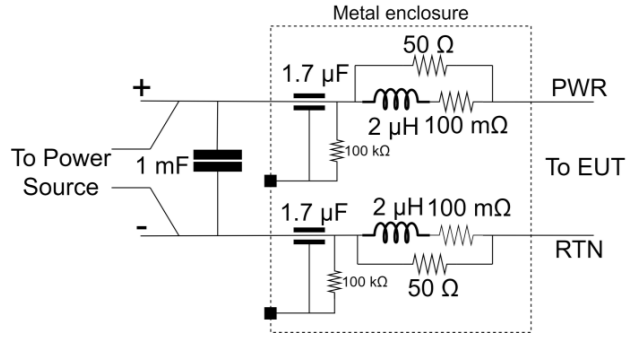


Fig. 2 LISN block diagram [7]

II. INTERLEAVING METHOD SIMULATION

The block diagram of the simulation can be seen in Fig 1. The task of the Line Impedance Stabilization Network (LISN) is to simulate the impedance of the satellite power bus and to isolate the Buck converters from external conducted disturbances [6], [7]. The LISN was made according to the Electromagnetic Compatibility (EMC) ECSS standard and its block diagram can be seen in Fig 2 [7]. The 1 mF capacitance represents the bus capacitance. The 2 µH coils and the 100 mΩ resistances simulate the series inductance and resistance of the satellite harness. At higher frequencies the two 1.7 µF capacitances will shunt the 100 kΩ resistors, and the two 50 Ω resistors will be in series, making their net resistance 100 Ω, which is the typical characteristic impedance of the satellite harness.

DOI: 10.36244/ICJ.2022.2.4

In the simulations the converter Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs) were driven by two 50 kHz square signals, the duty cycle of both signals were ~ 50 %. In Fig 1. we can see that the phase delay between the signals is 180°. In case of a signal with 50 kHz frequency (20 µs time period) this phase delay translates to 10 µs time delay. The input bus voltage of the converters was 25 V, their output voltage was 12 V and their max output power was 48 W. In the simulation the output current of the converters were set to the same value.

To characterize the behavior of the system, the input currents of the converters and their summarized current were observed in time domain and then were converted into spectrum with Fast-Fourier Transform (FFT) algorithm [8]. In Fig 3 and 4 we can see the simulation results. On top there is the summarized input current ripple of the converters and below is the current spectrum. From the figures we can see that the noise component on the 50 kHz switching frequency is completely diminished and the signal frequency is 100 kHz, the double of the original [9]. The main components of the spectrum are the upper harmonics of this doubled frequency. In this simulation the value of the choke inductances and the driving signal duty cycles were completely identical. These are ideal conditions because the components would have a finite tolerance. Also, in case of individually controlled converters the duty cycles would not exactly match. If we repeat the same simulation but with 1% choke inductance value tolerance and with slightly unmatched converter duty cycles, then we will get the results in Fig 5 and 6 [1].

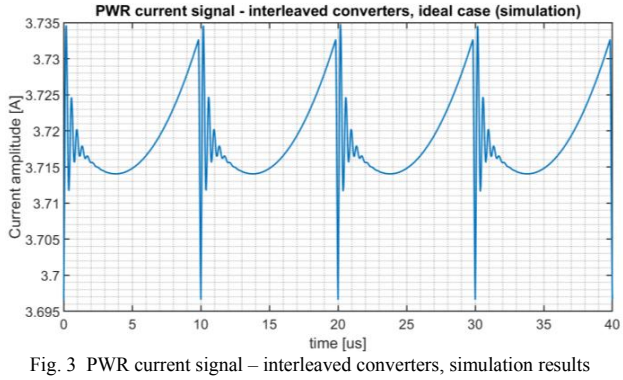


Fig. 3 PWR current signal – interleaved converters, simulation results

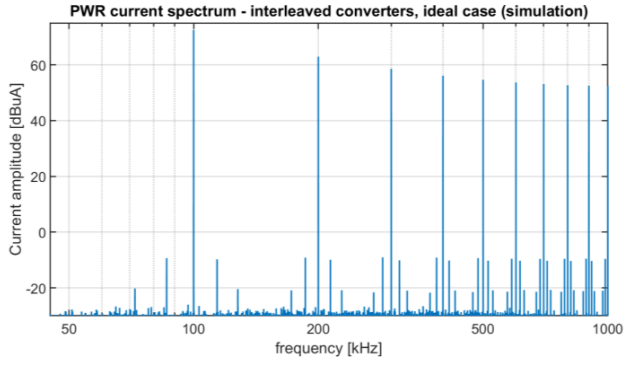


Fig. 4 PWR current amplitude spectrum – interleaved converters, simulation results

If more than 50 ns delay is added to the 10 µs time difference, then the input current ripples of the converters will not be in anti-phase in every period. This extra delay can be attributed to the improper phase-shift caused by the different turn-on times of the switches [10]. In Fig 5. we can see that the current ripple cancellation occurs only every 20 µs. The effects of this extra delay may be mitigated by differential mode (DM) filtering applied between the Power (PWR) and Return (RTN) (see Fig. 1) lines. In Fig 6. we can see, that due to the 20 µs time period the noise appears on the 50 kHz switching frequency and on its upper harmonics. From the measurements we will get similar results in time-domain as well as in current spectrum.

It is possible to set the time difference between the driving signals to zero ($\Delta\varphi = 0°$). In Fig 7. the resulting (PWR) line current signal can be seen. In this case the converter switch-on times are synchronized and their input current ripples are summarized in-phase on the PWR line.

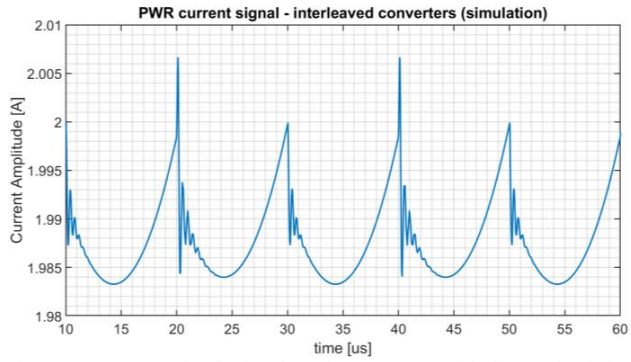


Fig. 5 PWR current signal – interleaved converters, nonideal case, simulation results

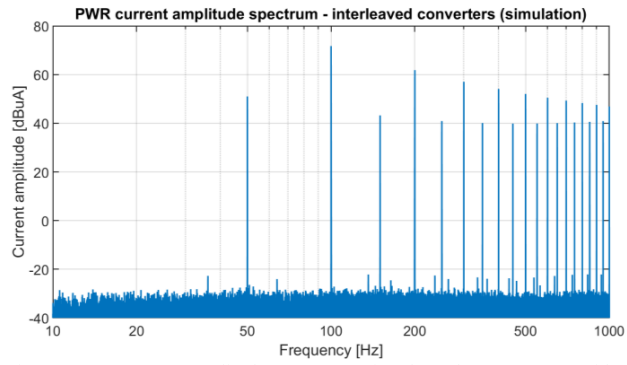


Fig. 6 PWR current amplitude spectrum – interleaved converters, nonideal case, simulation results

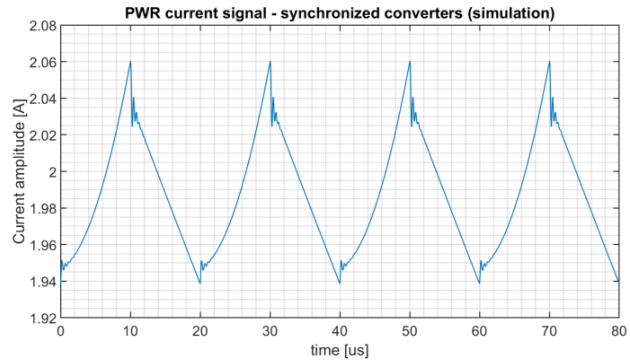


Fig. 7 PWR current signal – synchronized converters, simulation results

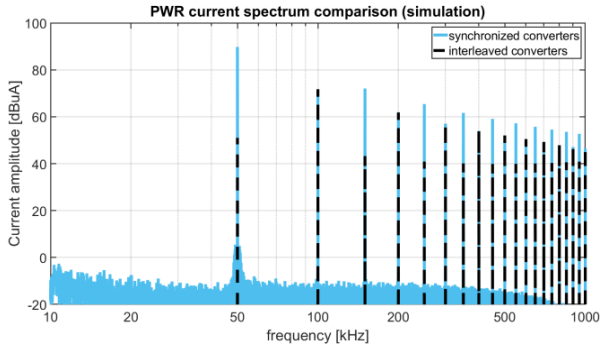


Fig. 8 PWR current spectrum comparison, simulation results

In Fig 8. we can compare the amplitude spectrum of the synchronized and interleaved converters. From this figure we can conclude that the noise component on the 50 kHz and on its odd upper harmonics (150 kHz, 250 kHz, 350 kHz…) is greatly reduced, especially on the switching frequency. The even upper harmonic components (100 kHz, 200 kHz, 300 kHz…) are amplified in case of the interleaved converters but due to the filtering of the LISN they mostly remain the same as in the case of synchronized converters. In these simulations we wanted to examine the effects of the interleaving directly and because of this we did not apply any filtering on the PWR and (RTN) lines excluding the LISN. In the next chapter we are going to present an EMC measurement setup with the same structure and the measurements described above will be conducted.

III. INTERLEAVING METHOD MEASUREMENT

For the measurements two Buck converters and a dedicated signal generator circuit was designed and built with the specification mentioned in the previous chapter. The block diagram in Fig 9. shows how the circuits are connected. For the regulation of the converters LM3524 Pulse Width Modulation (PWM) [11] control ICs were selected. The IC oscillators were synchronized by the outputs of the signal generator. With this setup we can set the phase angles of the converter switch-on times. The two major possibilities are the interleaving (φ2 = 180°) and the synchronization (φ2 = 0°) of the converters, but the phase delay can be set between these two values (see SYNC 2 in Fig 9.).

In Fig 10 we can see the entire EMC measurement setup. The LISN in this setup has the same structure as the one used in the simulations. The current signal of the PWR line was measured with a DC current probe, which was connected to an Agilent oscilloscope. The measurements were taken in time domain and later they were converted into spectrum with the FFT function of the oscilloscope [8]. The signal generator was fed from a different source and as it can be seen from Fig 9. the outputs were galvanically separated from the converters. In Fig 11. the assembled EMC measurement setup is shown.

The conducted emission of a switching mode power supply (SMPS) can be broken up into two major components: differential mode noise and common mode (CM) noise. The EMC ECSS standard contains regulations for both [7]. We are going to use these regulations to compare the measured current amplitude spectrum. The spectrum was examined only between DC and 1 MHz due to the limitations of the oscilloscope, but an SMPS with a 50 kHz switching frequency can have noise components in a 100 MHz bandwidth. However, as we have seen previously from the simulations it is possible to make estimations from these measurements too [9].

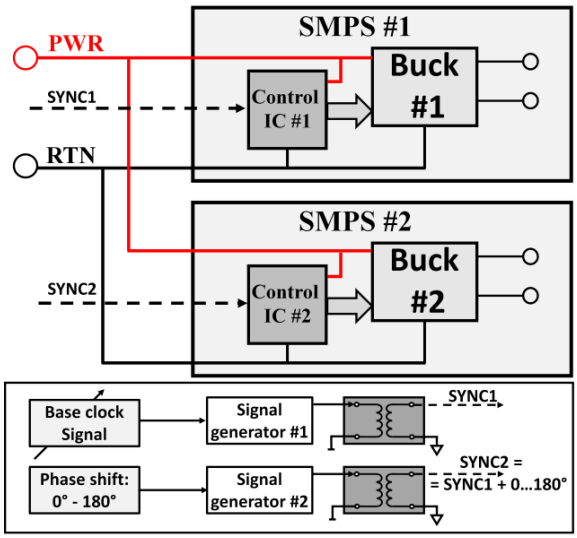


Fig. 9 Block diagram: Buck converters with the signal generator

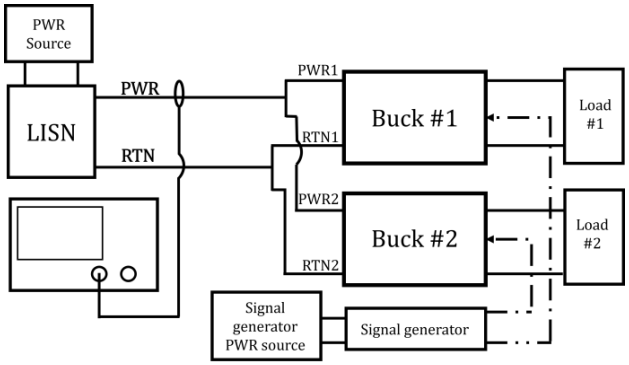


Fig. 10 Block diagram: EMC measurement setup

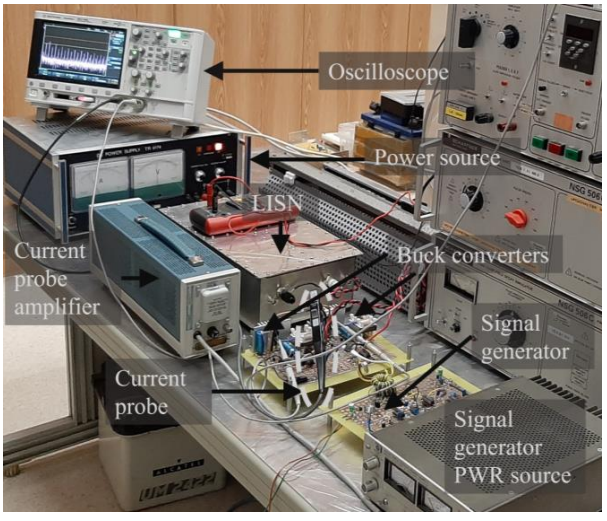


Fig. 11 EMC measurement setup assembled

The current amplitude spectrum measured on the PWR line can be seen in Fig 12. The spectrum was measured in case of interleaved and synchronized converters. From the measurements we can come to the same conclusion, as in the case of simulations: the noise on the 50 kHz switching frequency and on its odd upper harmonics are greatly reduced and the even harmonic components are mostly on the same level as in the synchronized case. These statements are depicted in more detail with the column diagrams in Fig 13 and Fig 14.

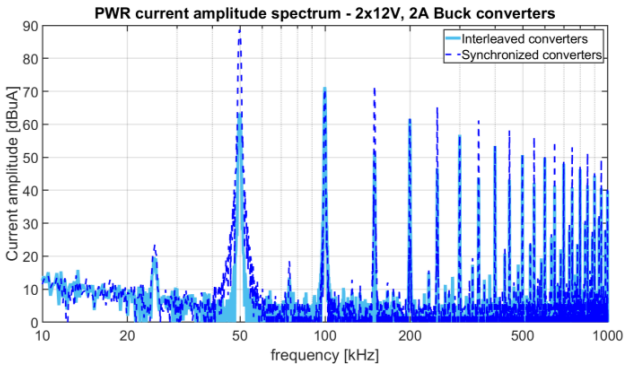


Fig. 12 Current amplitude spectrum measured on the PWR line

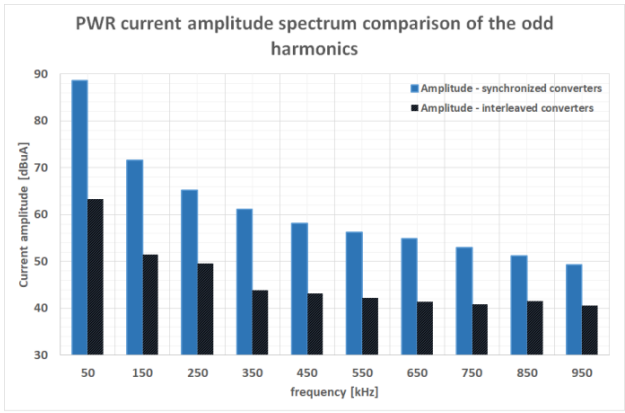


Fig. 13 Measured current amplitude spectrum comparison of the odd harmonics

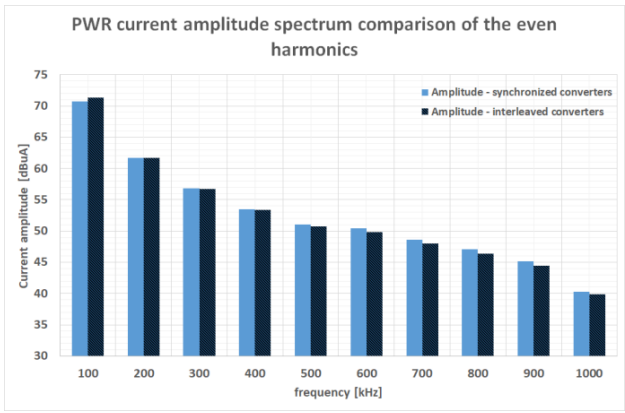


Fig. 14 Measured current amplitude spectrum comparison of the even harmonics

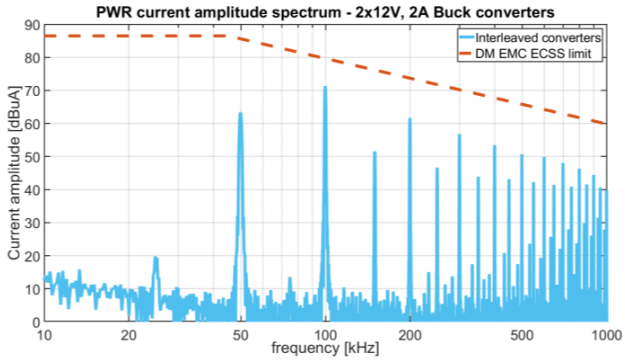


Fig. 15 Measured PWR current amplitude spectrum compared to differential mode limit (ECSS standard [7], p. 107, Fig. A-1)

The odd harmonic differences in Fig 13. are getting smaller as the frequency increases, because the amplitude of the components decreases exponentially in both control methods. The highest attenuation (25 dB) is achieved on the switching frequency, and on the 19th (950 kHz) component is around 9 dB. In Fig. 15. we compare the amplitude spectrum measured on the PWR line in case of interleaved converters against the differential mode limit issued in the ECSS EMC standard [7]. The measured PWR current spectrum is below the permitted level. The noise on the even harmonic components can be further filtered out with a capacitor between the PWR and RTN lines. Due to the strong attenuation of the switching frequency (and its odd harmonics), it is sufficient to filter the even harmonic components, which can be managed with a smaller capacitor value.

IV. Conclusion

In this article we examined the noise reduction effect of the interleaving method in case of a system with two DC-DC Buck converters. From the measurements we concluded that the switching frequency component and its odd harmonics can be decreased greatly, but not entirely. In the simulations we could recreate similar results by taking into account conditions, which make the systems non-ideal: component value tolerance, non-matching driving signal duty cycles and improper phase-shift. However, there are other factors, which may influence the noise reduction of the interleaving method. On-board of a satellite the payloads often require different voltages to function, but here we only studied two Buck converters with the same output voltages. If the output voltages are different, then the duty cycles and the peak-to-peak values of the input current ripples will be different. In such systems we can optimize the phase delay of the converters to improve the noise reduction of the interleaving [3].

REFERENCES

[1] P. Zumel, O. Garcia, J. A. Cobos and J. Uceda, "EMI reduction by interleaving of power converters," *19th Annual IEEE Applied Power Electronics Conf and Expo, 2004. APEC '04.*, 2004, pp. 688-694 vol.2, DOI: 10.1109/APEC.2004.1295894.

[2] M. Schuck and R. C. N. Pilawa-Podgurski, "Input current ripple reduction through interleaving in single-supply multiple-output dc-dc converters" *2013 IEEE 14th Workshop on Control and Modeling for Power Electronics (COMPEL)*, 2013, pp. 1-5, DOI: 10.1109/COMPEL.2013.6626436.

[3] J. Weinstein, S. C. Huerta, Z. Zhao, J. Mincey and Z. Moussaoui, "Optimized phase positioning for minimizing input filter requirements in single-input multiple-output DC-DC switch-mode power supplies" 2014 IEEE Applied Power Electronics Conf and Expo - APEC '14, 2014, pp. 455-459, DOI: 10.1109/APEC.2014.6803347.

[4] P. Takacs., "Conducted emissions measurement of on-board interleaved DC-DC converters", BSc thesis, Departure of Broadband Infocommunications and Electromagnetic Theory, Budapest University of Technology and Economics, Budapest, Hungary, 2021

[5] LTspice XVII, Analog Devices, www.analog.com

[6] A. Fazekas, "EMC analysis of the charge density meter for satellite on-board application (ESEO-LMP)" BSc thesis, Departure of Broadband Infocommunication and Electromagnetic Theory., Budapest University of Technology and Economics., Budapest, Hungary, 2016.

[7] Electromagnetic Compatibility, ECSS-E-ST 20-07C-Rev2, 2022

[8] D. Bellan, F. Marliani, S. A. Pignari and G. Spadacini, "Spectral analysis of conducted emissions of DC/DC converters" 2010 IEEE International Symp on EMC, 2010, pp. 490-494, DOI: 10.1109/ISEMC.2010.5711324.

[9] S. Ozeri, D. Shmilovitz, S. Singer and L. Martinez-Salamero, "The Mathematical Foundation of Distributed Interleaved Systems," in *IEEE Trans on Circuits and Systems*, vol. 54, no. 3, pp. 610-619, March 2007, DOI: 10.1109/TCSI.2006.886001.

[10] P. D. Antoszczuk, R. G. Retegui, N. Wassinger, S. Maestri, M. Funes and M. Benedetti, "Characterization of Steady-State Current Ripple in Interleaved Power Converters Under Inductance Mismatches" in *IEEE Transactions on Power Electronics*, vol. 29, no. 4, pp. 1840-1849, April 2014, DOI: 10.1109/TPEL.2013.2270005

[11] "LM2524D/3524D Regulating Pulse Width Modulator datasheet (Rev. E)", Texas Instrument, www.ti.com

Tamás Kőnig is currently a PhD student at the Department of Broadband Infocommunications and Electromagnetic Theory at the Budapest University of Technology and Economics (BME). He completed his MSc studies at BME in the Laboratory of Space Technology and his thesis topic was the design and building of a high-voltage quasi-resonant switching-mode power supply. He participated in the development of the Langmuir Probe experiment and the Power Distribution Unit of the ESEO satellite (2017-2018). He is conducting his PhD studies at the Laboratory of Space Technology. His research interest are the EMC examination of switching-mode power supplies and electrical power systems for space applications.

Lajos Nagy received the Engineer option Communication and PhD degrees, both from the Budapest University of Technology and Economics (BME), Budapest, Hungary, in 1986 and 1995, respectively. He joined the Department of Microwave Telecommunications (now Broadband Infocommunications and Electromagnetic Theory) in 1986, where he is currently an associate professor. He has been the head of Department of Broadband Infocommunications and Electromagnetic Theory in 2007. He is a lecturer on graduate and postgraduate courses at BME on Antennas and radiowave propagation, Radio system design, Adaptive antenna systems and Computer programming. His research interests include antenna analysis and computer aided design, electromagnetic theory, radiowave propagation, communication electronics, signal processing and digital antenna array beamforming, topics, where he has produced more than 100 different book chapters and peer-reviewed journal and conference papers. Member of Scientific Association for Infocommunications, official Hungarian Member and Hungarian Committee Secretary of URSI, Chair of the IEEE Chapter AP/ComSoc/ED/MTT.

A Novel Time Series Representation Approach for Dimensionality Reduction

Mohammad Bawaneh and Vilmos Simon

***Abstract*—With the growth of streaming data from many domains such as transportation, finance, weather, etc, there has been a surge in interest in time series data mining. With this growth and massive amounts of time series data, time series representation has become essential for reducing dimensionality to overcome the available memory constraints. Moreover, time series data mining processes include similarity search and learning of historical data tasks. These tasks require high computation time, which can be reduced by reducing the data dimensionality. This paper proposes a novel time series representation called Adaptive Simulated Annealing Representation (ASAR). ASAR considers the time series representation as an optimization problem with the objective of preserving the time series shape and reducing the dimensionality. ASAR looks for the instances in the raw time series that can represent the local trends and neglect the rest. The Simulated Annealing optimization algorithm is adapted in this paper to fulfill the objective mentioned above. We compare ASAR to three well-known representation approaches from the literature. The experimental results have shown that ASAR achieved the highest reduction in the dimensions. Moreover, it has been shown that using the ASAR representation, the data mining process is accelerated the most. The ASAR has also been tested in terms of preserving the shape and the information of the time series by performing One Nearest Neighbor (1-NN) classification and K-means clustering, which assures its ability to preserve them by outperforming the competing approaches in the K-means task and achieving close accuracy in the 1-NN classification task.**

***Index Terms*—Time Series Representation. Time Series Segmentation, Big Data, Dimensionality Reduction, Time Series Analysis.**

I. INTRODUCTION

Nowadays, owing to the rapid advancement of the core technologies of data acquisition including the cloud data centers, cell towers, and personal computers and smartphones, notably with the emerging of the Internet of Things (IoT) technology which automates the process of data collecting and storing, massive amounts of data are being stored continuously for future data mining tasks,which could contribute to the sustainable development goals (including Good Health, Sustainable Cities, and Economic Growth) [1, 2]. The amount of available data, either created, consumed, or stored, was estimated at 4.4 zettabytes in 2013, reaching 64.2 zettabytes in 2020, and is expected to reach more than 180 zettabytes in 2025 [3, 4, 5]. Recently, Wu et al. [6] have studied the relation between greening and big data by introducing the issues of big data from the greening point of view. They have identified three main domains which require greening. First, big data acquisitions necessitate significant energy consumption for data collecting as well as data transfer through networks. Second, storing massive data has called for more advanced technologies that are inefficient in terms of energy and resources. Third, the process of analytics of big data is usually computationally expensive, consuming time, energy, and resources. As a result, a dimensionality reduction technique can contribute to greening big data storage and analytics by conserving storage space while also reducing the computational complexity of the data analytics process.

A significant amount of the generated data are streaming data which is also known as time series data. Time series is a sequence of observations, where each observation is recorded sequentially with time [7]. Time Series data are used in various domains including finance and stock market [8, 9], voice recognition [10], online signature verification [11], failure prediction in high performance computing and cloud systems [12], earthquake forecasting[13], weather prediction [14], and intelligent transportation systems [15]. Consequently, an enormous amount of data are generated daily and requires special memory management. As previously stated, such massive data has two major consequences. First, a significant quantity of memory must be provisioned, consuming energy and resources. Second, because of the inherited high computation complexity, processing and analyzing high-dimensional data is challenging, making it difficult to analyze the time series in its raw form. To achieve that, many researchers have investigated time series representation approaches, with various ways offered to minimize time series high dimensionality by expressing the time series in a new representation form in a lower dimension space [16]. However, a common key concept for applying valuable time series representation is that the new representation of the time series must include the original characteristic features in order to preserve the important information of the raw time series (such as local trends information and basic data distribution). Furthermore, these features must be acquired while keeping the new representation as simple as possible. Moreover, because time series data comes from various domains and represents distinct behaviors, the representation approach should be applicable to numerous types of time series datasets. Therefore, the time series representation approach should be general and applicable to any dataset to be used as a preprocessing step. As a result of these transformation criteria, storage space will be saved and further processing and analysis of data will be accelerated. In this paper, we adopt these criteria to propose an effective offline time series representation approach termed

Mohammad Bawaneh is with the Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary, e-mail: mbawaneh@hit.bme.hu.

Vilmos Simon is with the Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest Uni- versity of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary, e-mail: svilmos@hit.bme.hu.

DOI: 10.36244/ICJ.2022.2.5

Adaptive Simulated Annealing Representation (ASAR). The proposed approach treats the time series representation as an optimization problem, with the aim of retaining the time series shape while lowering dimensionality. Moreover, because it is focused on tracking the local trends of the time series, the proposed approach is able to transform any sort of time series data with diverse characteristics and behavior.

Transforming the time series into a new representation has several advantages. When it comes to extracting information from time series data, several data mining tasks, such as classification and clustering [17, 18], may be used to analyze the time series data. As a consequence, with the requirement to measure similarity and examine historical time series data in order to apply effective classification or clustering tasks, transforming the time series as a preprocessing step would give minimal computing complexity and hence speedier results. Furthermore, certain similarity metrics may get skewed due to the distortion in the raw time series. As a result, changing the time series while retaining its fundamental characteristic features overcome this issue as well [19].

Time series representation methods that have been proposed in the literature have several flaws; we will discuss them in detail in the next section. Some of these methods transform the time series into symbolic form and by this lose the original structure, which makes it impossible to restore the shape of the time series. In addition, some methods lose the local trend information, which is crucial information for similarity measuring of time series data. Some variants of these have been proposed to include the trend information; however, this comes with a cost of insufficient compression ratio, which is considered one of the main objectives when representing time series in a new form. In this paper, ASAR is proposed to overcome the shortcomings of these methods by introducing a shape-based representation of time series. ASAR keeps the new form of the time series as simple as possible by transforming the data into a lower dimension but with the same shape as the raw time series together with the same data distribution. This way, it addresses the issue of keeping the original structure while compressing the data. Moreover, by preserving the shape of the time series, the local trend information is preserved, with no cost of including additional information in the new representation. The proposed approach ASAR is assessed and compared with some approaches from the literature in this paper by measuring the Compression Ratio (CN) to determine which approach saves the most memory. Furthermore, classification and clustering tasks are used to assess ASAR's capacity to maintain time series information (i.e., such as the local trends) and to demonstrate the process acceleration feature. The following are the key contributions of this paper:

- The new representation of time series ASAR can significantly reduce dimensionality while retaining the shape of the time series. This conserves storage space without losing the information required for future data mining operations. Moreover, the high compression ratio that can be achieved by ASAR accelerates future data mining operations.
- The ASAR approach views time series representation as an optimization issue with the objective of maintaining the raw time series shape. This is achieved by tracking local trends in the raw time series and expressing these trends by the least number of segments. As a result, ASAR has no restrictions on the type, shape, distribution, or source domain of time series data.

This paper is organized as follows. Section 2 includes a review of prior similar studies from the literature. The proposed approach is explained in detail in Section 3. Section 4 presents the findings of the experimental analysis. Finally, in section 5, the paper's conclusion is provided.

II. RELATED WORKS

In the last two decades, the applications domains that apply time series analysis have grown tremendously. In addition, the rapid advancement in data acquisition technologies offered an enormous amount of data which in turn could be mined to form significant knowledge. As a consequence, numerous time series representation methods were developed to overcome the challenges of the data's high dimensionality [20, 16]. Aghabozorgi et al. [19] classified the time series representation methods into four main categories: data adaptive, non-data adaptive, model-based, and data dictated representation methods. This section provides a brief overview of these categories and the most significant approaches presented in the previous two decades.

In data adaptive methods, the segmentation of the time series is done with varied length segments. Singular Value Decomposition (SVD) was one of the earliest methods proposed in time series dimensionality reduction [21]. It can be used to represent multivariate time series data. SVD deals with the multivariate time series as an $(m*n)$ matrix. It applies a space rotation process to the best least-squares fit direction by factorizing the matrix into three other matrices ($A = U\Sigma V^T$). U is $m*m$ unitary matrix, Σ is $m*n$ rectangular diagonal matrix with diagonal non-negative elements called singular values, and V^T is $n*n$ unitary matrix. The dimension of the matrix is reduced by removing the least significant singular values in Σ and the corresponding entries in U and V^T. The disadvantage of SVD is that it has high computation complexity $O(mn^2)$. Years later, the Adaptive Piecewise Constant Approximation (APCA) was proposed [22]. APCA segments the time series into constant segments but with varying lengths. The new representation is simply the records of the endpoints for each segment with the mean value of the segment in the original raw time series. With a computation complexity of $O(n)$, APCA is a faster method than SVD. However, a significant disadvantage of APCA is that it loses the trend information since two segments with two different trends may have the same mean values. Gullo et al. [23] have proposed the Derivative time series Segment Approximation (DSA) representation model. DSA model transforms the raw time series into the derivative estimation by computing the first derivative of each sample. Then it segments the derivative estimation into variable-length segments, where the breaking criterion is that the points that have close slopes (close first derivative values) are in the same segment. In other words, the segment keeps expanding while

the absolute difference between the new sample and the mean value of the previous samples within the segment is less than a certain threshold. Finally, the new representation is formed by pairs representing the segments. Each pair consists of the timestamp of the last point in the segment, and an angle demonstrates the average slope of this segment.

Non-data adaptive methods segment the time series with fixed-length segments. One of the widely used time series representation methods under this category is the symbolic representation called the Symbolic Aggregate approXimation (SAX) [24, 25]. SAX normalizes the time series to a zero mean distribution and standard deviation of 1, keeping the different time series within the same offset. Then the time series is transformed into the Piecewise Aggregate Approximation (PAA) representation [26], which in turn reduces the dimensionality. PAA divides the time series into a number of equal-sized frames. Then for each frame, the mean value of the points within the frame is calculated, and finally, the sequence of the mean values of all frames will be the new PAA representation. As a result of the normalizing process, the time series follows a Gaussian distribution. In the next step, the authors divide the time series into equal-sized areas under the curve of the Gaussian distribution (the same size as the PAA representation's frame). Finally, they assign a symbol for each area which will be later assigned for all samples within this area. Based on the sequence values obtained by the PAA representation, the time series is transformed into a sequence of symbols called a word. Similar to APCA, SAX has a drawback of losing the trend information since segments have different trends but similar slope values will be assigned by similar symbols. There are several variants that have been proposed as SAX extensions. Lkhagva et al. [27] have proposed to use the minimum and maximum values within the segment in addition to the mean value to overcome the drawback of SAX. However, this will triple the dimension reduced by SAX. Another Variation is proposed by Sun et al. [28] in their SAX-TD method. SAX-TD adds the trend information of each segment to the SAX representation by calculating the distance between the segment's ending points which they called the trend distance. Consequently, the dimension is double that reduced by SAX. Another extension, SAX with Standard Deviation (SAX_SD), has been proposed [29]. The authors improved SAX by adding the standard deviation feature in addition to the mean value in order to study the spread of the values within the segment and to improve the similarity measure. In [30], Multivariate Symbolic Aggregate Approximation (MSAX) was proposed to represent multivariate time series data. Some applications contain more than one variable explaining the same behavior. Therefore, MSAX integrates the information of the different time series in one symbolic representation. MSAX first checks the dependency between the variables. If they are independent of each other, the data are normalized. However, in the case of dependent variables, a linear transformation must be applied. Then, all the time series in the matrix are represented using the PAA method. Last, discretization is applied resulting in a symbol matrix. As a final step, the symbols in the matrix are transformed into a sequence of symbols with a length equal to the columns, where each entry is represented by compressing the symbols in all rows (all the time series) in the corresponding column.

Model-based representation methods transform the time series stochastically. Time Series Bitmaps belongs to this category [31]. Time Series Bitmaps uses the time series extracted features and their frequencies to color a Bitmap. This visualization of the similarities between time series offers the users a fast discovery of the clusters, classes, anomalies, and other shape-based tasks. This is done by first transforming the continuous time series into discrete time series by applying SAX. Then, the frequencies of the sub-words in the SAX representation are counted, where the desired level of recursion defines the length of the sub-word. These frequencies are mapped into the corresponding pixel of the grid, where the grid contains pixels that represent all possible sub-words based on the desired level. The frequencies are normalized by dividing them by the largest value to handle the length variety between the time series. The final step is the color mapping of these frequencies into the grid, which offers the ability to compare the time series. It is not recommended to use bitmaps representation for a single time series as it does not offer any information. Another drawback of Bitmaps is that the structure of the raw time series is hidden and cannot be captured.

In data dictated methods, the compression ratio is not defined in advance where it is dependent on the raw time series behavior. The Clipped representation is an example of this category [32, 33]. Clipped represents the time series as binary values. The raw time series' samples above the population's mean will be represented by 1, whereas those below the mean will be represented by 0. The new binary representation is compressed to a new sequence that contains the lengths of the subsequences with the same value. It is unnecessary to mention the sample value in addition to the length as a pair because it is a binary representation. Hence, including the first value is enough where the rest of the values will be only toggling between 0 and 1. Zhan et al. [34] proposed the Feature-based Clipped Representation (FCR). FCR divides the time series into equal-length segments. Then it finds the trends' turning points within each segment and their corresponding importance indices using the method presented in [35]. The turning points are then chosen based on their importance and converted into binary values using the clipped representation, which will be compressed to a new sequence that contains the lengths of the subsequences with the same value. The clipped representation here compares the values to the segment's mean instead of the population's mean. Another example of this category is the symbolic representation of the Fragment Alignment Distance (FAD) method [36]. FAD estimates the derivative of time series using the DSA method [23]. This derivative estimation contains the trend information. After that, FAD converts this derivative sequence into a symbolic sequence R by setting a threshold and comparing it with the derivative estimation value of each sample. If the value is less than the threshold, the point has a small change compared to the previous point, and they will be assigned with the same symbol. However, if the value is bigger, the point has a big change compared to the previous point, and so a different

symbol will be assigned for this point. Finally, FAD transforms the resulted symbolic representation series R into feature series consisting of pairs of values. Each pair represents the symbol of a similar subsequence and the length of this subsequence. Another method that belongs to this category was proposed in the paper [37] which is called Adaptive Particle Swarm Optimization Segmentation (APSOS). APSOS deals with the time series segmentation as an optimization problem. The goal of the optimization is to minimize the error function between the raw time series and the segmented time series. To find the samples that best segment the series, they have adapted the particle swarm optimization algorithm to find the best segments' endpoints. APSOS is able to capture the trend information of the time series; however, it has high computation complexity $O(n^2)$, which makes it difficult to use with the high daily acquired streaming data.

The proposed approach in this paper is part of the data dictated methods since it is based on tracking the local trends in the raw-time series, and consequently, the compression ratio is dependent on the time series behavior. It is inspired by the APSOS approach by dealing with the time series segmentation as an optimization problem. The following section introduces the proposed approach in detail.

III. Adaptive Simulated Annealing Representation (ASAR)

A brief summary of the significant approaches proposed in the literature for representing time series was introduced in the previous section. These approaches suffer from different drawbacks. Some approaches are time-consuming due to the high computational complexity required to create the representation of the raw time series. On the other hand, some of those solutions with low computational complexity failed to capture the local trends information. Furthermore, some approaches do not offer a high enough compression ratio, where the high compression ratio is one of the crucial features of a time series representation; therefore, it became the key objective of our work. The Adaptive Simulated Annealing Representation (ASAR) is introduced in this paper to overcome these issues. ASAR's objective is to represent the time series in a new form to achieve a high compression ratio, this way saving the storage space and at the same time preserving the shape of the time series, which will keep the essential features and prevent information loss. Inspiring by the APSOS approach [37], ASAR deals with the time series representation as an optimization problem. This optimization aims to find the instances in the raw time series that can describe the shape in the possible best way, ignoring the rest of the instances. In the following subsection, we define the time series representation as an optimization problem.

A. Formulating Time Series Segmentation as an Optimization Problem

Each time series contains several local trends, forming a time series shape. For example, two time series may have the same shape, which means that they follow the same local trends. However, the time of occurrence of the local trends does not have to be the same. As mentioned earlier, ASAR is proposed to reduce the time series dimensions while maintaining the time series shape. For this purpose, a heuristic algorithm can be utilized. Heuristic algorithms are optimization algorithms that can find an approximated optimum global value for a particular function. Accordingly, in order to use a heuristic algorithm to apply time series representation, the time series representation must be formulated first as an optimization problem with the objective of reducing the time series dimensions while preserving the shape. Let us assume that X is a time series of length n and is defined as:

$$X = \{X_1, X_2, ..., X_n\}$$

Our target is to find a new time series R, representing X time series shape with a reduced dimensionality. The new representation R can be defined as follows:

$$R = \{R_1, R_2, ..., R_k\} \tag{1}$$

where $k \ll n$. To illustrate, Figure 1 shows the objective of the proposed approach using a synthetic time series. The length of the raw time series (depicted by the blue line) is 1000, whereas it can be reduced to 22 samples while preserving the shape of the raw time series (the orange line). It must be noted that this is just an illustrative example of the approach's objective, not the result of the ASAR's transformation. The segment from the time series R is defined

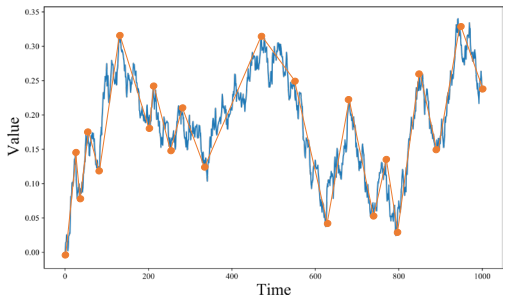


Fig. (1) An illustration example of the time series segmentation result, note that the blue line represents the raw time series, while the orange one represents the new time series representation.

as the line connecting two consecutive points in the new representation. Hence, R will contain $k-1$ segments. This segment is obtained by recording two timestamps from the raw time series as endpoints and neglecting the timestamps between them. However, the segment may still be used to estimate the value for each timestamp of the raw time series (even the neglected ones). This estimation can be specified by the line equation (the line connecting the two endpoints). Let us assume that RX represents a time series for the estimated values of the raw time series from the point of view of the new representation R, Then, the RX_i's approximate corresponding value of X_i can be computed as follows:

$$RX_i = \frac{1}{(e-s)}[(i-s)X_e + (e-i)X_s] \tag{2}$$

where i is the index of the point that will be estimated, s is the starting point of the segment where i locate, and e is the endpoint of this segment (s,i, and e are timestamps from the raw time series). The proof is as follows:

Based on the straight line equation:

$$RX_i = m \times i + b \tag{3}$$

where m is the slope of the line, and b is the y-intercept. The slope m is calculated as follow:

$$m = \frac{(X_e - X_s)}{(e - s)} \tag{4}$$

now substitute 4 in 3:

$$RX_i = \left(\frac{(X_e - X_s)}{(e - s)}\right) \times i + b \tag{5}$$

Hence, the y-intercept b can be defined as:

$$b = RX_i - \left(\frac{(X_e - X_s)}{(e - s)}\right) \times i \tag{6}$$

by using the point e:

$$b = X_e - \left(\frac{(X_e - X_s)}{(e - s)}\right) \times e \tag{7}$$

now substitute 7 in 5:

$$RX_i = \left(\frac{(X_e - X_s)}{(e - s)}\right) \times i + X_e - \left(\frac{(X_e - X_s)}{(e - s)}\right) \times e \tag{8}$$

by expanding out the brackets we get equation 2:

$$RX_i = \frac{1}{(e - s)}[(i - s) \times X_e + (e - i) \times X_s]$$

The approximating values ($R_i \quad \forall\ s < i < e$) will be used to calculate the Mean of Squared Errors (MSE) for each segment between the raw time series and the corresponding approximate values in the new representation. MSE is the average of the squared errors and is defined as:

$$MSE(s, e) = \frac{1}{e - s + 1} \times \sum_{i=s}^{e} (X_i - RX_i)^2 \tag{9}$$

MSE serves as an indicator of how much the segment that has a starting point s and an endpoint e aligns the samples in the range (s, e) from the raw time series. The MSE values then will be compared to indicate the superiority between different possible segments to represent the data. In other words, ASAR computes the MSE for different segments (same starting point but different endpoints) to find the best segment that has the minimum MSE among them, which points to the best alignment between the new segment and the corresponding samples from the raw time series. For simplifying the implementation and for the purpose of direct calculations, equation 2 is substituted to equation 9, which gives us the following formula:

$$MSE(s, e) = \frac{1}{e - s + 1} \times \sum_{i=s}^{e} \left(X_i - \frac{(i - s)X_e}{(e - s)} - \frac{(e - i)X_s}{(e - s)}\right)^2 \tag{10}$$

The proposed ASAR algorithm aims to segment the raw time series based on the local trends so that the points that follow the same trend will be covered by one segment. However, once the trend changes significantly, a new segment should be used to cover the next points in the raw time series. Moreover, since different segments can represent a trend, the selected segment should be the best-aligned one with the points that follows the same trend. From this point of view, we have defined our objective function to find the best segment's endpoint as follows:

$$e = \underset{state}{\operatorname{argmin}}\ MSE(s, state) \tag{11}$$

In the first segment, the starting point will be the first point in the raw time series. So basically, this objective function searches for the endpoint of the segment. Once the optimal endpoint is found and recorded, our method starts to search for the next endpoint, considering the previous endpoint as the new segment's starting point.

B. Simulated Annealing for ASAR

The Simulated Annealing (SA) heuristic algorithm [38] is utilized in our method to find the new representation of the time series which can maintain its shape. SA is chosen as the heuristic algorithm in this paper due to its ability to overcome the issue of being stuck at some local optima during the global optimum solution search process. In addition, SA is a robust and general algorithm as it makes no constraints on the type of data. This allows for proposing a general time series representation algorithm that can be applied for streaming data coming from various domains. Moreover, in case the time for the searching process is due to end, SA returns the best-known solution. In other words, there is always a best-known solution even if the time was not enough to complete the search process.

The general principle of SA and its use in this paper can be clarified as follows: the computational optimization by SA is a probabilistic technique with the objective of finding the global optimum value for a particular function within a solution search space. The algorithm starts searching for the optimum value by moving between the possible solutions randomly. However, each transition is evaluated, and only those who have a high transition probability will be considered. The probability is controlled by two parameters, a change in the system Energy ΔE, and a system temperature T. The solutions resulting in smaller system energy are better than those solutions with greater energy. Therefore, SA accepts the solutions that result in smaller system energy. However, worse solutions may still get accepted with a certain probability controlled by the change in the system energy and temperature. The system temperature is used to reduce the likelihood of accepting worse solutions as the solution search space is investigated. Accepting a worse solution provides a more thorough search for the ideal global solution. Simulated Annealing operates in the following manner. The temperature gradually declines from a positive starting point to zero. At each time step, the Simulated Annealing randomly picks a solution similar to the current one, evaluates its quality, and progresses to it based on the probability of picking better or worse solutions.

The time series representation has been formulated as an optimization problem in order to apply the Simulated Annealing algorithm (see equation 11). The solution search space has been adapted in this study to cover only a part of the raw time series rather than the entire space. In other words, because the time series contains several local trends, only a window of the raw time series is inspected each time in order to locate the best segment. As a result, the window of the raw time series is investigated each time to locate the optimal endpoint that depicts a local trend. The discovered endpoint is then utilized as the starting point for the next segment search, with the same window size serving as the search space. As the entire solution space (the entire raw time series) is not required in each segment inquiry, this reduces the computing complexity of ASAR. In addition, for the reason that any local trend consists of consecutive samples, the transition between states in the Simulated Annealing is made incremental rather than random in this paper. It should be emphasized that the window size does not determine the size of the segments but rather the size of the solution search space (i.e., the search space for the segment endpoint). As a result, ASAR is not sensitive to this window size setting because it is utilized to reduce calculation complexity. We recommend setting a large enough window size that can suit the majority of the local trends (i.e., the window size is larger than the length of the majority of the local trends). Nonetheless, not having a large enough window size does not cause a problem in preserving the shape of the time series; however, local trends with lengths greater than the window size will be represented by more than one segment.

1) Energy Function: The Mean of Squared Errors defined in equation 10 is used as the energy function in this paper. The change in the energy function when moving from a state ($state$) to a state ($state + 1$) is defined as follows:

$$\Delta E_{state} = MSE(s, state - 1) - MSE(s, state) \quad (12)$$

2) Cooling Schedule: The other parameter which controls the acceptance probability of the new $state$ is the temperature. The temperature should decline gradually, which is controlled by the cooling schedule. In this paper, the cooling schedule is defined linearly as follows:

$$T_{state} = \alpha \times T_{state-1} \quad (13)$$

where T is the system temperature, and α is cooling parameter. In this paper, we adopt the definition provided in the paper [39] to calculate the α value.

$$\alpha = \left(\frac{T_{initial}}{T_{final}}\right)^{\left(\frac{1}{w}\right)} \quad (14)$$

where w is the window size used for the search space. The final temperature is typically set close to 0. However, the initial temperature is problem-dependent and has to be studied and set based on the use case. This definition of α ensures that the temperature starts with a high value and decreases gradually towards the low final value. In other words, at the beginning of the process, the temperature is set to be high, resulting in a high probability of accepting states. However, as the process gets through, the temperature value decreases, resulting in a lower probability of accepting solutions.

3) Acceptance Probability: The probability function used in this paper is defined as:

$$Pr_{state} = \begin{cases} 1 & \Delta E_{state} > 0 \\ e^{\frac{\Delta E_{state}}{T_{state}}} & \Delta E_{state} < 0 \end{cases} \quad (15)$$

Pr_{state} contains two different factors; the system temperature T_{state}, and the change in energy function ΔE_{state}. The probability is directly proportional to the system temperature and the change in energy. Table I explain the effects of these factors.

TABLE (I)
THE EFFECTS THAT EACH FACTOR CAN CAUSE ON THE PROBABILITY VALUE.

Factor	Magnitude	Effect on probability
T	High	High
	Low	Low
ΔE	High	High
	Low	Low

4) Acceptance Criteria: SA searches for the segment that will keep the change in the energy positive as it indicates that the new MSE is better than the previous one, meaning that it is a good move. However, in case of a bad move, i.e., the energy is negative, the SA may still make the move but with a certain probability to ensure finding the global minima where the segment's MSE reflects the best segment alignment with the raw data. Once the state is rejected, the previous state will be used to record a new value in the new representation R as it indicates the best endpoint that makes the best segment alignment. This is expressed as follows:

$$state = \begin{cases} Accepted & \Delta E_{state} > 0 \\ Accepted & Pr_{state} > rand(0,1) \\ Rejected\ (R_c = X_{state-1}) & Pr_{state} < rand(0,1) \end{cases} \quad (16)$$

where c is the timestamp of the new record in the new representation, starting from 1 and increasing by 1 with each new record until it reaches the new representation length k. To summarize, the ASAR approach searches for the points that can explain the local trends and neglects the rest to reduce the data dimensions while preserving the shape of the time series. Table II shows the ASAR algorithm's pseudocode.

IV. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

In this section, the proposed approach is compared to three significant approaches from the literature, the PAA, the SAX, and the FAD approaches. Since representation approaches are not direct methods to extract information, a validation experiment is designed in this paper to test the ability of ASAR to preserve the information of the raw time series by applying similarity search and detection tasks using time series classification and time series clustering methods. In addition, to test the effectiveness of ASAR, it is compared with the competing approaches in terms of the compression ratio in order to test the storage space saving supremacy. Moreover, the time performance of each approach for applying classification or clustering tasks is compared to demonstrate which approach accelerates the data mining process the most.

TABLE (II)
THE PSEUDO CODE IMPLEMENTATION OF THE PRO-POSED ASAR ALGORITHM.

Algorithm: Adaptive Simulated Annealing Representation (ASAR)
Input: Time Series $X = (X_1, X_2, ..., X_n)$
Output: Time Series $R = (R_1, R_2, ..., R_k)$

1. $k = 1$, $s = 1$, $state = s + 3$,$R_k = X_1$
2. For $T = T_{max}$:T_{min} do:
    If ($\Delta E_{state} > 0$):
        $state = state + 1$
    Else if ($Pr_{state} > rand(0, 1)$)
        $state = state + 1$
    Else
        Break
    End if
  End for
3. $k = k + 1$
   $R_k = X_{state-1}$
   $s = state - 1$
   Go to 2

A. Assessment algorithms

In order to evaluate the ability of ASAR to maintain the time series information, it is tested and compared with other approaches based on similarity search tasks. Standard classification and clustering algorithms are applied as the paper's objective is not the similarity search tasks themselves but the dimensionality reduction. Therefore, we chose to apply the well-known algorithms, One Nearest Neighbor (1-NN) for the classification and K-means for the clustering.

1) One Nearest Neighbor classification (1-NN): In K Nearest Neighbor (K-NN) classification [40], the tested instance is classified based on the classes of the closest k instances. In other words, the algorithm checks the class of closest k instances (using a similarity metric) and applies majority voting to predict the tested instance class. Since the classification task is not the main objective of this paper, the One Nearest Neighbor (1-NN) has been chosen in this paper as the most basic, straightforward, and standard method to check the similarity between time series. Moreover, it provides a fairer comparison as it does not require parameter tuning, leading to unbiased results. We use 50% of the dataset under study as a training dataset and 50% as a testing dataset.

2) K-means Clustering: One of the well-known and most used clustering algorithms is the K-means algorithm [19]. It is a partitioning clustering approach that separates the data into k clusters, intending to minimize the distance between the instances and the cluster center and maximize the distance between the instances from different clusters. In this paper, K-means is used to provide a more thorough analysis of the ability of ASAR to preserve the shape of the time series by testing its accuracy of clustering the data based on the time series shape. Since K-means requires setting the number of clusters k in advance, we use the actual number of the classes of the labeled dataset under study.

B. Assessment Criteria

To assess the classification and clustering results of the competing approaches, the F-measure (or F-score) [41] is used in this paper. It is an accuracy measure determined using the test's precision and recall. The precision is the number of the true predicted positives divided by the number of all predicted positives. The recall is the number of the true predicted positives divided by the number of actual positives. The F-measure is defined as follows:

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (17)$$

The compression ratio (CR) in the new representation of the time series is computed to assess the degree of dimensionality reduction achieved by the competing approaches and demonstrate the superiority in storage space savings. The compression ratio explains the proportion of the reduction in the length of the time series to the original length and can be calculated as follows:

$$CR = \frac{n - k}{n} \times 100\% \quad (18)$$

where n is the raw time series length, and k is the length of the time series in the new representation form.

C. Dataset Description

The UCR Time Series Classification Archive [42] is employed in this paper to evaluate the proposed approach. The data in the archive is z-normalized and originates from several domains, providing variety for the time series form. The experiment in this study has been applied to eight datasets. The datasets are chosen to have varying lengths and to exhibit a variety of characteristics to evaluate ASAR's ability to handle diverse types of time series. The data in the UCR repository is divided into two sets: training and testing. In this experiment, we choose the testing set since its bigger size provides for more robust results. Table III shows the dataset information used in the experiment.

TABLE (III)
THE INFORMATION OF THE SELECTED DATASETS FROM THE UCR ARCHIVE.

Name	Dataset Size	Time Series Length	Number of Classes
HandOutlines	1000	2709	2
StarLightCurves	1000	1024	3
Lightning-2	60	637	2
OSU Leaf	200	427	6
ShapeletSim	180	500	2
WormsTwoClass	180	900	2
Yoga	1000	426	2
Trace	100	275	4

D. Parameters Tuning

One or more parameters must be tuned for the competing solutions. We used the parameter sweeping technique to tune the parameters as in the original articles [36, 26, 25]. FAD requires adjusting two parameters: the threshold used to determine symbol changes and the number of symbols. The threshold is changed by 0.01 increments between 0 and 0.2, whereas the number of symbols is adjusted by one increment between 3 and 7. A threshold of 0.02 with a number of symbols of 5 has shown the best performance for the classification and the clustering tasks.

Since the time series is divided into equal-sized frames, PAA requires tuning the frame size. The frame size is adjusted by one increment between 2 and 100. A frame size of 8 has shown the best classification and clustering results.

In SAX, two parameters must be tuned: the frame size and the alphabet size (the number of symbols to represent the data). As in PAA, the frame size is adjusted by one increment between 2 and 100, while the alphabet size is adjusted by one increment between 2 and 10. The best performance has been shown using a frame size of 4 with an alphabet size of 8.

For the proposed approach ASAR, two parameters need to be tuned. The first parameter is w, the window size used for the search space. The second parameter is the initial temperature (we use 0.1 for the final temperature as it must be close to 0). We adjust w by one increment between 10 and 150. We must note that this window size does not generate equal-sized frames but rather a search space for the segment's endpoint. Regarding the initial temperature, it is adjusted by 0.5 increments between 0.5 and 5. A window size of 25 with an initial temperature of 2.5 has shown the best performance in the experiment.

E. Time Series Representation Using ASAR

As previously stated, ASAR is intended to track time series local trends. Its objective is to reduce the time series' dimensions while preserving its shape. This was accomplished by retaining the data points that best explain the time series and its trends and eliminating the others. We show an example from each dataset to demonstrate the ASAR's ability to represent time series with this objective. Figure 2 shows the raw time series of each sample on the left and its ASAR representation on the right. It can be seen that ASAR representations are totally following the raw time series shapes with much lower lengths, which confirms the fulfillment of the paper's objective.

F. Effectiveness Evaluation

The proposed approach is compared to FAD, PAA, and SAX. In this research, four distinct assessments are used. To assess ASAR's efficiency in preserving the shape and time series information, we use 1-NN classification and k-means clustering. The compression ratio (CR) is used to measure the efficiency of decreasing dimensions to conserve storage space (see equation 18). Furthermore, the efficacy of ASAR in facilitating accelerated data mining processes is presented by comparing the time required to execute classification and clustering tasks using raw time series and the competing representation techniques. To guarantee a robust outcome, the F-measure (equation 17) for 1-NN classification and k-means clustering is calculated by averaging 50 experimental runs. Likewise, the time necessary to complete these activities is averaged across 50 attempts. As mentioned in section II, ASAR belongs to the data dictated methods, which means that the compression ratio is not previously defined and fixed for all time series but is dependent on the time series behavior. Therefore, CR here indicates the dataset's average compression ratio (the average of the compression ratio for all time series in the dataset). Since the classification and the clustering tasks are not the main objectives of this paper, we calculate the relative results of the F-measure for the representation approaches and the raw time series. This allows us to see how effective is a representation method compared to the case when the time series is used in its raw form. Table IV shows the F-measure relative results for the 1-NN classification. The table shows that ASAR comes in third place after PAA and SAX and performs better than FAD.

On the other hand, Table V shows the F-measure relative results for the K-means clustering task. ASAR achieved the best performance among the competing methods. It has increased the clustering accuracy by 8%.

When it comes to the main objective of this paper, ASAR has achieved the maximum compression ratio (equation 18) among the competing representation approaches. Table VI shows the compression ratios and the average compression ratio for each representation approach as a percentage (%).

This high compression ratio accelerates the data mining process since the training dataset dimension is reduced. To show and compare the efficiency of the dimensionality reduction, we have calculated the time needed to perform the 1-NN classification and the K-means clustering tasks using the data representation provided by the various methods. The experiments in this paper have been conducted on a platform with an Intel(R) Xeon(R) Silver 4215 CPU with clock speed of 2.50GHz and 2.49 GHz (2 processors) with 8 GB RAM, running Windows 10 (64-bit). Python programming language was used to implement all the approaches. Table VII shows the performance time in seconds. It can be seen that by using ASAR, the data mining process can be accelerated the most, which proves its advantage in fast information extraction besides the storage space saving.

G. Results Discussion

According to Figure 2, it is pretty clear that ASAR is able to fulfill the objective of this paper by significantly reducing the dimensionality while preserving the time series shape. This confirms that ASAR is an applicable and reliable dimensionality reduction approach for the data mining tasks (such as time series classification, clustering, and anomaly detection) that considers the original shape of the time series as a crucial feature . All examples show clearly that the ASAR representation follows the local trends of the raw time series except for the ShapeletSim dataset due to the high frequency of trends reversal. However, tables IV and V show that using ASAR has resulted in no significant information loss in the ShapeletSim dataset by achieving an accuracy of 94% and 100% in the classification and clustering tasks, respectively.

To test the information preservation capability of the representation methods, the 1-NN classification and the K-means clustering were employed in this paper. The results for the 1-NN classification (table IV) locate ASAR in the third place among the competing approaches. The average F-measure of ASAR is 96% of the accuracy obtained by using the raw time series. This shows that ASAR preserves the time series information. On the other hand, Table V shows that ASAR has achieved the highest accuracy among the competing approaches. It outperformed the results by 8% when compared to the raw time series used for K-means clustering. While the main objective of this paper is dimensionality reduction, Table VI shows the compression ratio obtained using each

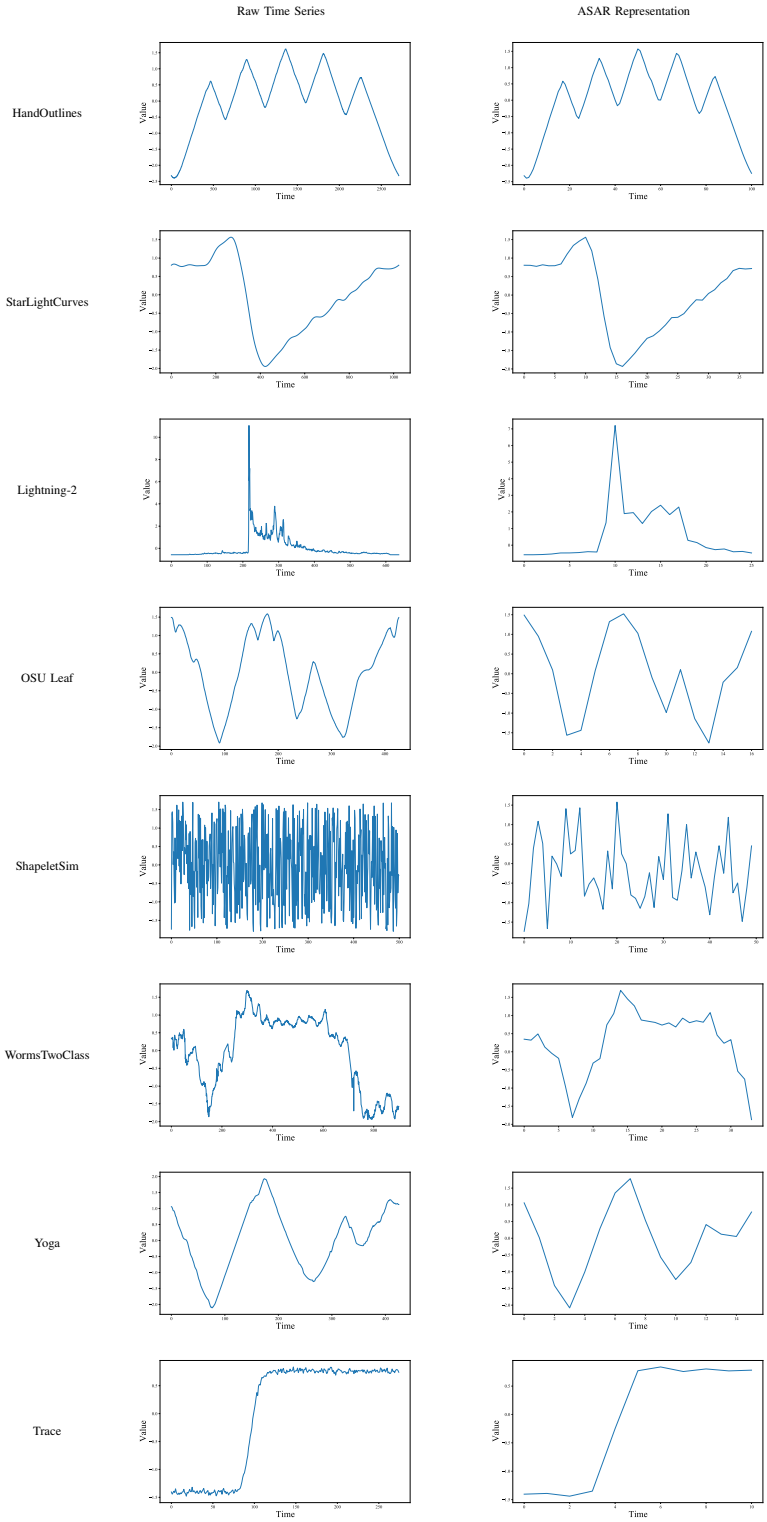


Fig. 2.: ASAR representation examples compared to the raw ones.

TABLE (IV)
F-MEASURE RELATIVE RESULTS FOR 1-NN CLASSIFICATION.

Method	Raw Time Series	FAD	PAA	SAX	ASAR
HandOutlines	1.00	0.64	1.00	1.00	0.99
StarLightCurves	1.00	0.97	1.00	1.00	1.00
Lightning-2	1.00	0.74	1.15	0.97	0.99
OSU Leaf	1.00	0.75	1.00	0.98	0.80
ShapeletSim	1.00	1.12	1.16	1.20	0.94
WormsTwoClass	1.00	0.92	1.00	1.02	0.98
Yoga	1.00	0.97	1.00	0.99	0.82
Trace	1.00	0.83	1.06	1.02	1.19
Average	1.00	0.87	1.05	1.02	0.96

TABLE (V)
F-MEASURE RELATIVE RESULTS FOR K-MEANS CLUSTERING.

Method	Raw Time Series	FAD	PAA	SAX	ASAR
HandOutlines	1.00	0.86	1.14	1.29	1.21
StarLightCurves	1.00	1.33	1.08	1.25	1.33
Lightning-2	1.00	0.93	0.96	0.85	0.89
OSU Leaf	1.00	0.94	0.88	0.94	1.00
ShapeletSim	1.00	1.02	1.02	0.98	1.00
WormsTwoClass	1.00	0.91	1.00	1.05	1.00
Yoga	1.00	0.93	0.82	0.93	1.04
Trace	1.00	0.90	1.15	0.90	1.15
Average	1.00	0.98	1.01	1.02	1.08

TABLE (VI)
THE COMPRESSION RATIO AS A PERCENTAGE (%).

Method	FAD	PAA	SAX	ASAR
HandOutlines	99.8	87.5	75	96.3
StarLightCurves	98.1	87.5	75	96.2
Lightning-2	4.9	87.4	74.9	94.8
OSU Leaf	56	87.4	74.9	95.5
ShapeletSim	0	87.4	75	88.2
WormsTwoClass	17.5	87.5	75	95.5
Yoga	52.7	87.3	74.9	95.8
Trace	10	87.3	74.9	95.4
Average	42.4	87.4	75	94.7

TABLE (VII)
THE TIME NEEDED (IN SECONDS) TO PERFORM 1-NN CLASSIFICATION AND K-MEANS CLUSTERING USING THE DATA REPRESENTED BY FAD, PAA, SAX, AND ASAR.

Method	Raw Time Series	FAD	PAA	SAX	ASAR
1-NN Classification	11.7	11.5	3.3	4.3	2.3
K-means Clustering	11.9	11.2	3.5	4.4	2.5

representation approach. With an average compression ratio of 94.7%, ASAR surpasses the competing approaches in reducing the dimensionality of the time series. PAA and SAX come in second and third places, respectively. However, FAD shows an unstable compression ratio among the datasets due to its high reliance on the data behavior, resulting in not guaranteed dimensionality reduction. Furthermore, the data mining tasks acceleration feature was demonstrated by measuring the time needed to apply the 1-NN classification and the K-means clustering. Table VII shows that these tasks have been performed the fastest by using the ASAR representation. ASAR took around 70% of the time needed to perform the tasks when compared to the PAA representation, which comes in second place. It is even less for others. Furthermore, it took around 20% of the time needed to perform these tasks using the time series in its raw form.

In summary, ASAR has achieved the best results in terms of the compression ratio. In addition, it has achieved the best results in terms of the data mining process acceleration and the K-means clustering. Moreover, even though it comes in third place in terms of the 1-NN classification, it has achieved a pretty good result, which shows the ability to preserve the time series information. This is also proved by ASAR's superiority in the K-means clustering results. These results were accomplished independently of the data behavior and domain, which provides extra advantage for ASAR as it can be used without any constraints on the type of data. The latter feature was proved by employing data from different domains with a diversity of behaviors and time series lengths (see Table III and Figure 2).

V. CONCLUSION AND FUTURE WORKS

In this paper, a novel time series representation approach has been proposed. The Adaptive Simulated Annealing Representation (ASAR) approach treats the time series representation as an optimization problem. Its objective is to segment the time series based on the tendencies by recording the instances that best explain the tendencies and neglecting the rest. By tracking the tendencies in the time series, ASAR was able to transform the time series into new dimensions while preserving the shape and the information. An experiment was designed to test its ability of maintaining the information, reducing dimensionality, and accelerating the data mining process. The experimental results have shown that ASAR outperforms FAD, PAA, and SAX approaches in terms of the compression ratio, the time needed to perform 1-NN classification and K-means clustering, and in the K-means clustering accuracy. It has also achieved high accuracy results in the 1-NN classification. These results assure that ASAR is able to conserve storage space and accelerate the data mining process while preserving the shape and the information of the time series. In addition, the experimental results have shown that ASAR is independent of the data type, behavior, domain, or length.

Some domains provide multivariate time series data (such as speed, flow, and occupancy in the intelligent transportation systems domain), which usually describe the same process. A possible future work is to extend ASAR to represent these multivariate time series data in a unified representation. Another possible future work is to define a new similarity measure based on the ASAR representation by utilizing the shape-preserving feature. Since this similarity measure will be tailored to the ASAR representation, this extension could achieve better results for the data mining tasks than using other similarity measures such as Euclidean distance or Dynamic Time Warping.

ACKNOWLEDGEMENT

The research reported in this paper and carried out at the Budapest University of Technology and Economics has been supported by the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology.

REFERENCES

[1] Q.-T. Doan, A. Kayes, W. Rahayu, and K. Nguyen, "Integration of iot streaming data with efficient indexing andstorage optimization," *IEEE Access*, vol. 8, pp. 47 456–47 467, 2020, DOI: 10.1109/ACCESS.2020.2980006.

[2] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389–2406, 2018, DOI: 10.1109/COMST.2018.2812301.

[3] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," *Journal of business research*, vol. 69, no. 2, pp. 897–904, 2016, DOI: 10.1016/j.jbusres.2015.07.001.

[4] D. R.-J. G.-J. Rydning, "The digitization of the world from edge to core," *Framingham: International Data Corporation*, p. 16, 2018.

[5] A. Holst, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025," https://www.statista.com/statistics/871513/worldwide-data-created/, 2021, accessed: 2021-06-30.

[6] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: Greening big data," *IEEE Systems Journal,* vol. 10, no. 3, pp. 873–887, 2016, DOI: 10.1109/JSYST.2016.2550538.

[7] C. Chatfield, *Time-series forecasting*. CRC press, 2000.

[8] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, "Deep learning for financial applications: A survey," *Applied Soft Computing*, p. 106384, 2020, DOI: 10.1016/j.asoc.2020.106384.

[9] S. M. Idrees, M. A. Alam, and P. Agarwal, "A prediction approach for stock market volatility based on time series data," *IEEE Access*, vol. 7, pp. 17 287–17 298, 2019, DOI: 10.1109/ACCESS.2019.2895252.

[10] M. G. Tulics and K. Vicsi, "Automatic classification possibilities of the voices of children with dysphonia," *Infocommunications Journal,* vol. 10, no. 3, pp. 30–36, 2018, DOI: 10.36244/ICJ.2018.3.5.

[11] M. Saleem and B. Kovari, "Online signature verification using signature down-sampling and signer-dependent sampling frequency," *Neural Computing and Applications*, pp. 1–13, 2021, doi: DOI: 10.1007/s00521-021-06536-z.

[12] B. Mohammed, I. Awan, H. Ugail, and M. Younas, "Failure prediction using machine learning in a virtualised hpc system and application," *Cluster Computing*, vol. 22, no. 2, pp. 471–485, 2019, DOI: 10.1007/s10586-019-02917-1.

[13] A. Bhatia, S. Pasari, and A. Mehta, "Earthquake forecasting using artificial neural networks," *Int. Arch. Photogram. Rem. Sens. Spatial. Inform. Sci.*, pp. 823–827, 2018, DOI: 10.5194/isprs-archives-XLII-5-823-2018.

[14] E. Soares, P. Costa Jr, B. Costa, and D. Leite, "Ensemble of evolving data clouds and fuzzy models for weather time series prediction," *Applied Soft Computing*, vol. 64, pp. 445–453, 2018, DOI: 10.1016/j.asoc.2017.12.032.

[15] A. M. Nagy, B. Wiandt, and V. Simon, "Transient- based automatic incident detection method for intelligent transport systems," *Infocommunications Journal*, vol. 13, no. 3, pp. 2–13, 2021, DOI: 10.36244/ICJ.2021.3.1.

[16] S. J. Wilson, "Data representation for time series data mining: time domain approaches," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 9, no. 1, p. e1392, 2017, DOI: 10.1002/wics.1392.

[17] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.- A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery,* vol. 33, no. 4, pp. 917–963, 2019, DOI: 10.1007/s10618-019-00619-1.

[18] A. Belhadi, Y. Djenouri, K. Nørvåg, H. Ramampiaro, F. Masseglia, and J. C.-W. Lin, "Space–time series clustering: Algorithms, taxonomy, and case study on urban smart cities," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103857, 2020, DOI: 10.1016/j.engappai.2020.103857.

[19] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering–a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015, DOI: 10.1016/j.is.2015.04.007.

[20] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Mining and Knowledge Discovery,* vol. 26, no. 2, pp. 275–309, 2013, DOI: 10.1007/s10618-012-0250-5.

[21] D. Wu, A. Singh, D. Agrawal, A. El Abbadi, and T. R. Smith, "Efficient retrieval for browsing large image databases," in *Proceedings of the fifth international conference on Information and knowledge management,* 1996, pp. 11–18.

[22] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Locally adaptive dimensionality reduction for indexing large time series databases," in *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, 2001, pp. 151–162, DOI: 10.1145/375663.375680.

[23] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco, "A time series representation model for accurate and fast similarity detection," *Pattern Recognition*, vol. 42, no. 11, pp. 2998–3014, 2009, DOI: 10.1016/j.patcog.2009.03.030.

[24] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003, pp. 2–11, DOI: 10.1145/882082.882086.

[25] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: a novel symbolic representation of time series," *Data Mining and knowledge discovery*, vol. 15, no. 2, pp. 107– 144, 2007, DOI: 10.1007/s10618-007-0064-z.

[26] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[27] B. Lkhagva, Y. Suzuki, and K. Kawagoe, "New time series data representation esax for financial applications," in *22nd International Conference on Data Engineering Workshops (ICDEW'06)*. IEEE, 2006, pp. x115–x115, DOI: 10.1109/ICDEW.2006.99.

[28] Y. Sun, J. Li, J. Liu, B. Sun, and C. Chow, "An improvement of symbolic aggregate approximation distance measure for time series," *Neurocomputing*, vol. 138, pp. 189–198, 2014, DOI: 10.1016/j.neucom.2014.01.045.

[29] C. T. Zan and H. Yamana, "An improved symbolic aggre- gate approximation distance measure based on its statistical features," in *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*, 2016, pp. 72–80, DOI: 10.1145/3011141.3011146.

[30] A. M. Nagy and V. Simon, "A novel data representation method for smart cities' big data," in *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities: Designing for Sustainability.* Springer International Publishing, 2022, pp. 97–122, DOI: 10.1007/978-3-030-84459-2_6.

[31] N. Kumar, V. N. Lolla, E. Keogh, S. Lonardi, C. A. Ratanamahatana, and L. Wei, "Time-series bitmaps: a practical visualization tool for working with large time series databases," in *Proceedings of the 2005 SIAM international conference on data mining. SIAM*, 2005, pp. 531–535, DOI: 10.1137/1.9781611972757.55.

[32] C. Ratanamahatana, E. Keogh, A. J. Bagnall, and S. Lonardi, "A novel bit level time series representation with implication of similarity search and clustering," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2005, pp. 771–777.

[33] A. Bagnall, E. Keogh, S. Lonardi, G. Janacek et al., "A bit level representation for time series data mining with shape based similarity," *Data Mining and Knowledge Discovery,* vol. 13, no. 1, pp. 11–40, 2006, DOI: 10.1007/s10618-005-0028-0.

[34] P. Zhan, H. Xu, and L. Chen, "Fcad: Feature-based clipped representation for time series anomaly detection," in *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE, 2020, pp. 206–210, DOI: 10.1109/ICISCAE51034.2020.9236862.

[35] P. Zhan, C. Sun, Y. Hu, W. Luo, J. Zheng, and X. Li, "Feature-based online representation algorithm for streaming time series similarity search," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 34, no. 05, p. 2050010, 2020, DOI: 10.1142/S021800142050010X.

[36] M. Zhang and D. Pi, "A new time series representation model and corresponding similarity measure for fast and accurate similarity detection," *IEEE Access*, vol. 5, pp. 24503–24519, 2017, DOI: 10.1109/ACCESS.2017.2764633.

[37] H. Kamalzadeh, A. Ahmadi, and S. Mansour, "A shapebased adaptive segmentation of time-series using particle swarm optimization," *Information Systems,* vol. 67, pp. 1–18, 2017, DOI: 10.1016/j.is.2017.03.004.

[38] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983, DOI: 10.1126/science.220.4598.671.

[39] S. Sakamoto, K. Ozera, A. Barolli, M. Ikeda, L. Barolli, and M. Takizawa, "Implementation of an intelligent hybrid simulation systems for wmns based on particle swarm optimization and simulated annealing: performance evaluation for different replacement methods," *Soft Computing*, vol. 23, no. 9, pp. 3029–3035, 2019, DOI: 10.1007/s00500-017-2948-1.

[40] Y.-H. Lee, C.-P. Wei, T.-H. Cheng, and C.-T. Yang, "Nearest-neighbor-based approach to time-series classification," *Decision Support Systems*, vol. 53, no. 1, pp. 207–217, 2012, DOI: 10.1016/j.dss.2011.12.014.

[41] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006, DOI: 10.1016/j.patrec.2005.10.010.

[42] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr time series classification archive," https://www.cs.ucr.edu/~eamonn/time_series_data/, July 2015.

Mohammad Bawaneh received his M.Sc. degree in Computer engineering from Yarmouk University (YU)-Jordan in 2017. He is currently a Ph.D. candidate at Budapest University of Technology and Economics – Department of Networked Systems and Services (BME-HIT) in the Multimedia Networks and Services Laboratory (MEDIANETS). His research interests include time series data mining and analysis, and machine learning and data analytics for smart cities and intelligent transportation management systems.

Vilmos Simon received his PhD from the Budapest University of Technology and Economics (BME) in 2009. Currently he is an Associate Professor at the Department of Networked Systems and Services and Head of the Multimedia Networks and Services Laboratory. He has done research on mobility management and energy efficiency in mobile cellular systems and self-organized mobile networks, recently his research interests include machine learning and data analytics for smart cities and intelligent transportation management systems. He published 50+ papers in international journals and conferences, and acts as a reviewer or organizer for numerous scientific conferences. He serves currently as the Corporate liaison vice president for the Connected and Automated Mobility Cluster of Zala.

Wide Band Spectrum Monitoring System from 30MHz to 1800MHz with limited Size, Weight and Power Consumption by MRC-100 Satellite

Yasir Ahmed Idris Humad and Levente Dudás

***Abstract*—Today, the usage of radio frequencies is steadily increasing based on the continuous development of modern telecommunication technologies, and this, in turn, increases the electromagnetic pollution not only on Earth but also in space. In low Earth orbit, electromagnetic pollution creates some kind of difficulty in controlling nano-satellites. So it is necessary to measure the electromagnetic pollution in the Low Earth Orbit. The basic aim of this paper is to present the capability of designing and developing a PocketQube-class satellite 3-PQ 5 x 5 x 15 cm as a potential continuation of SMOG-1, the fourth satellite of Hungary. The planned scientific payload of MRC- 100 is a wideband spectrum monitoring system for radio frequency smog in the frequency range of 30-2600 MHz on Low Earth Orbit (600 Km). In this paper, we have executed qualifying measurements on the whole system in the frequency range of 30-1800 MHz (first phase), and we calibrated its broadband antenna with a measurement system. We present the capabilities of the wideband spectrum monitoring system to measure radio frequency signals, with the limited size, weight, and power consumption of the designed system. The working spectrum measurement system was tested on the top of the roof of building V1 at BME University and An-echoic chamber, we were able to show that there is significant radio frequency smog caused by the upper HF band, FM band, VHF band, UHF band, LTE band, GSM band, 4G band, and UMTS band. This is relevant to the main mission target of MRC-100.**

***Index Terms*—Educational Student Satellite, PocketQube, Radio frequency Smog, Spectrum Monitoring System.**

I. INTRODUCTION

STUDENT satellite development, which is now common in a variety of applications, is categorized as nano-satellites. Due to their small sizes, minimal costs, and shorter manufacturing times, student satellites have shown to be a great alternative for large satellites in a variety of applications including space exploration. Nano-satellites, as opposed to traditional space missions, rely on commercial off-the-shelf (COTS) components which reduce prices and speed development. Typically, the term ”nano-satellite” refers to satellites sized between 1 and 10 kg. The modern proposed class is the Pocket Qube Satellite, which restricts developers to a volume of approximately 5 x 5 x 5 cm for a single unit and a mass range of 0.1 to 1 kilogram. Many Pocket Qube Satellite experiments can be found in the Microwave Remote Sensing Laboratory at the Department of Broadband Infocommunications and Electromagnetic Theory at BME University. [12]. [8]

MRC-100 is a 3-PQ (PocketQube) class student satellite $(5 \times 5 \times 15)cm$ with a total mass of 750 grams. SMOG-1 is 1-PQ $(5 \times 5 \times 5)cm$ with 175 g total mass. SMOG-P was the first and smallest operational satellite in the world during its lifetime. MRC-100 is considered a possible continuation of SMOG-P, ATL-1, and SMOG-1, which were created by the students of BME university and are classified as second, third, and fourth-class Hungarian PocketetQube satellites that are integrated into the BME educational system [1]. [6]. [10]. [9]. First, the hardware used for the measurements is presented. MRC-100 main subsystems are: COM, EPS, OBC, and **SP**. The 3D model of MRC-100 student satellite can be seen in Fig. 1.

The proposed trajectory for the MRC-100 is a polar, circular, and sun-synchronous Low Earth Orbit with a distance of 600 km apogee and perigee. The planned launch will be in December 2022.

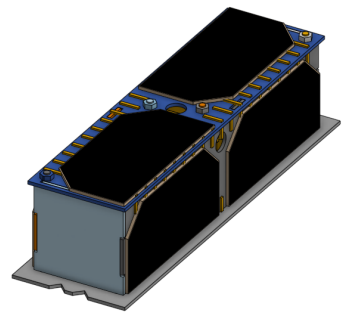


Fig. 1.: Three dimensional view of MRC-100.

Yasir Ahmed Idris Humad is a Ph.D student at the dept. of Broadband Infocommunications and Electromagnetic Theory, faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics (BME), Budapest, Hungary. (e-mail: yasirahmedidris.humad@edu.bme.hu)

As a supervisor, Levente Dudás Ph.D. is a communicational and system engineer for the MRC-100 student satellite project at the dept. of Broadband Infocommunications and Electromagnetic Theory, faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics (BME). (e-mail: dudas.levente@vik.bme.hu)

DOI: 10.36244/ICJ.2022.2.6

The distance between Sun and Earth is about 150 million km. Around the Earth, the averaged power density equals to $1360\frac{W}{m^2}$ [2]. [11]. MRC-100 will be covered by solar panels made by AzurSpace [3]: 8 pieces of three-layers 80 mm $\times$ 40 mm cells.

Because of the atmosphere, the solar power density on Earth's surface is $1000\frac{W}{m^2}$ (due to the ozone layer). A 15 cm satellite has a 1.088 W incoming power. The three-layer solar cells of the MRC-100 have a 28 percent efficiency and a $40mm \times 80mm$ dimension, resulting in a DC (Direct Current) power of 0.8 W. MRC- 100's LEO lasts 100 minutes, with 60 minutes in the light and 40 minutes in the dark. As a result, due to the MRC-100's essentially random movement, while orbiting around 800 mW, the average DC input is 0.68 W, with a 1.7 W peak (on LEO DC input will be 36% more). On-board systems on the MRC-100 will have single-point- failure tolerant and cold-redundant.

The peak power of 1.7 W is estimated using the three-layer solar cell dimension $(80 \times 40)mm$ and the solar cell cut-off edge$(13.5 \times 13.5)mm$ for 1U cube $(100 \times 100)mm$ as seen in Eq.(1) - (5).

$$\frac{(80 \times 40 - 13.5 \times 13.5) \cdot 2}{100 \times 100} = 60\% \tag{1}$$

The solar power density on Earth's surface is 1000 $\frac{W}{m^2}$, for 10 cm^2 cube it equals to 10 $\frac{W}{cm^2}$

$$\text{The total D.C power } = 10\ W \times 60\% = 6\ \text{W} \tag{2}$$

$$\text{The peak D.C input } = 6\ W \times 28.5\% = 1.71\ \text{W} \tag{3}$$

$$\text{Average D.C power } = 1.71\ W \cdot \frac{4\ sides\ of\ cube}{6\ sides\ of\ cube} = 1.14\ W \tag{4}$$

$$\text{Average D.C input } = 1.14\ W \times 60\% = 0.684\ \frac{W}{90\,\text{min}} \tag{5}$$

The 3-PQ electrical power system is referred to as the EPS. The 3-PQ surfaces is covered by 8 pieces of solar panels, which serve as a source of energy. The electrical power system's main responsibilities are to independently control the operating point of the solar panels in order to achieve maximum input DC power, charge the Lithium-ion accumulator in order to work on the dark side of the Earth, and provide a stable +3.3 V power supply voltage to all subsystems.

The on-board computer is referred to as the OBC. The OBC is in control of the operation of the on-board subsystems, such as SP (spectrum monitoring system) and COM (communication system). The OBC is responsible for data collecting and handling on-board (these will be the Communication radiated telemetry data).

The COM (communication system) is responsible for establishing a two-way data link between MRC-100 and the ground stations. Because of the size of the onboard antenna, which should be opened from the cube, this radio transmission is on the 70 cm UHF (radio amateur band).

The SP (spectrum monitoring system) is the MRC-100 main payload. The mission of this monitoring system is to monitor the (upper HF band, FM band, VHF band, UHF band, 5G band, GSM band, and UMTS band) radiated from the Earth in (LEO) orbit. As a result, this power is considered as a lost power.

II. MRC-100 PAYLOADS

There are several payloads on-board the satellite: **spectrum analyzer (30 - 2600 MHz)**, active magnetic attitude control, horizon + Sun camera, GPS + LoRa downlink (satellite identification), 1 Mbit/s S-band down-link, total ionizing dose measurement system, automatic identification system receiver for vessel traffic services, UHF-band LoRa-GPS Tracking, memory based single event detector, special thermal insulator test. The 3D model of MRC-100 student satellite subsystems can be seen in Fig. 2.

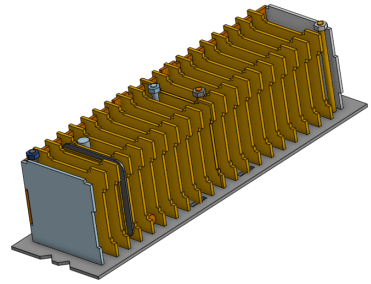


Fig. 2.: Three dimensional view of MRC-100 Subsystems.

III. THE MAIN PAYLOAD OF MRC-100 SATELLITE

The main payload of MRC-100 will be a wide band spectrum monitoring system (SP). the spectrum monitoring is a single-chip radio transceiver (in receiver mode) from Silicon Laboratories SI 4464. This receiver is working from 119 to 960 MHz with 1 to 850 KHz bandwidth. [4]. The monitored frequency band is 30 - 1800 MHz. The block scheme of the single-chip transceiver (SI4464) is in Fig. 3.

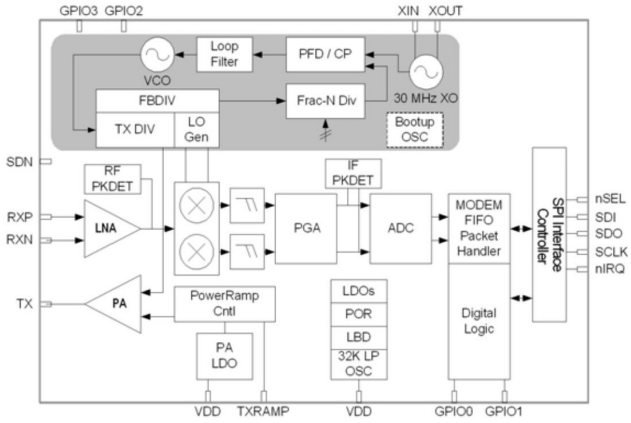


Fig. 3.: Single-chip radio transceiver as a spectrum monitor. [4].

The receiver part of this transceiver is a conventional superhetrodyne receiver with digital IF (Intermediate Frequency) unit. The SI4464 chip contains a wide-range fractional-PLL (Phase Locked Loop) as a local oscillator, a wide-band LNA (Low Noise Amplifier) and mixer, PGA (Programmable Gain Amplifier) as IF amplifier with low pass filter, I-Q ADCs(Analogue to Digital Converters) and digital OOK-FSK MODEM (MOdulator and DEModulator). [4]. [5]. In the case of the spectrum monitor, the mission is to tune the carrier frequency of the receiver from 119 to 960 MHz and read the RSSI (Received Signal Strength Indicator) register of the receiver chip. Fig. 4. Shows the Locked frequencies and the frequencies gab of the local oscillator with fractional-PLL (Phase Locked Loop).

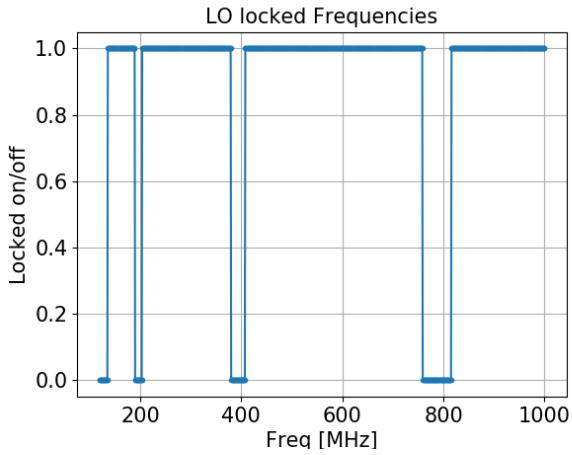


Fig. 4.: SI1060 as Local Oscillator.

IV. WIDEBAND SPECTRUM ANALYZER

The main payload of MRC-100 will be a wide band spectrum analyzer $40mm \times 40mm$ dimension, on the upper HF band, FM band, VHF band, UHF band, 5G band, GSM band and UMTS band (30 to 1800) MHz frequency range.

The spectrum monitoring system of MRC-100 is based on a RF microcontroller from Silabs SI1060. It contains a C8051F930 micro-controller in a single Quad Flat No-Lead (QFN) package and a SI4464 digital radio module. Fig. 5. shows the block diagram of the MRC-100's spectrum monitoring system.

The receiver side of the above mentioned transceiver (SI4464) can operate in RF scanning mode. At a given frequency and bandwidth, the SI4464 chip measures the RSSI (Received Signal Strength Indicator) level with enough dynamic range and 1 dB accuracy. According to the working frequency range of the receiver part (119-960)MHz, this wideband (30-1800)MHz must be divided into sub-bands.

In the RF Scanning mode, the communication antenna is first connected to the RF BPF (Band Pass Filter), then the RF signal passed through the RF switch, next to a LNA and the active mixer, after that it will pass through the IF/ BPF (Band Pass Filter) and a LNA and finally to the receiver input.

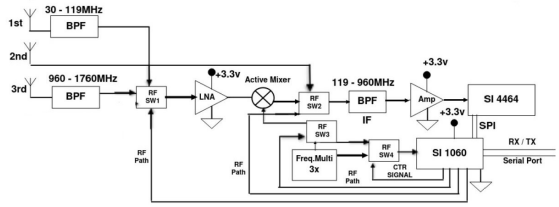


Fig. 5.: Block Diagram of Wideband Spectrum Monitoring System.

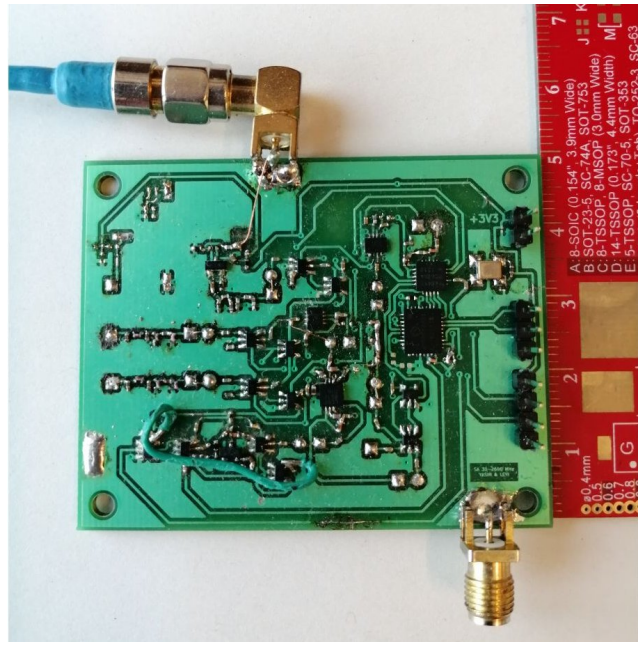
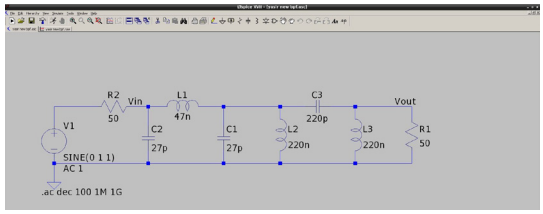


Fig. 6: The experimental model of a Wide Band spectrum monitoring system.

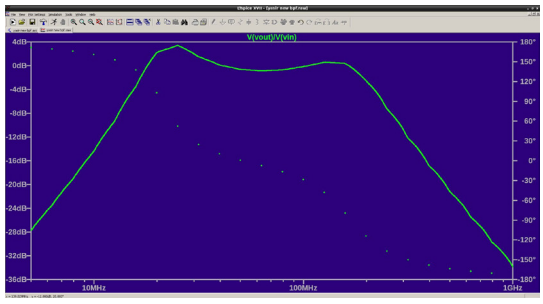
Several electronic components are present: the SMA connector, TCXOs (temperature compensated crystal oscillators) as a reference oscillator, and (SI4464) as a radio transceiver chip is on the right up. the RF switches, an active mixer, and a LNA (Gain Block Ampilifier) in the middle. The monopole antenna connection for the three sub-bands in the left and the frequency multiplier in the left down. as shown in Fig. 6.

A. First Band 30-119 MHz

In the case of the first band, the task is to adjust the receiver carrier frequency from 119 to 960 MHz and read the values of the RSSI register of the receiver chip (Received Signal Strength Indicator), but the first band frequencies are lower than the range of the receiver chip. So by tuning the local oscillator at 820MHz to up-convert the received signal to the receiver chip frequency range. the monopole antenna is connected directly to the RF BPF (Band Pass Filter) controlled by the first RF Switch stage in order to filter the first band signals. as shown in Fig. 7.



(a) First Band Filter Circuit

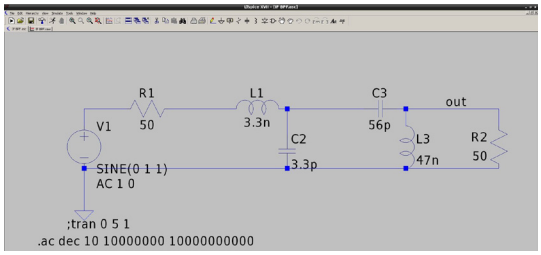


(b) First Band Filter

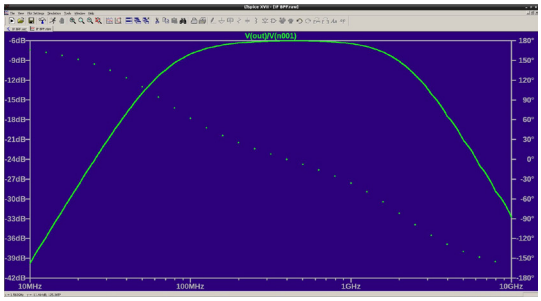
Fig. 7.: 2 Figures of First Band Pass Filter.

B. Second Band 119-960 MHz

The task in the second band is to adjust the receiver carrier frequency from 119 to 960 MHz and read the values of the RSSI register of the receiver chip (Received Signal Strength Indicator), in the case of the second band the monopole antenna is connected directly to the IF BPF (Band Pass Filter) controlled by the second RF Switch stage in order to filtered the second band signals, because the second band's frequencies are within the receiver chip's range. as shown in Fig. 8.



(a) Second Band Filter Circuit

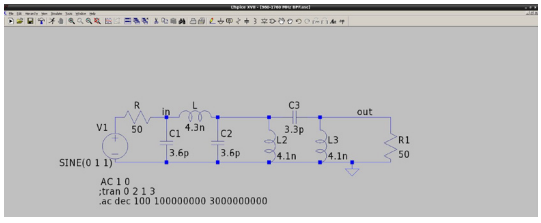


(b) Second Band Filter

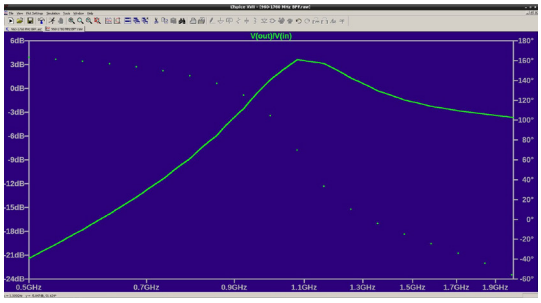
Fig. 8.: 2 Figures of Second Band Pass Filter.

C. Third Band 960-1800 MHz

In the case of the third band, the task is to adjust the receiver carrier frequency from 119 to 960 MHz and read the values of the RSSI register of the receiver chip (Received Signal Strength Indicator), but the third band frequencies are higher than the receiver chip's range . So by tuning the local oscillator at 841MHz to down-convert the received signals to the receiver chip frequency range. the monopole antenna is connected directly to the RF BPF (Band Pass Filter) controlled by the third RF Switch stage in order to filtered the third band signals. as shown in Fig. 9.



(a) Third Band Filter Circuit



(b) Third Band Filter

Fig. 9.: 2 Figures of Third Band Pass Filter.

The wideband spectrum monitor's antenna is a 470 mm monopole used in a wide frequency range without any matching circuit. The real part of the input impedance is between 1Ω and $1k\Omega$, depending on the actual frequency. The only way to calibrate the measurement system is: in our an-echoic chamber, in front of a log-periodic wide-band antenna, fed with known transmit power at fixed distance, the RSSI values had been recorded versus the frequency in 1 MHz step. As a result, the spectrum receiver has a calibration vector. This calibration vector is stored in the GND signal processing software. MRC-100 will downlink the original time-stamped RSSI values versus frequency, the correction based on the calibration vector will be done at the ground station (post processing).

As shown in Fig. 10.and Fig. 12.The radiation pattern of the on-board antenna with 45 degree and Fig. 11. Antenna on-board cube-skeleton.

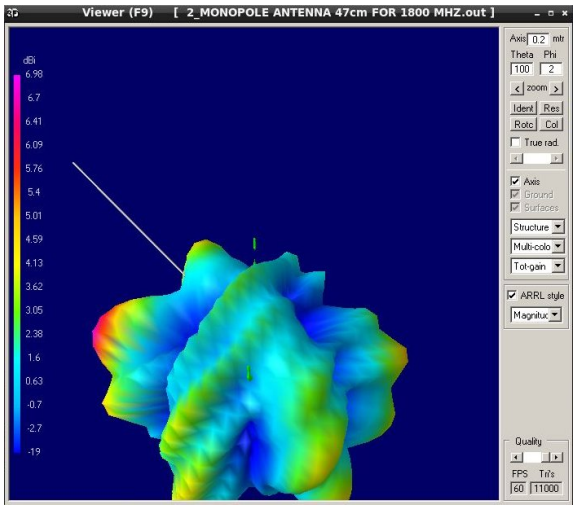


Fig. 10.: Three dimensional radiation pattern of the antenna with 45 degree.

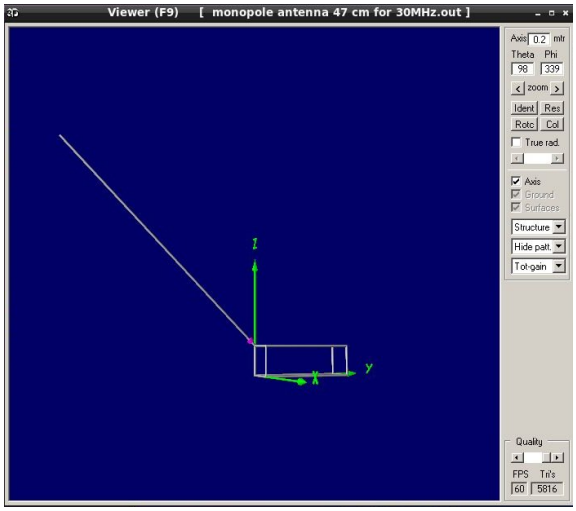
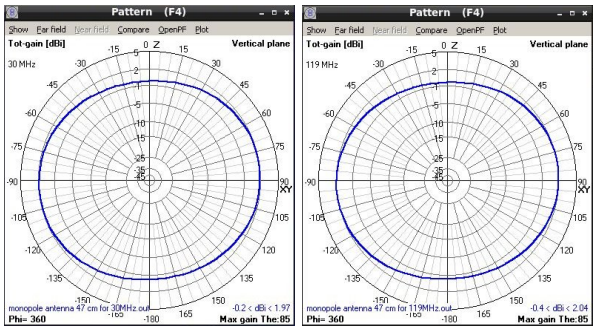
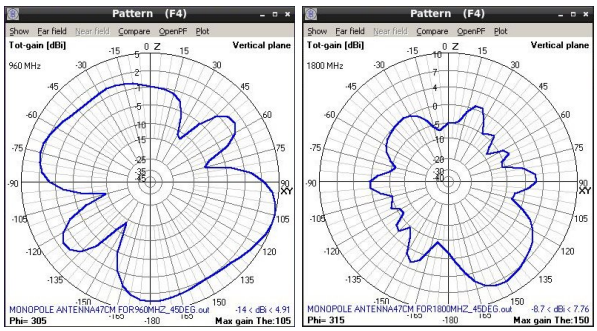


Fig. 11.: Antenna on-board cube-skeleton.



(a) 30 MHz Antenna Radiation Pattern (b) 119 MHz Antenna Radiation Pattern



(c) 960 MHz Antenna Radiation Pattern (d) 1800 MHz Antenna Radiation Pattern

Fig. 12.: 4 Figures of Antenna Radiation Patterns with 45 degree.

V. SPECTRUM MEASUREMENT RESULTS

A wide band spectrum monitoring system of MRC-100 satellite tested on the top of the roof of building V1 in BME to measure the upper HF band, FM band, VHF band, UHF band, LTE band, GSM band, 4G band and UMTS band signal levels. The experimental model of the measurement system is mentioned in Fig. 6.

The maximum distance between the satellite and the ground station (communication in zero degree elevation angle - horizon) as in Fig. 13. is based on the equation(6) at 600 km apogee/perigee of the orbit.

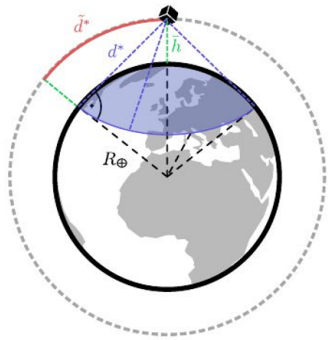


Fig. 13.: Diagram of MRC-100 Horizon. [7]

$$d = \sqrt{(R+h)^2 - R^2} \tag{6}$$

Where h = 600 km, R = 6,371 km and d = 2830 km (where d is the maximal distance between the satellite and the ground station).

The speed of a satellite in a circular orbit, it is calculated by equation (7).

$$v = \sqrt{\frac{g \cdot R}{1 + H/R}} = 7.55 \ \frac{km}{s} \tag{7}$$

g is the gravitational acceleration on the surface of the Earth.

The used bandwidth of the monitoring receiver is 800 kHz (the receiver chip maximal bandwidth is 850 kHz), controlled by the configuration software. So it is necessary to calculate the measurement minimum time for each band as shown in TableI, the minimal time depends on the Resolution Bandwidth (RBW) and the step frequency.

TABLE I
MINIMAL TIME [S] TO COMPLETE RF SCANNING.

	RBW [kHz]	GMSK [kbit/s]	1st Band [s]	2nd Band [s]	3rd Band [s]
0	1.5	1	1582	1495	14933
1	3	2	396	3737	3733
2	6	4	99	934	933
3	12	8	24	233	233
4	24	16	6	58	58
5	48	32	1.5	14	14
6	96	64	0.3	3.6	3.6
7	192	128	0.096	0.9	0.9
8	384	256	0.024	0.22	0.22
9	768	512	0.006	0.057	0.057

$$Minimal\ Time = \frac{f_{\max} - f_{\min}}{f_{step}} \cdot \frac{1}{RBW} \cdot 10 \tag{8}$$

$$DataRate_{GMSK} = 2^{RBW} \tag{9}$$

$$RBW = 1.5 \cdot DataRate \tag{10}$$

$$Step\ Frequency = \frac{RBW}{4} \tag{11}$$

TABLE I
MINIMAL TIME TO COMPLETE BAND SCANNING.

Band	RBW KHz	GMSK(DR) kb/s	Step Frequency KHz	Minimal Time second
FM	192	128	48	0.022
DVB-T	192	128	48	0.009
LTE	192	128	48	0.011
4G	192	128	48	0.022
GSM	192	128	48	0.048
UMTS	192	128	48	0.005

The total measurement points of the wideband spectrum (30 - 1800)MHz, when the Resolution Bandwidth (RBW) is 192 kHz and the f_{step} is 48 kHz can be calculated by equation 12.

$$Total\ Points = \frac{f_{\max} - f_{\min}}{f_{step}} = 36875\ points \tag{12}$$

The maximum flash memory of the OBC (On-Board Computer) of MRC-100 is 8 MB, this means 36875 measurement points will be saved on 36 KB (Because every 1 RSSI value equals to 1 Byte memory size).

The minimum time of measuring the RSSI values of the wideband spectrum (30 - 1800)MHz, when the Resolution Bandwidth (RBW) is 192 kHz and the Step Frequency is 48 kHz can be estimated by equation 13.

$$Minimal\ Time = TotalPoints \cdot \frac{1}{RBW} \cdot 10 = 1.92\ s \tag{13}$$

According to the estimated time in Eq. 13, we can analyze the wideband spectrum every 23 seconds.

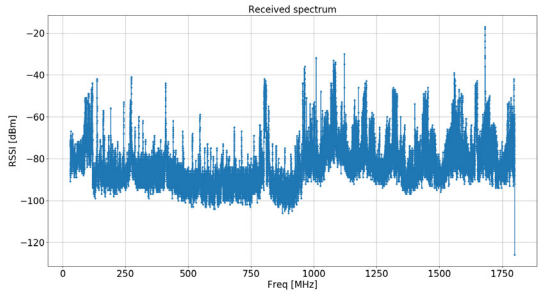


Fig. 14.: The hole Band Received Signals.

The transceiver chip measures the spectrum of the first band 30MHz to 119MHz, as shown in Fig.15 . There is a huge amount of RF power radiated from several FM local transmitters (88 - 108)MHz.

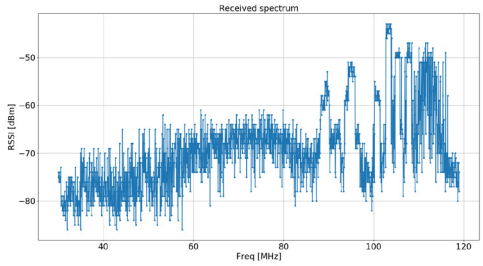
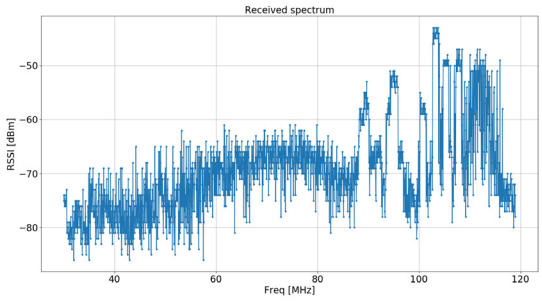
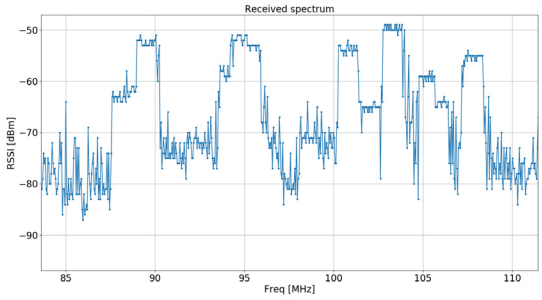


Fig. 15.: The First Band Received Signals.



(a) First Band



(b) FM Received Spectrum

Fig. 16.: 2 Figures of FM Received Spectrum

As shown in Fig.17 . the transceiver chip measures the spectrum of the second band (119 - 960)MHz, There is a huge a mount of RF power radiated from several transmitters.On 406, 500, 610, 640, 750, and 770 MHz the local TV transmitter's signals are very visible.. On 800 MHz the LTE is also highly visible and on (950 - 960)MHz the GSM band is highly visible. The digital terrestrial TV channel (DVB-T) has an 8 MHz bandwidth, hence the usable bandwidth is 800 kHz. Because the (DVB-T) bandwidth is 10 times highest than the bandwidth of the receiver chip, the only way is to modify the level of the measured data by increasing the RSSI values by 10 decibels.

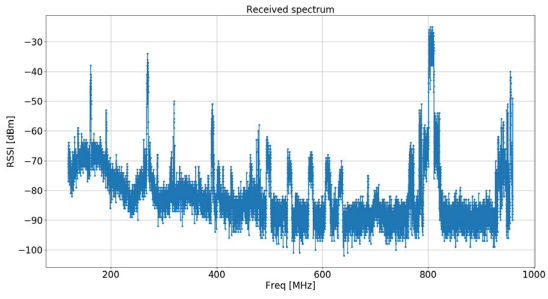
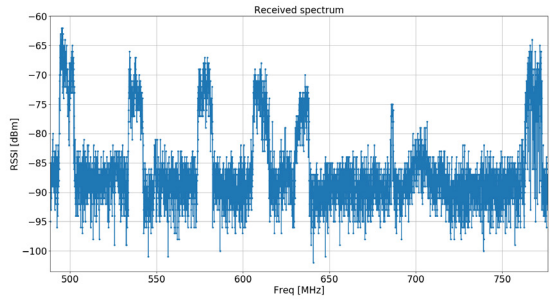
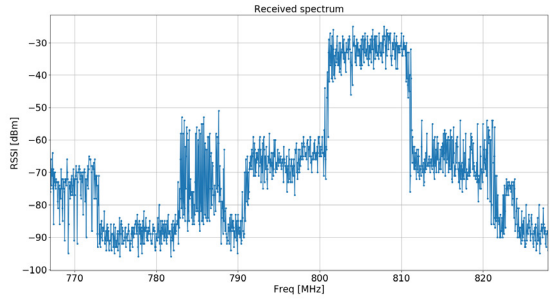


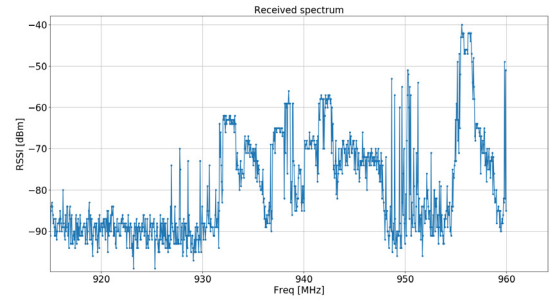
Fig. 17.: The second Band Received Signals.



(a) DVB-T Received Spectrum



(b) LTE Received Spectrum



(c) GSM Received Spectrum

Fig. 18.: 3 Figures of The Second Band Received Spectrum

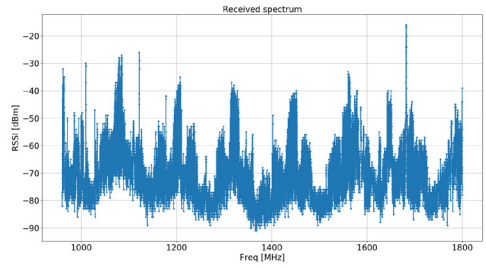


Fig. 19.: The Third Band Received Signals.

As shown in Fig.19 . the transceiver chip measures the spectrum of the third band 960MHz to 1800MHz, There is a huge amount of RF power radiated from several UMTS local transmitters between(1.77 - 1.8)GHz. The effect of the second

harmonics of the local oscillator of the third band is highly visible at 1682MHz as shown in Fig. 21.

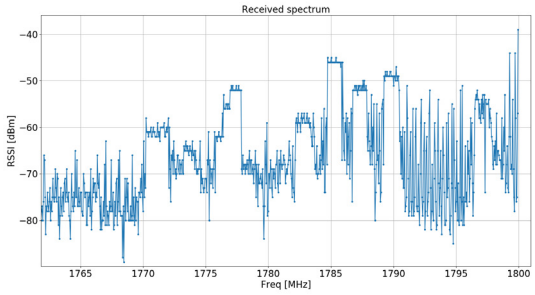


Fig. 20.: UMTS Received Spectrum.

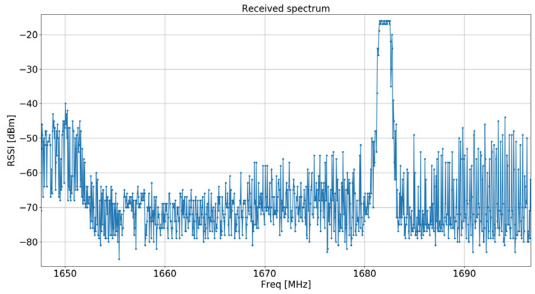


Fig. 21.: The Second Harmonic of the Local Oscillator.

As determined by the measurements, the SI4464 chip can sense RF signal level in -110...-10 dBm range linear in dB scale: the RSSI level can be calculated by the RSSI register value: $RSSI_{dBm} = \frac{RSSI_{reg}}{2} - 130$ modified the calibration vector.

VI. Conclusion

MRC-100 is in the designing and developing phase and will be launched in December 2022. The wideband spectrum monitoring system now is working and is able to measure RSSI (Received Signal Strength Indicator) values on three different RF bands according to the presented measurement results, it can be used as a conventional scalar spectrum analyzer with less than 120 mA current consumption from +3.3 V nominal regulated bus voltage and $40mm \times 40mm$ PCB (Printed Circuit Board) size. The system has enough sensitivity and enough dynamic range to be a payload on the MRC-100 3-PQ (PocketQube) satellite to measure RF (Radio Frequency) smog of the upper HF band, FM band, VHF band, UHF band, LTE band, GSM band, 4G band, and UMTS band over the globe in (LEO) orbit.

References

[1] http://152.66.80.46/smog1/satellites.pdf

[2] Dudás, L., Varga, L., & Seller, R. (2009, August). The communication subsystem of Masat-1, the first Hungarian satellite. In Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2009 (Vol. 7502, pp. 184-193). SPIE. DOI: 10.1117/12.837484

[3] http://www.azurspace.com/images/0003429-01-01_DB_3G30C-Advanced.pdf

[4] https://www.silabs.com/documents/public/data-sheets/Si4464-63-61-60.pdf

[5] Dudás, L., Pápay, L., & Seller, R. (2014, April). Automated and remote controlled ground station of Masat-1, the first Hungarian satellite. In 2014 24th International Conference Radioelektronika (pp. 1-4). IEEE. DOI: 10.1109/Radioelek.2014.6828410

[6] https://gnd.bme.hu/smog

[7] Takács, D., Markotics, B., & Dudás, L. (2021). Processing and Visualizing the Low Earth Orbit Radio Frequency Spectrum Measurement Results From the SMOG Satellite Project. INFOCOMMUNICATIONS JOURNAL, 13(1), 18-25. DOI: 10.36244/ICJ.2021.1.3

[8] Dudás, Levente; Gschwindt, András; Filling the Gap in the ESA Space Technology Education, 4th International Conference on Research, Technology and Education of Space February 15-16, 2018, Budapest, Hungary ISBN 978-963-313-279-1, https://space.bme.hu/wp-content/uploads/2019/01/Proceedings_abstracts_HSPACE2018.pdf

[9] Dudas, L., & Gschwindt, A. (2016, May). The communication and spectrum monitoring system of Smog-1 PocketQube class satel- lite. In 2016 21st International Conference on Microwave, Radar and Wireless Communications (MIKON) (pp. 1-4). IEEE. DOI: 10.1109/MIKON.2016.7491999

[10] Dudás, L., & Varga, L. (2010). Masat-1 COM. Antenna Systems & Sensors for Information Society Technologies COST Action IC0603, Dubrovnik.

[11] Levente Dudas, Laszlo Szucs, Andras Gschwindt, The Spectrum Monitoring System of Smog-1 Satellite, 14th Conference on Microwave Techniques, COMITE 2015, Pardubice, pp. 143-146, ISBN:978-1-4799- 8121-2. DOI: 10.1109/COMITE.2015.7120316

[12] Ahmed, Y., Babiker, W., TagElsir, A., & Daffalla, M.M. (2015). Design and Selection Criterion of Imaging Device as Payload for Cube Satellite. https://www.isnet.org.pk/pdf/research-003.pdf

Yasir Ahmed Idris Humad received his Bachelor's Degree in Electronic Engineering (Telecommunication) from AL-Neelain University, Faculty Of Engineering in 2012. He continued his graduate studies at AL-Neelain University Faculty of Engineering and he received his MSc in Electronic Engineering (Data and Communication Networks) in 2015. He enrolled as a researcher assistant at the Institute of Space Research and Aerospace (ISRA) Sudan in 2014. While working in ISRA, his research effort has been in the area of CubeSat Systems Design. In 2020, he granted a scholarship to complete his Ph.D. in electrical engineering at BME, Hungary. His current research areas include CubeSat and PocketQube type satellite development; analog RF hardware, antenna design; automated and remote-controlled satellite control station development; Software Defined Radio-based signal processing for satellites.https://gnd.bme.hu/smog

Levente Dudás At the Budapest University of Technology and Economics, he got his MSc in electrical engineering in 2007 and his Ph.D in 2018 in Radar and Satellite Applications of Radio and Antenna Systems (BME). He is currently a senior lecturer at BME's Department of Broadband Infocommunications and Electromagnetic Theory, an electrical engineer in the Microwave Remote Sensing Laboratory, and the president of the BME Radio Club. Microwave Remote Sensing (RADAR) lab. is working on: active and passive radars; CubeSat and PocketQube satellite development; analog RF hardware and antenna design; automated and remote-controlled satellite control station development; Software Defined Radio based signal processing for satellite and radar applications are just a few of his research interests. http://radarlab.hvt.bme.hu/, https://gnd.bme.hu/

A practical framework to generate and manage synthetic sensor data

Zoltán Pödör[1], and Anna Szabó[2]

***Abstract*—A huge number of sensors are around us and they generate different kinds of data. Data owners, e.g. the companies need IT environments and applications to handle these datasets. The collected data often contain sensitive information about the operation of the companies and the production processes. Therefore, artificial sensor data are strongly needed in the development and testing phase of these applications.**

In this paper, we introduce a complex application with three main modules to manage synthetic sensor data. The first component is the data generator module, which is capable of creating synthetic sensor data according to the user-defined distributions and parameters. The second module is in charge of storing the generated data in a flexible relational database, developed by us. The third component ensures the filtering and the visualization of the collected or generated data. A common interface was created to bring together the components and to provide a unified interface for the users. The adequate user management was an important aspect of our work. Accordingly, four different user types and authorities were defined.

***Index Terms*—synthetic sensor data; data generation; database for sensor data; data visualization**

I. INTRODUCTION

Nowadays various types of sensors are available to measure various things around us. They can be applied in our everyday lives to map the attributes of our environment or to measure our health conditions. In addition to this, they can be part of a smart home or smart devices. Apart from these possibilities, sensors can be utilized in industrial environment as well. According to the challenges of IoT and Industry 4.0 an expanding number of the companies and factories apply sensors [1] [2]. These devices measure different kinds of quantitative and qualitative attributes connected to the production, including the actual manufacturing processes and other related attributes (e.g. current consumption) [3]. Collected data can help us to improve the efficiency of the product, to reduce the cost, the amount of waste and to increase the income and the profit. Companies often require special, individual software solutions and applications connected to their data processing and displaying [4].

Security issues - connected to IoT and Industry 4.0 -, such as privacy, access control, information storage and management and the reliability of data management software are the main challenges [5] [6]. Besides the security and privacy questions, the reliability of the developed software is an important problem [7].

Bugs in the code or an incorrectly implemented analysis method can cause different kinds of problems, e.g. a huge loss for the company. Different aspects of software testing are important parameters of developing software that are free from bugs and problems [8]. Test data are the basic elements of the reliable and well-qualified software testing methods [9]. In many cases, the collected data can be sensitive, especially in an industrial environment. For example, the machine and the product data of a given company or the personal data of the workers can be considered sensitive. In many cases, the person or the company who ordered the software cannot disclose the sensitive data to the developers, only in the form of transformed, encrypted data [10].

There are two main ways of creating artificial data in order to handle this problem. First one is data masking [11] when real data are replaced with generated data with a high, measurable level of similarity. The name of these artificial data is semi-synthetic, or hybrid data. Another type of artificial data is the full-synthetic data that is created by an algorithm, and it is usually used for test datasets of production or operational data [12]. In this paper, we focus on the problems of functionality testing.

Synthesis of data is a simple simulation with the primary aim of generating data according to a given model [13]. The elementary simulations can be: (1) sequences that generate various increments; (2) randomizers that create random values using any well-known distributions; (3) mathematical functions that describe the form of generated data; (4) noisers which are generators of lists with missing values, range filters, etc. In this paper we focus on the randomizers with some general distributions. The details and descriptions of different kind of data generator applications are introduced in Section II.B. These and the above mentioned solutions and applications focus only on the data generating method, without the possibility of included, direct and efficient data storing and data processing. The data handler applications target special kind of visualization and analysis possibilities [14], including the pre- and post-processing methods. They usually do not have the ability to create artificial sensor data. They usually use outer databases, which store the actual used data and they focus only on data visualization and data analysis. The uniqueness of our solution is that these three important components (data generator, database and data handler) are connected into one complex application with an appropriate user management according to the stored datasets.

[1] Zoltán Pödör is with Eötvös Loránd University, Faculty of Informatics (e-mail: pz@inf.elte.hu)

[2] Anna Szabó is with Eötvös Loránd University, Faculty of Informatics (e-mail: h6co1g@inf.elte.hu)

DOI: 10.36244/ICJ.2022.2.7

For that reason, on one hand, we offer a Python-based solution to create artificial sensor data according to different conditions (range, measurement frequency, measurement unit, accuracy, etc.) to help the testing period of the software development process. The generated sensor data is stored in the database module that is an integrated part of our application. The data generator was implemented to test the functionality of the developed software with special distributions of the generated synthetic sensor data. On the other hand, our solution is a complex framework with a database to store the generated data and a visualization surface to check, visualize the generated data. Another big advantage is the common surface above the modules that contains all available services of the application, and it grants the services only to the authorized users from the artificial sensor creation through the data generation to storing and visualization. The data visualization and the basic analysis modules allow to process data, from different, outer sources, but stored in the application's database.

Our main aim is to support different kinds of data-based application development and testing phase with artificial sensor data samples. The problem of sensitive data can be handled by the included data generation module during the software development period. Our application can be used in different areas. The integrated database and visualization modules provide the opportunity to handle the data that is either artificially generated or comes from an external source. The module-based structure provides easy expandability in terms of the number of the modules and their structure while the user-friendly interface provides easy handling and use for non-IT professionals because it does not require any special prior knowledge. In the future we plan to include other types of databases and to develop the visualization and analysis modules next to the data generator module.

II. THE APPLICATION COMPONENTS

The application contains three main modules (Fig. 1.). First one is the flexible database to store the generated artificial data with all connected properties, like timestamp, unit, location, devices etc. Another database was created to store separately the data of the different users.

Second one is the data generator module that is responsible for the generation of the artificial sensor data according to the user-defined parameters and distributions.

Third part is the filtering and visualization module that intends to select and visualize the appropriate data. This module gives the opportunity to create some basic statistical properties of the generated data besides the visualization which can help us to compare the artificial data with the real data.

These three modules are embedded into a web-based interface to provide a unified user interface, easy application handling, and user-friendly system for all potential users. The common surface includes all available services of our application but only for the authorized users.

Because of the data security difficulties, user management was an important aspect of our work. Four different types of users were created. The tool admin can handle the sensors (create, delete and modify them), the data generator can use and set the data generator module, the data handler uses the analysis and visualization module while the user admin manages the data (name, e-mail address, roles etc.) of the system's users and registers new members. In the next chapters we will introduce these components in detail.

A. The database

Not only the sensors but also the systems which can handle them (store, analyse the collected data, display the raw and the

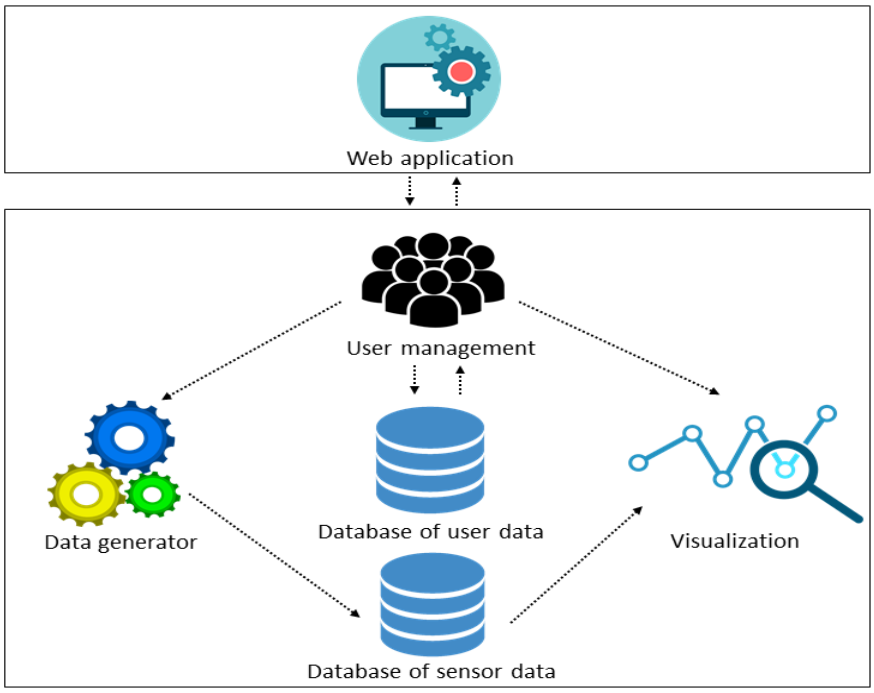


Fig. 1. Structure of the application

processed data) are important. A basic element of such a system is an efficient, reliable and flexible database to store the collected sensor data. Relational and non-relational databases have a lot of advantages and some disadvantages too and both of them offer the opportunity to store the collected sensor data [15]. The structure of a non-relational database is not fixed (unstructured), which means that there are no relation schemas, connections, and the form and the structure of data can be easily modified [16]. The IoT and the sensor based applications often use non-relational databases, because they have a lot of advantages, e.g. the free structure of the database [17]. One of the advantages of the relational databases that they are capable of performing more complex queries and filtering [18] [19]. Our main aim is to create and store artificial sensor data to help the testing of the functionality of the developed software and not to serve big queries. Because of the above-mentioned reasons, we decided to create and use a flexible relational database in MSSQL environment. It is the first module of our complex system, which can store the data derived from the data generator module.

The structure of the database can be seen in Fig. 2 shown as an entity-relationship diagram. The following tables were defined:

- ***Sensor_types*** table: general characterization of the different sensor types defined by a unique identifier (*id*) and each type has a practical name (*name*). *Max_value*s and *min_values* define the possible minimum and maximum values measured by the given sensor type. The *measurement_length* defines the length of one measurement and the *measurement_accuracy* defines the measurement precision of the actual sensor type.
- ***Sensors*** table: it describes the parameters of a given sensor, which belongs to a certain type. Each one of the sensors is defined by a unique identifier (*id*) and has a practical name (*name*). The *alarm_low_boundary* and the *alarm_over_boundary* attributes define the lowest and the highest measured values which might be practically correct. The *frequency* defines the regularity of the measurement. Furthermore, the year of production (*year_prod*) and the date of calibration (*calib_date*) are stored in this table.
- ***Measurement_types*** table: describes the parameters of the measured data. Each piece of data has a unique identifier (*id*) and has a practical name (*name*). The unit of the measured values is stored as well (*unit*).
- ***Measured_values*** table: it contains the measured sensor values, which are defined by two attributes: the *date* and the *sensor.id*. The attribute *measured_value* stores the measured sensor values. The attribute *valid* is a binary parameter with only two values, 1 and 0. Value 1 shows that the measured value is correct, which means that it is between the *sensors.alarm_low_boundary* and *sensors.alarm_over_boundary* values from Sensors table, and the *error_rate* attribute is NULL. Value 0 shows that the measured value is incorrect. In this case, the value of *error_rate* is the value of signed distance between the measured value and the given boundary. The values of attributes *valid* and *error_rate* are created automatically by a trigger.
- ***Devices*** table: describes the parameters of the devices which contain the sensors. Each device is defined by a unique identifier (*id*) and has a practical name (*name*).

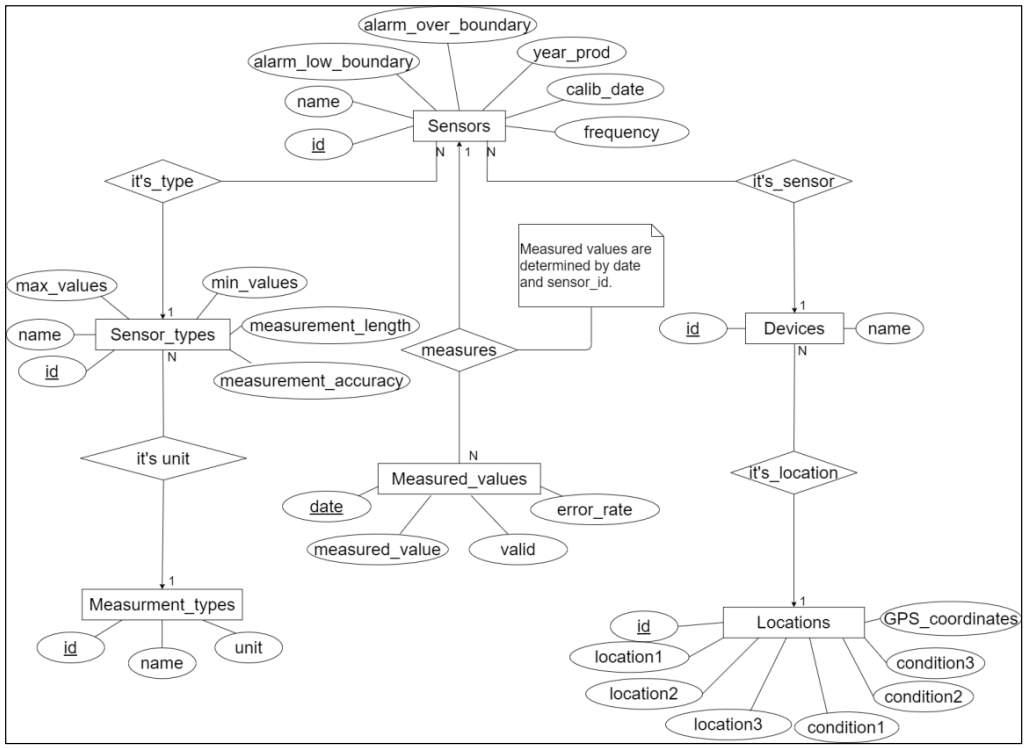


Fig. 2. Entity-relationship diagram of the database structure

- ***Locations*** table: describes the parameters of the location of the devices and sensors. Each place and location is defined by a unique identifier (*id*). Attributes *location1*, *location2* and *location3* store the details of the location, e.g. the name of the location or company. The attribute *GPS_coordinates* contains the accurate location in GPS format. *Condition1, condition2* and *condition3* provide an opportunity to store information about the conditions of the location, e.g. if the sensor or the device is an indoor or outdoor gadget.

The next connections were defined between the tables:

- Each of the sensors has one and only one sensor type. In other words, there are sensor families, with some well-defined general parameters in which we can define a concrete sensor, which belongs to the given sensor family. This connection is a one-to-many relationship, because one sensor family can contain more sensors and one sensor always belongs to only one family.
- Each of the sensor types includes the unit and the type of the measurement. It is a one-to-many relationship, because a given sensor type measures only one thing, but one type of measurement can measure more than one sensor type.
- The measured values are connected to the sensor which measured them. This is a one-to-many connection, because a measured value always belongs to only one given sensor, which measured it, but one sensor can measure more than one data.
- All sensors belong to a device, which contains the given sensor. This connection is a one-to-many relationship, because a device can contain a lot of sensors, but one sensor is always in one device.
- It is important to store the conditions, e.g. the place of the sensor. It is a one-to-many connection, because a given device is always in one place, but more than one device can be in one place.

Our database is suited for storing any measured sensor data or storing the data, which are generated by the application's second module, the data generator. Of course, in the future, if it is necessary or practical, we can replace our relational database with another type of database.

B. The data generator

This module supports the development and testing phase of the applications, which use sensitive data, because it generates artificial data according to the user-defined parameters. These artificial data are similar to the original but sensitive data. There are some solutions for generating synthetic or artificial data, but they usually have other, special goals with their solutions. For example, it is possible that they are made for a special task. Zimmering et al. [20] created a novel method for a generation process of data to compare the selected machine learning methods. Tam et al. [21] have developed an automatic process for creating input files (connected to building sensors) to a fire model simulation. Norgaard et al. [22] proposed a supervised generative adversarial network architecture to create synthetic sensor data connected to health monitoring.

In some synthetic generation, tools need a sample of the real-data as an input. They learn from the original dataset, and they generate the new clone datasets based on this information. CTGAN [23] is an open-source project from MIT which is a collection of Deep Learning-based Synthetic Data Generators. Synsys [24] is a system written in Python and it uses Hidden Markov Models to generate sensor event sequences and it accepts an existing dataset as input and generates a similar synthetic dataset.

There are some solutions which do not need a sample dataset but need a schema to describe the real-data. We have taken to examples to show it. Log-synth [25] generates data in accordance with a given schema. It contains only two distributions which are the normal and random walk. The user can define starting time, frequency and the parameters of the given distribution in the schema file. Iosynth [26] is similar to log-synth, but it allows for more flexibility in terms of defining the type of data to be generated (in a schema file). Furthermore, it has more implementations of distributions to choose from. We can choose either fixed interval sampling, normal or exponential distributions. Both solutions can create JSON files as an output.

The above-mentioned data generators are simple solutions, which implies that they do not have either an included database, user management or a web-based surface to handle the whole application. They do not give the opportunity to create artificial sensors and devices connected to the data synthetization.

Our main aim is to generate artificial sensor data to support the testing phase of software and their functionalities which were developed to handle the original sensor data. It was an important aspect of creating this application not to use real data in the generation method, only their, like Iosynth and log-synth.

The main part of the generator module was implemented in Python language. A public Github project, which was developed by a Korean developer team, was integrated into our module to help us to create data. It is called Mandrova which means "make it" in English, but it means "make sensor data" in the context of sensors [27]. Developers can generate values with many kinds of distributions with the help of this project, but only three of them were used:

- normal distribution with the mean (μ) and the standard deviation (σ) parameters:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

- exponential distribution with lambda parameter:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, where\ x \geq 0 \\ 0,\ where\ x < 0 \end{cases} \tag{2}$$

- gamma distribution with α and β parameters:

$$f(x) = \begin{cases} \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, where\ x > 0 \\ 0, \qquad\qquad where\ x \leq 0 \end{cases} \tag{3}$$

These three types were chosen in our current solution, because they are among the most general distributions [28]. The surveyed data generators, which use the distribution of the data,use less or the same distributions as our application. Log-synth [25] contains only normal distribution and random walk. Using Iosynth [26] we can choose normal or exponential distributions and fixed interval sampling. We are planning to build in other distributions in the future to extend the possibilities of this module.

The generation of the values is based on four basic parameters. At first, an existing sensor must be chosen for which the values will be generated. All sensors (which are stored in the database, except for those that have not been connected to a sensor type) are listed and can be selected. The second criterion is the type of statistical distribution with appropriate parameters. The user also has to decide how much data should be generated. Finally, a timestamp has to be given, which is the date of the latest created data. Before the actual data generation process, some important information is automatically queried from the database. The table ***sensor_types*** guarantees the following attributes of the chosen sensor: *measurement_accuracy*, *max_values* and *min_values*. *Measurement_accuracy* defines the accuracy of the stored data. If this record is missing, the rounding value is one by default (as one decimal). When the generated value exceeds the *max_values* or it is less than the *min_values* parameters, the generated data's new error indicator values are 9999 or -9999 (it depends on whether the value is over the maximum or under the minimum). This notation helps us to determine and handle "measurements" which are certainly wrong. If both *min_values* and *max_values* fields are empty in the database, all incoming data are accepted. If only one of them is missing, all the generated values are correct which are under or over the existing limit.

In accordance with the above-mentioned conditions, the generator module's algorithm has three main steps. The first one is to generate the artificial data based on the user-defined distribution and connected parameters. The second one is the checking method, to decide whether the generated value is valid or not according to the parameters *min_values* and *max_values*. If it is necessary, the algorithm changes the generated values to the error indicator values. The last step is the data rounding on the basis of the parameter *measurement_accuracy*. After that, the module stores the new records in the database.

C. Filtering and visualization

Data filtering and visualization are the most important basic tasks of data handling and they are essential for further data analysis [28]. The third module's visualization part is a JavaScript-based component and it gives the opportunity to handle and visualize the generated datasets and to analyse them later in Python language following further development. Currently, the module has two main functions: filtering data by different aspects, like timestamp, sensors, places etc. and visualizing and creating basic statistical properties of datasets. This second function allows a basic comparison between the generated and the original, real data.

A user-friendly interface was created to filter and visualize the stored data. It was an important aim not to develop this module only for IT specialists. Therefore, a huge number of automatized solutions were built in to help users.

Users can filter by devices, sensor types or sensors and enter a starting and an ending date. The device selection is optional, but if a device has been chosen, only those sensor types and sensors are available which are connected to the selected device. Otherwise, all stored sensor types are available to users. Sensor type selection is required because our goal was to visualize only those sensors which have the same type and for example the same unit. Obviously, the user must choose at least one sensor. The last two parameters, the timestamps are optional. After setting these parameters, the application queries all records from the database according to the selected parameters. Our aim was to visualize more than one sensor from one sensor type, but we had to handle the problem of different timestamps of different sensors. To solve it, an algorithm was created and implemented. For visualization, we had to use a two-dimensional table, which stores the values that the user wants to display. The first column stores timestamps after which there is one column for each sensor. This table is loaded up with data by the above-mentioned algorithm whose main task is to check if the selected sensors made measurements at a given time or not. The algorithm uses all the timestamps when any of the selected sensors (in the given time period) made a measurement. Timestamps are stored twice, in two arrays. In the first one, we store all the available dates in chronological order. All of them are stored in the second one too but they are grouped by sensors. The algorithm iterates over the first array and compares the elements to the other array's timestamps. If they are equal, it means that a measurement was made at this time by the current sensor, and we store this measured value in the given sensor's column. Otherwise, we insert a null value into the appropriate sensor in the given timestamp. Such a table can be seen below (Table I.). The timestamps are in the first column, and it can be seen in the others that the three sensors (named S1, S2, S3) did not make measurements at the same time, so there are null values in these fields.

TABLE I
AN EXAMPLE FOR THE UPLOADED TWO-DIMENSIONAL TABLE

date	S1	S2	S3
2021-11-02 14:00:00	null	17.03	null
2021-11-02 14:05:00	20.17	null	10.58
2021-11-02 14:10:00	null	23.11	null
2021-11-02 14:15:00	22.05	null	9.87

For displaying the selected data, Google Charts was used which is an API that helps developers to display data on diagrams simply. The easiest way to use it is to load some libraries and embed a JavaScript code in the application [30].

The application has a user-friendly web-based interface to provide a uniform display for the using of all three main modules and their functionalities. The available application menus depend on the actual user's authorities. This solution ensures that users can only use those functions which are available to them. To guarantee this, user management is an important part of the application.

III. USER ADMINISTRATION

Data security and the adequate user authorities are basic requirements for such an application [31]. According to this, our software can be used only by registered users. The attributes of the users are stored in an independent database. It consists of two tables. One is for the users and the other is for the different roles. A user can have more than one role (Fig. 3).

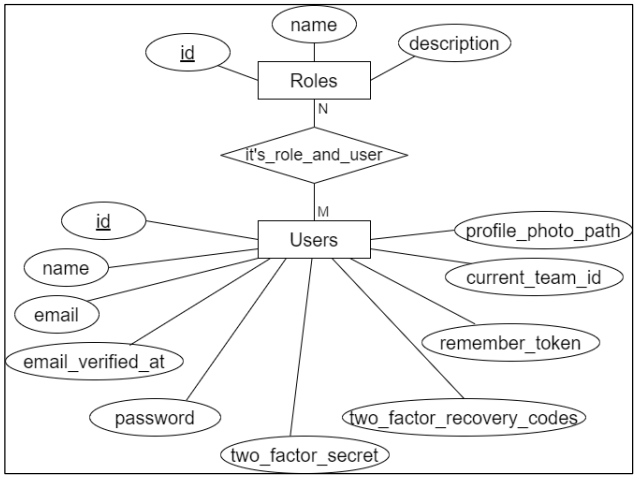


Fig. 3. Users and roles

Four different roles are defined. These roles are separated from each other, and they are related to well-defined tasks. First is ***tool admin*** who manages the data of measurement types, sensors, sensor types, devices and locations. It is the tool admin's responsibility to modify the attributes of a given element, e.g. the alarm boundaries of a sensor, the year of the production, the date of the calibration and the measurement frequency.

Due to the relationship between the database tables, there may be anomalies after a modification or creation, and we have to handle them. A special menu was created for this. The application can identify and inform the tool admin about the next anomalies:

- There is no measurement type connected to a sensor type,
- A sensor is not connected to a sensor type,
- A sensor is not connected to a device,
- A device does not have a location.

The application can identify these anomalies, and it creates a list about the problems for users to correct them. The interface of this menu is shown in Fig. 4.

The second type of user is ***data generator***, who can generate the artificial sensor data by specifying various parameters. The details are in Section II B.

Data handler manages the visualization interface. This role gives the opportunity to give some filtering conditions. Afterwards, the diagram, which is created by the program based on the given conditions, can be viewed by the user.

The last user is the ***user admin***, who manages the data of the users and registers new users. User admin can handle the personal data of the users, like name, e-mail address and it is his/her responsibility to set the authorities of the users. The current surface of the application corresponds to the logged user's authorities. It means that a given user can reach only those functions that come under his or her authority. For example, a data generator user can only use the functions connected to artificial data generation, but not other functions, like visualization, sensor and user management. Of course, a user can be assigned multiple roles who can reach and use all related functions in the application.

IV. A DEMONSTRATION USE-CASE

In this section, we would like to show the functionality of our application from the data generation to data filtering and visualization. To do this, different kinds of users were defined with the appropriate roles and authorities. At first, we generated some artificial values with our data generator module. Before generating, some test elements (sensors, sensor types etc.) were inserted into our database. In this test, one sensor, named Tempr2 was selected, which is a thermometer, and the other parameters were entered. We chose the normal distribution and 1000 values were generated where the starting date was 2022. 01. 07 23:00:00. (Fig. 5). At the end of the process, a message tells us if the generation was successful or not. We manually checked the generated data in the database.

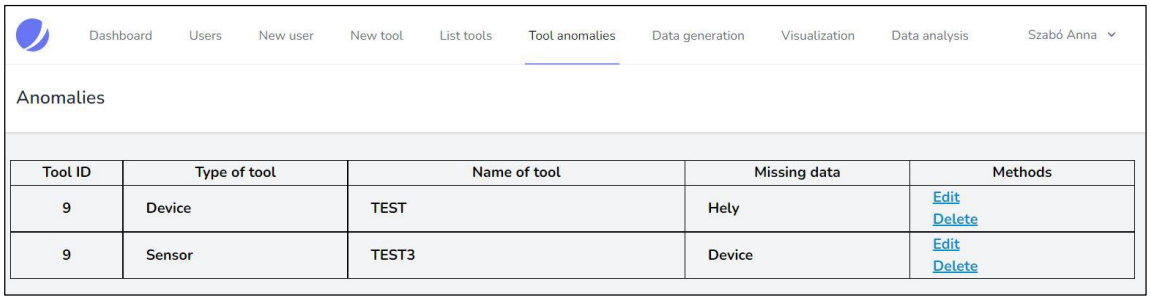


Fig. 4. Anomalies

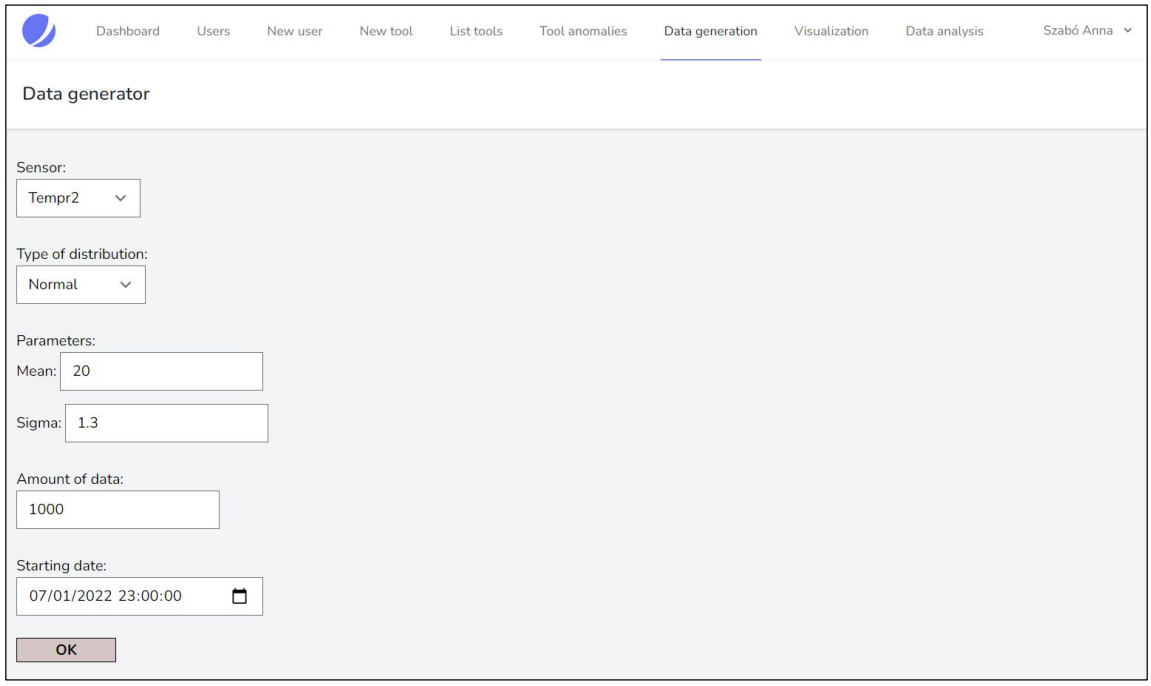


Fig. 5. Interface of a test data generation

After a huge number of values had been generated, we started to test the visualization component. As it can be seen in Fig. 6, humidity sensor type was selected and two sensors appeared in the box which means that there are two sensors in the database connected to the chosen sensor type. We also entered the starting and the ending date, and then the line diagram was created by the program as shown in Fig. 7. Besides, we tested the filter under the diagram that also worked properly as we could easily change the interval of the dates. In addition, some random values were selected from the diagram and we checked in the database if they matched, and we experienced that all of them were the same at the given times.

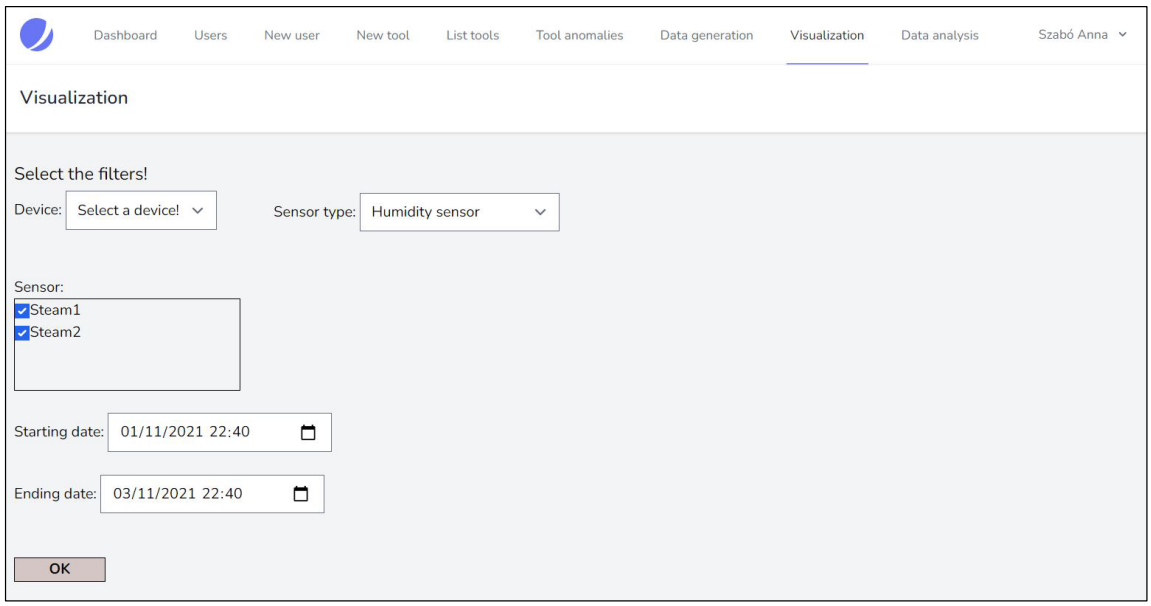


Fig. 6. Filtering before visualization

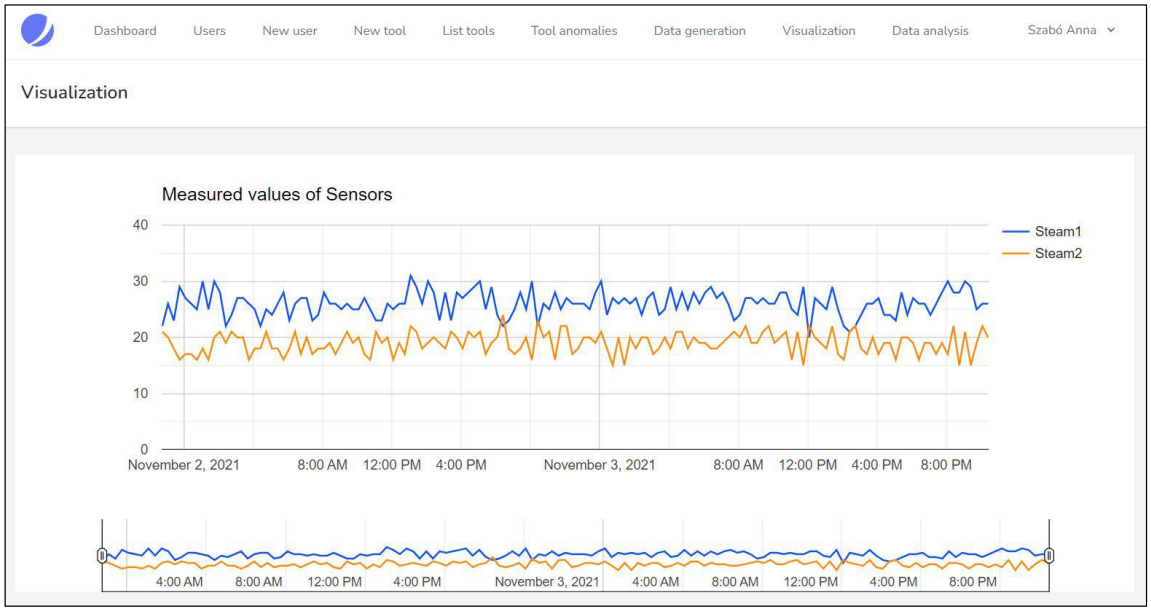


Fig. 7. Line diagram of measured values

V. CONCLUSION

In this paper we created and implemented an artificial sensor data generator as part of a complex sensor data handling application. The application includes a flexible database structure (designed by us) to store the data, the above-mentioned data generator module and a visualization module to filter and visualize the selected data. The main idea behind the generator module is that there are a huge number of applications which are developed to handle sensitive, industrial sensor data. This module is able to create artificial sensor data sample to support the development and testing phase of the software and their functionalities. A common web-interface guarantees access to the whole application.

User management was an important part of the software. Accordingly, the four defined roles and the connected authorities define the available application interface and functions for a given user. The connected data is stored in a separated database.

The examined similar data generators are simple solutions since they do not have included database and user-friendly surface. The parametrization and the use of them is possible only in command line or in special environment. User administration is the other specialty of our solution that defines the access to the artificial devices, to the sensor, and to the generated data.

The architecture and the module-based design of the application allows for later expandability and further steps of development. There are three possible distributions in our current generator module, but we can expand this circle with other important distributions to broaden the range of possibilities of this module. The visualization module currently contains simple analytic functions to create some basic statistical parameters and we are planning the expansion in this direction. The range of the user roles can be expanded too in the future. The current version of our framework – thanks to the well based database structure, user administration and module structure - is a good basis for the future making it possible for us to extend the sphere of the modules and their functionalities.

ACKNOWLEDGMENT

The research was supported by the project No. 2019-1.3.1-KK-2019-00011 financed by the National Research, Development and Innovation Fund of Hungary under the Establishment of Competence Centers, Development of Research Infrastructure Programme funding scheme.

REFERENCES

[1] G. Dalmarco, F. R. Ramalho, A. C. Barros, and A. L. Soares, "Providing industry 4.0 technologies: The case of a production technology cluster," *The Journal of High Technology Management Research*, vol. 30, no. 2, 2019, DOI: 10.1016/j.hitech.2019.100355.

[2] A. Koncz, and A. Gludovatz, "Calculation of indirect electricity consumption in product manufacturing," *International Journal of Energy Production and Management*, vol. 6, no. 3, pp. 229–244, 2021, DOI: 10.2495/EQ-V6-N3-229-244.

[3] S. A. Hashmi, C. F. Ali, and S. Zafar, "Internet of things and cloud computing-based energy management system for demand side management in Smart Grid," *International Journal of Energy Research*, vol. 45, no. 1, pp. 1007–1022, 2020, DOI: 10.1002/er.6141.

[4] J.-Q. Li, F. R. Yu, G. Deng, C. Luo, Z. Ming, and Q. Yan, "Industrial Internet: A Survey on the Enabling Technologies, Applications, and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1504-1526, 2017, DOI: 10.1109/COMST.2017.2691349.

[5] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the internet of things: Perspectives and challenges," *Wireless Networks*, vol. 20, no. 8, pp. 2481–2501, 2014, DOI: 10.1007/s11276-014-0761-7

[6] K. R. Sollins, "Iot Big Data Security and Privacy Versus Innovation," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1628–1635, 2019, DOI: 10.1109/JIOT.2019.2898113

[7] A. Garg and A. Arora, "Software Reliability–A Review", *International Journal of scientific research and management*, vol. 4, no. 7, 2016.

[8] H. Tahbildar and B. Kalita, "Automated Software Test Data Generation: Direction of Research," *International Journal of Computer Science & Engineering Survey*, vol. 2, no. 1, pp. 99–120, 2011, DOI: 10.5121/ijcses.2011.2108

[9] M. Esnaashari and A. H. Damia, "Automation of software test data generation using genetic algorithm and reinforcement learning," *Expert Systems with Applications*, vol. 183, 2021, DOI: 10.1016/j.eswa.2021.115446.

[10] M. A. Calles, "Protecting Sensitive Data," in *Serverless Security,* Berkeley, CA, USA: Apress, 2020, pp. 257-283, DOI: 10.1007/978-1-4842-6100-2_10.

[11] A Net 2000 Ltd., "Data Masking: What You Need to Know" 2016.

[12] N. Laskowski, "What is synthetic data? - definition from whatis.com," SearchCIO, 12-Feb-2018. [Online]. Available: https://www.techtarget.com/searchcio/definition/synthetic-data. [Accessed: 09-Mar-2022].

[13] S. Popic, B. Pavkovic, I. Velikic, and N. Teslic, "Data Generators: A short survey of techniques and use cases with focus on testing," *2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin)*, pp. 189–194, 2019, DOI: 10.1109/ICCE-Berlin47944.2019.8966202.

[14] O. Embarrak, Data Analysis and Visualization Using Python, Apress Berkeley, CA, 2018 DOI: 10.1007/978-1-4842-4109-7

[15] K. Fraczek, M. Plechawska-Wojcik, "Comparative Analysis of Relational and Non-relational Databases in the Context of Performance in Web Applications," in BDAS 2017 DOI: 10.1007/978-3-319-58274-0_13

[16] B. Anderson and B. Nicholson, "SQL vs. NoSQL databases: What's the difference?," *IBM*, 15-Jun-2021. [Online]. Available: https://www.ibm.com/cloud/blog/sql-vs-nosql. [Accessed: 19- Jan-2022].

[17] S. Kontogiannis, C. Asiminidis and G. Kokkonis, "Comparing Relational and NoSQL Databases for carrying IoT data", *Journal of Scientific and Engineering Research*, 2019.

[18] A. Gopani, A. Choudhary, S. Bhattacharyya, and S. Goled, "10 most used databases by developers in *2020," Analytics India Magazine*, 12-Jan-2022. [Online]. Available: https://analyticsindiamag.com/10-most-used-databases-by-developers-in-2020/. [Accessed: 19-Jan-2022].

[19] "Engines ranking," DB. [Online]. Available: https://db-engines.com/en/ranking [Accessed: 19-Jan-2022].

[20] B. Zimmering, O. Niggemann, C. Hasterok, E. Pfannstiel, D. Ramming, and J. Pfrommer, "Generating artificial sensor data for the comparison of unsupervised machine learning methods," Sensors, vol. 21, no. 7, 2021, DOI: 10.3390/s21072397.

[21] W. C. Tam, E. Y. Fu, R. Peacock, P. Reneke, J. Wang, J. Li, and T. Cleary, "Generating synthetic sensor data to facilitate machine learning paradigm for prediction of Building Fire Hazard," *Fire Technology,* 2020, DOI: 10.1007/s10694-020-01022-9.

[22] S. Norgaard, R. Saeedi, K. Sasani, and A. H. Gebremedhin, "Synthetic Sensor Data Generation for Health Applications: A Supervised Deep Learning Approach," in Annual International Conference of the EMBC, Honolulu, HI, USA, 2018, pp. 1164-1167, DOI: 10.1109/EMBC.2018.8512470.

[23] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni. "Modeling tabular data using Conditional GAN". 2019, DOI: 10.48550/arXiv.1907.00503.

[24] J. Dahmen and D. Cook, "SynSys: A Synthetic Data Generation System for Healthcare Applications," *Sensors*, vol. 19, no. 5, 2019.

[25] Tdunning, "TDUNNING/log-synth: Generates more or less realistic log data for testing simple aggregation queries.," *GitHub*. [Online]. Available: https://github.com/tdunning/log-synth. [Accessed: 09-Mar-2022].

[26] Rradev, "Rradev/iosynth: Iosynth is IOT device/sensor simulator and synthetic data generator.," *GitHub*. [Online]. Available: https://github.com/rradev/iosynth. [Accessed: 09-Mar-2022].

[27] Makinarocks, "Makinarocks/Mandrova: An Awesome Synthetic Sensor Data Generator for Python3," *GitHub*. [Online]. Available: https://github.com/makinarocks/Mandrova. [Accessed: 02-Mar-2022].

[28] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, 4th Edition, Hoboken, New Jersey, USA: John Wiley & Sons, Inc., 2011

[29] S. K. Peddoju, and H. Upadhyay, "Evaluation of IoT data visualization tools and techniques," in *Data Visualization*, Singapore: Springer, 2020. pp. 115-139. DOI: 10.1007/978-981-15-2282-6_7

[30] "Using google charts | google developers," Google. [Online]. Available: https://developers.google.com/chart/interactive/docs. [Accessed: 02-Mar-2022].

[31] K. Tabassum, A. Ibrahim, and S. A. El Rahman, "Security Issues and Challenges in IoT," in ICCIS, Sakaka, Saudi Arabia, 2019, pp. 1-5, DOI: 10.1109/ICCISci.2019.8716460.

Zoltán Pödör received the M.Sc. degree in Mathematics and Computer Science from the University of Szeged in 1999. He wrote his PhD thesis on the extension opportunities of time series analysis based on special method at University of West Hungary in 2014. Between 2006 and 2020 he worked at the University of Sopron, Hungary in various positions including Head of Institute of Informatics and Economics. Since 2020 he has been associate professor at Eötvös Loránd University, faculty of Informatics. His current research interests cover time series analysis, data mining techniques in practice and handling and processing of sensor data. He has more than 70 publications.

Anna Szabó has been a BSc student at Eötvös Loránd University, Szombathely, Hungary in Computer Science since 2019. Her current interests are storage, handling, visualization and analysis of different kinds of data (mainly sensor data).

On the Quality of Experience of Content Sharing in Online Education and Online Meetings

Tushig Bat-Erdene, Yazan N. H. Zayed, Xinyu Qiu, Ibrar Shakoor, Achref Mekni, Peter A. Kara, *Member, IEEE*, Maria G. Martini, *Senior Member, IEEE*, Laszlo Bokor, *Member, IEEE*, and Aniko Simon

***Abstract*—The turn of the decade introduced a new era of global pandemics to the world through the appearance of COVID-19, which is still an active crisis at the time of this paper. As a countermeasure, the phenomena of home office and online education became not only widely available, but also mandatory in many countries. However, the performance, reliability and general usability of such real-time activities may be severely affected by unfavorable network conditions. In both contexts, content sharing is now a common practice, and the success of the related use cases may fundamentally depend on it. In this paper, we present our surveys and subjective studies on the Quality of Experience of content sharing in online education and online meetings. A total of 6 surveys and 5 experiments are detailed, addressing topics of student experience, user interface settings, sharing options of lecturers and employees of the private sector, the perceivable effects of network impairments and the related long-term adaptation, the rubber band effect of slide sharing, the overall perceived quality and the separate quality aspects of media loading times, and the preference between visual quality, average frame rate and frame rate uniformity. The findings of the subjective studies do not characterize the use cases of the investigated topics on a general, widely-applicable level, as only a single online platform is involved throughout the experiments. However, their experimental configurations are reinforced by comprehensive surveys and many results indicate statistically significant differences between the selected test conditions.**

***Index Terms*—Quality of Experience, Quality of Service, online meeting, online education, video quality, video resolution, loading time.**

I. INTRODUCTION

DUE to the ongoing global pandemic SARS-CoV-2 – also known as COVID-19 – the employees of more and more companies and institutions perform their daily occupation-related activities from the safety of their homes. Similarly, as the virus appeared in every corner of the world – threatening the lives of millions – education suddenly shifted towards its online variations, as an attempt to battle this crisis. In numerous countries, online education is still the only reasonable option in 2021, since even at the time of writing this paper, although vaccines are already available, yet the disease remains to be dealt with – particularly due to the continuously evolving variants. Additionally, new threats are on the rise, such as the 2022 human monkeypox outbreak, and other pandemics may emerge as well.

Remote education via modern technology is far from being a completely novel phenomenon. In fact, media (i.e., radio and educational films) was already utilized for educational purposes more than a hundred years ago [1]–[3]. In the age of the Internet, we have a vast array of techniques to choose from. There are multiple types of self-learn, self-study software, pre-recorded lectures are available online – either publicly or solely to the students of the institution – and classes, lectures are interactively held via online communication platforms. However, the latter is a real-time educational service, and therefore, its perceived quality highly depends on network conditions. Of course, quality in this context refers to media quality, yet unfavorable network conditions may indeed affect the educational quality of such online classes. Unfortunately, there are so many factors that can degrade network conditions during real-time online education. It only makes matters worse when resource-demanding dynamic multimedia – and not just static slides – is being shared, such as the introduction of the usage of certain technical tools via a camera. Network impairments during the different types of content sharing may have a severe effect on online education. Yet, throughout longer portions of online lectures with shared multimedia, students may adapt to smaller extents of such impairments.

In the context of home office, a notable percentage of online activities happen in real time. Probably the most commonly known form of such real-time activities is the online meeting. In online meetings, content sharing is relatively frequent. In most cases, the shared content is a sequence of slides, but other contents may be shared as well, such as a video or the window of a specific application, or even the entire screen. However, when such action is started, the content is not necessarily available instantaneously to the other participants of the meeting. The amount of this delay may depend on a variety of factors, like the type of the content and the associated bandwidth requirements. The initial delay of content sharing may not only affect user experience, but in a professional context, it may also cause further undesirable effects – for example, missing important information related to the subject at hand. Moreover, when a video is shared, playback may be subject to the rubber band effect (i.e., playback is not uniform in terms of frame speed), especially right after it becomes available to the observers.

Tushig Bat-Erdene, Yazan N. H. Zayed, Xinyu Qiu, Ibrar Shakoor, Achref Mekni, Peter A. Kara and Laszlo Bokor are with the Budapest University of Technology and Economics, Budapest, Hungary. E-mail: {bat-erdene.tushig, yzayed, qiuxinyu, ibrarshakoor, achrefmekni}@edu.bme.hu, {kara, bokorl}@hit.bme.hu

Peter A. Kara and Maria G. Martini are with Kingston University, London, United Kingdom. E-mail: {p.kara, m.martini}@kingston.ac.uk

Aniko Simon is with Sigma Technology, Budapest, Hungary. E-mail: aniko.simon@sigmatechnology.se

Manuscript received December 26, 2021.

DOI: 10.36244/ICJ.2022.2.8

Therefore, the Quality of Experience (QoE) of online education and online meeting platforms in general has become more relevant than ever. In this paper, we address the contexts of online education and online meetings through a series of surveys and subjective studies. The work we carried out over the past two years covers various topics that are relevant to the phenomena mentioned above. We particularly focused on the degradation of video content sharing QoE via network impairments, the rubber band effect, initial loading delay and frame rate variation. The surveys not only provide useful insights into the investigated topics, but also supported the experimental configurations of the subjective studies.

Regarding the online meeting platform of choice, one could repeat a certain experiment over the most commonly used platforms to carry out an exhaustive performance comparison. Instead, we used a single platform for all the tests, and the surveys were designed for that specific platform as well. Hence, the primary focus of the work was on the investigated research questions of the QoE-related phenomena and not on the capabilities of various meeting platforms. For our surveys and tests, we selected Microsoft Teams, since it is the default meeting platform of the institutions of all the authors of this paper.

As for the methodology of the subjective tests, the experiments were always implemented as a Teams meeting between the test participant and the conductor of the test. Having multiple test participants simultaneously is typical in online education and meetings; however, in order to avoid any issue or irregularity that may originate from such circumstance and thus distort the obtained results, in the scope of this paper, multi-participant scenarios are not addressed. Additionally, our approach of having only a single test participant in a call enabled test stimulus randomization; each test participant was provided a unique sequence to assess.

The students, lecturers and employees of the private sector who completed our surveys and participated in our tests reside in many different countries, including (but not limited to) Austria, China, France, Germany, Hungary, Italy, Jordan, Poland and the United Kingdom. Information on demographics (age and gender) is provided in the *Results* subsection of each subjective study. A total number of 303 individuals completed our surveys and 88 individuals participated in our studies. As there were 4 subjective studies, the results of each study were based on the ratings of either 20 or 24 test participants. While this may be perceived as a limitation of the work, statistically significant rating differences were achieved for multiple experiments nonetheless. Although the same statement is not applicable to other tests, the collected data initiates novel research questions for future scientific efforts in the field of QoE.

The remainder of this paper is structured as follows: Section II reviews the scientific literature related to both primary topics. Sections III and IV present our surveys and subjective studies on online education and on online meetings, respectively. Section V concludes the paper and highlights the potential continuations of the addressed topics.

II. RELATED WORK

Studies related to online education have boosted their relevance significantly during the past years due to the ongoing global pandemic. Many works particularly address online education separately from the perspectives of students and teachers [4]–[6], compare the most frequently used online platforms [7]–[10], and investigate the phenomenon of e-learning [11]–[13]. The work of Husniyah *et al.* [4] concludes that numerous teachers avoid real-time online engagements in order to elude the effects of unfavorable network conditions, and that alternative solutions are often preferred (e.g., pre-recorded lectures). On the other hand, the publications of Mukhtar *et al.* [14] and Dhawan [15] call attention to educational issues caused by the lack of immediate feedback – applicable to both real-time and asynchronous education. Yet the results of Barbour *et al.* [11] indicate that students are likely to prefer evading real-time communication (i.e., via microphone and/or camera) and communicate via chat instead. The research of Coman *et al.* [5] states that the most significant challenge regarding online education is the threat of potential technical issues – which may impose particular learning problems in the context of early childhood education [16] – and urges institutions to develop training sessions for teachers. Scarlat *et al.* [17] also emphasize this redirection of efforts. The recent papers of Chen *et al.* [18], [19] highlight that students focus more on the quality of real-time interaction since the outbreak of the pandemic, and that personal factors do not directly influence satisfaction. However, personal factors do correlate with motivations connected to learning under the circumstances of the current era, as signified by the work of Nurhopipah *et al.* [20]. While the "sense of presence" in cutting-edge research is primarily applied to novel glasses-free 3D technologies [21], [22], the paper of Chessa *et al.* [23] addresses this topic in the contexts of conventional online education and virtual-reality-assisted training, the findings of which correlate with the theoretical framework of Shea *et al.* [24]. On a more general level, Bao [25] concludes the essential need for the five principles of appropriate relevance, effective delivery, sufficient support, high-quality participation and contingency plan preparation in large-scale online education; and Prasetyo *et al.* [26] characterize the relevant constructs of system quality, information quality, perceived usefulness, perceived ease of use, user interface, behavioral intentions and actual use. Finally, as education in many parts of the world is now slowly reverting back from virtual to presential formats, post-lockdown studies are continuously emerging, such as the work of Kassahun [27].

The performance of video conferencing platforms [28]–[35] – regardless of their usage – is relevant to the investigated contexts. Multimedia QoE, in general, is fundamentally based on image quality, resolution and frame rate [36]–[38], but it is also affected by a plateau of other aspects, phenomena and effects, such as the memory effect [39], the contrast effect [40] and the labeling effect [41]. The effect of the initial delay on the QoE of real-time video streaming [42]–[44] can be looked at as one of the major motivators for modern adaptive streaming solutions. However, it is not only the system that

is adaptive, but the users as well, since works on the topic of QoE over time [45]–[47] indicate that personal tolerance may evolve against quality degradation – if the extent of degradation is not severe enough to render the specific use case useless, of course. Thus far, according to the best knowledge of the authors, the adaptation to quality degradation caused by network impairments has not been studied in the contexts of online meeting platforms yet, especially in the context of real-time educational multimedia. The same is applicable to the rubber band effect of slide sharing, the initial loading times of multimedia contents, and the frame rate values and fluctuations of such videos.

III. Surveys and Subjective Studies on Online Education

A. Survey on Student Experience

The survey focused on the subjective perception of online education via Teams. The questions utilized a 7-point symmetric rating scale [48] to assess satisfaction regarding the investigated aspects (*very unsatisfied*, *unsatisfied*, *slightly unsatisfied*, *neutral*, *slightly satisfied*, *satisfied* and *very satisfied*).

1) Questions: The questions of the survey were related to performance, addressing both delay-sensitive (i.e., real-time) and delay-tolerant (i.e., downloading), short-term (e.g., taking part in an oral exam) and long-term (e.g., participating in a lecture) tasks. The students who completed our survey had to assess the following: (Q1) reliability of downloading study-related materials, (Q2) speed of downloading study-related materials, (Q3) reliability of uploading study-related materials, (Q4) speed of uploading study-related materials, (Q5) taking part in oral exams, (Q6) being updated by lecturers regarding tasks and deadlines (i.e., information sharing outside the lectures), (Q7) sharing content during real-time activities (e.g., a student presentation during lecture), (Q8) taking part in lectures and (Q9) the comparison of online education to contact classes (e.g., the rating *very satisfied* indicates that online classes are much better from the perspective of the student). We also asked students about the weekly number of disconnections they suffer during online lectures.

2) Results: The survey was completed by 46 university students (25 B.Sc., 19 M.Sc. and 2 Ph.D. students). 80.4% of the students connects to the real-time lectures via laptops, and 93.5% uses wireless Internet connection. The results are shown in Table I, using the numerical equivalent of the scale (e.g., 3 corresponds to *very satisfied*). The questions related to downloading and uploading (Q1–Q4) received favorable ratings, and similar observations are applicable to information sharing outside lectures (Q6). The short-term activity of taking part in oral exams (Q5) obtained comparably lower ratings, and content sharing (Q7) and lectures (Q8) received the lowest ratings, particularly the latter. This may be connected to the fact that the students who completed this survey suffer from disconnections during lectures nearly 2 times per week on average. As for the preference (Q9), 58.7% of the students prefer online education – despite the potential issues – and only 13.04% chose regular contact classes. Generally, the results are indeed favorable, but the pitfalls of real-time activities – particularly lectures, which are the heart and soul of online education – are notable.

TABLE I
Results (in %) of the survey on student experience.

	-3	-2	-1	0	+1	+2	+3
Q1	0	4.35	2.17	19.57	21.74	32.61	19.57
Q2	0	2.17	2.17	19.57	15.22	41.3	19.57
Q3	0	0	2.17	32.61	13.04	41.3	10.87
Q4	0	0	2.17	26.09	21.74	43.78	15.22
Q5	0	2.17	8.7	30.43	10.87	36.96	10.87
Q6	2.17	0	6.52	26.09	15.22	34.78	15.22
Q7	0	2.17	4.35	34.78	30.43	19.57	8.7
Q8	0	4.35	13.04	30.43	26.09	13.04	13.04
Q9	4.35	4.35	4.35	28.26	23.91	19.57	15.22

B. Survey on User Interface Settings

The survey enquired about the user interface settings of Teams during lectures, particularly whether students have the list of participants or the chat open.

1) Question: The survey included a single multiple-choice question, asking about the preferred state of the right side of the application screen during lectures where the lecturer shares educational content. The students had to choose either *list of participants*, *chat* or *none*. In this context, the answer should reflect the state that is personally preferred, without special events. A special event can be that at one point of the lecture, the lecturer posts the link of a website or the title of a paper on the chat, which can prompt students to temporarily open the chat if it is not open by default.

2) Results: The survey was completed by 33 university students (16 B.Sc., 14 M.Sc. and 3 Ph.D. students). The dominant preference was the list of participants (18), followed by chat (14), and only a single student stated that none of these is preferred. We considered extending the survey with more test participants, however, the obtained results already provided sufficient evidence to support the experimental setup of the subjective studies, particularly the one on the effects of network impairments.

C. Survey on General Content Sharing Options

The survey addressed the personal preference of lecturers regarding the different possible options for general content sharing via Teams.

1) Question: The survey included a single multiple-choice question, asking about the preferred method of content (any educational content) sharing via Teams. The possible answers were the following: Desktop sharing (fullscreen); Window sharing; Presentation sharing directly via Teams; MS Whiteboard; Freehand; Webcam; Audio only.

2) Results: The survey was completed by 74 university lecturers from institutions where Teams is the default platform for online education. The dominant preference was window sharing (38), followed by fullscreen desktop sharing (26). Sharing directly via Teams (1), MS Whiteboard (2), webcam (7) and audio-only (1) methods were also preferred by some, while no preference was registered for Freehand. However, this survey covers all intents of content sharing in online education.

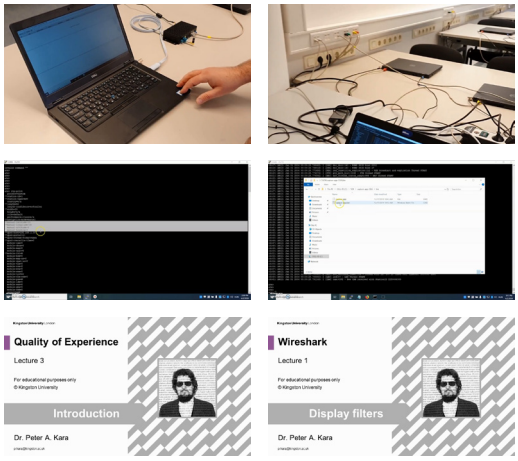


Fig. 1. Source video sequences of the study on the effects of network impairments.

Therefore, in order to have results more focused on video sharing – which is the topic of the subjective studies – we repeated the survey with the appropriate alterations.

D. Survey on Video Content Sharing Options

The survey addressed the personal preference of lecturers regarding different options for video content sharing via Teams.

1) Question: The survey included a single binary question, asking about the preferred method of video sharing (i.e., sharing the playback of a video file) via Teams. The possible answers were the following: Desktop sharing (fullscreen); Window sharing.

2) Results: The survey was completed by 48 university lecturers from institutions where Teams is the default platform for online education. The dominant preference was window sharing (33), followed by fullscreen desktop sharing (15). Therefore, in the subjective studies, we utilized the option of window sharing.

E. Subjective Study on the Effects of Network Impairments

The task of the test participants was to assess the quality of content sharing, comparing the perceived audiovisual quality of artificially degraded test cases (with added delay and packet loss values) to reference videos (i.e., where the transmission was not degraded additionally).

1) Experimental Setup: Based on the answers collected by the survey on content sharing options, the content was shared via application window (i.e., a video player running in fullscreen mode). This was also very convenient for using the same computer for setting the parameters of network conditions, without the test participants noticing it (which could have been an issue during single-screen fullscreen sharing).

For the pair comparison, a 5-point Degradation Category Rating (DCR) scale [49] was used, which registers both perceptibly and annoyance. The scores were recorded on the meeting chat. It was deemed a life-like scenario to have the chat open while viewing the shared content, since according to

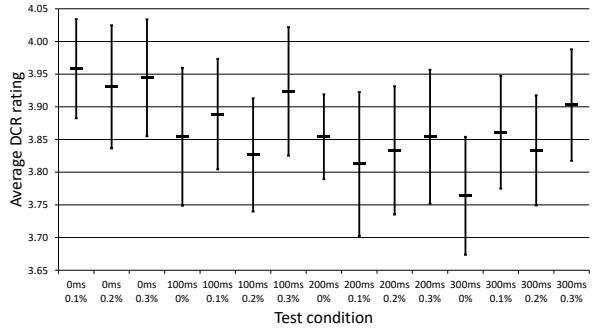


Fig. 2. Average of the DCR scores obtained for the study on the effects of network impairments.

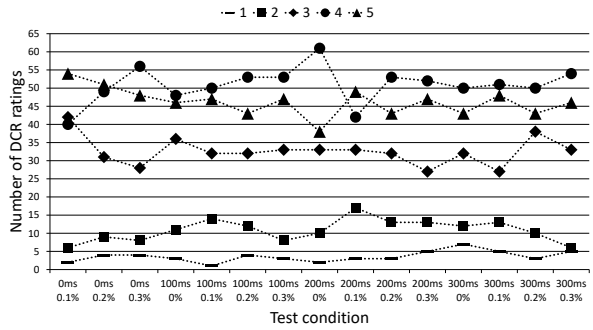


Fig. 3. Rating distribution of the study on the effects of network impairments.

the survey on user interface settings, the majority of students have either the participant list or the chat open during lectures.

The test sequences were separated by 5-second grey separation screens. After a stimulus pair, the test participant had to register the score during the separation screen. The network conditions were also changed during this period, according to the values defined by the test conditions.

The degradation of Quality of Service (QoS) parameters – the impacts of which are relevant to all forms of transmission contexts [50] – were simulated via Clumsy[1]. For both delay and packet loss, 4 values were chosen: 0 ms, 100 ms, 200 ms and 300 ms; 0%, 0.1%, 0.2% and 0.3%, respectively. There was a total of 16 test cases, as every single combination was included. Evidently, the condition with 0 ms delay and 0% packet loss was the reference. As in all our studies, the test condition order was uniquely randomized for each and every test participant, and the test conditions were applied to every single source sequence.

The 6 source video sequences were provided by the academic co-authors (i.e., university lecturers) of the paper. They are all 30-second long, 30-fps, 720p (1280×720 pixels) videos. As for the contained educational material itself, they were selected with the aim of content diversity. Sequence `A` and `B` were recorded by a hand-held device (i.e., smart phone), `C` and `D` were captured by a desktop recording software – the entire desktop was recorded – and `E` and `F` were rendered. In all of these videos, the lecturer continuously talks throughout

[1] http://jagt.github.io/clumsy/index.html

the entire duration. In sequence A, the lecturer introduces a specific device connected to a laptop, while holding the camera in a given position and angle (with minimal hand tremor) with one hand, and making hand gestures (i.e., pointing at certain devices) with the other. In sequence B, the camera is moved around in a laboratory, and it automatically refocuses when necessary. In sequence C, a piece of program code in a command terminal is explained, and some lines are highlighted by the lecturer. In sequence D, the lecturer navigates between different folder windows on the computer, and a program is launched. In sequence E, a fullscreen slideshow is presented, containing a single change between the slides. In sequence F, the same slide is shown throughout the entire video. A demonstrative screenshot of each source video sequence is shown on Figure 1.

2) Results: A total of 24 test participants completed the subjective tests (16m, 8f, avg. age 23). All of them were active university students (11 B.Sc., 11 M.Sc. and 2 Ph.D. students). 23 used wireless Internet connection and only 1 used a wired connection. Regarding devices, 19 students used laptops, 3 of them connected via a desktop computer, 1 participated through a tablet and 1 through a smart phone.

Figures 2 and 3 show the average and the distribution of the obtained scores, respectively. In the latter, the same rating options are connected over the series of test conditions to increase the quality of data communication. This approach is applicable to all the other cases of data series visualization in this paper as well. For every test condition, 144 ratings were collected, as a single test condition was rated by 24 test participants over 6 source contents. From the 144 ratings, even in the case of 0 ms delay and 0.1% packet loss, only 54 did not report perceivable differences, and the same was 46 for 100 ms delay and 0% packet loss. Regarding toleration, 32.6% reported annoyance to a given extent, but the rest indicated the total lack of irritation caused by quality degradation. In general, the results obtained for the test conditions do not differ significantly; the only statistically significant difference was between 0 ms delay with 0.1% and 0.3% packet loss values and between 300 ms delay and 0% packet loss. As for the source contents, the difference in scene dynamics is well-reflected in the results, as the two videos with slides (E and F) achieved the highest average ratings, 3.99 and 4, respectively. These are followed by the desktop recordings (C and D), at 3.89 and 3.85, respectively, and then by the camera captures (A and B), at 3.71 and 3.73, respectively.

The network parameter values in this experiment were added artificially to the already existing conditions as extra load, and almost every single test participant used wireless connection to take part in the study. Hence, variation over time regarding the real network conditions was possible, which can significantly affect the actual performance. However, this was an intentional decision within the experimental setup, in order to investigate realistic conditions. As for the added packet loss, it was not implemented in a strictly uniform manner (i.e., every n^{th} packet is dropped), thus its effect on content sharing was not deterministic, unlike the straightforward additional delay. The values were selected based on conclusions of the scientific literature [47], [51]–[55] and preliminary testing regarding

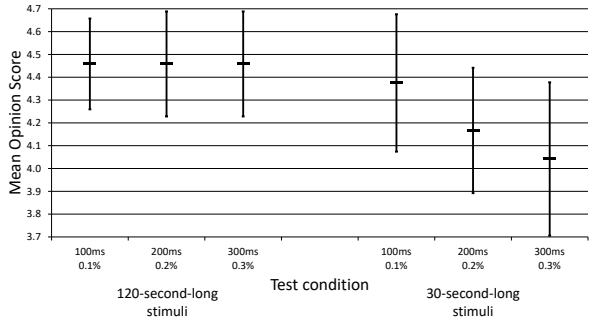


Fig. 4. MOS of the study on the adaptation to network impairments.

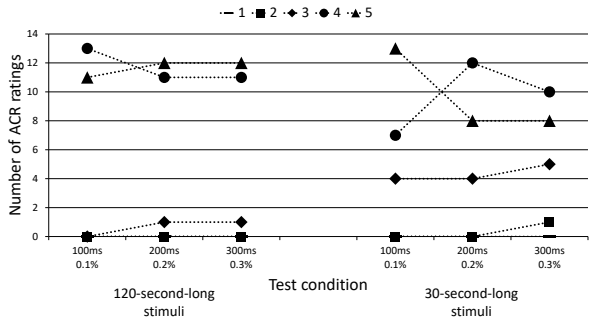


Fig. 5. Rating distribution of the study on the adaptation to network impairments.

just-noticeable differences (JNDs). Furthermore, no error correction was simulated. Finally, the experiment combined the assessment of the perceived video and audio quality. Based on the consistency of the ratings achieved by the different content types, we can assume that visual quality played a more significant role in the evaluation of the overall quality.

F. Subjective Study on the Adaptation to Network Impairments

Similarly to the previously introduced subjective study, the task of the test participants was to assess the quality of content sharing. However, in this study, we addressed the effect of content length and variation as well. Through such, the students' adaptation to suboptimal network conditions was investigated.

1) Experimental Setup: In the tests, the number of test conditions was limited to 3: 100 ms delay with 0.1% packet loss; 200 ms delay with 0.2% packet loss; 300 ms delay with 0.3% packet loss. These 3 test conditions were assessed in 2 test scenarios. In one, only a single 120-second-long video was played. In the other one, 4 30-second-long sequences were shown, separated by 5-second-long separation screens. The overall quality of the stimuli was to be evaluated via a single 5-point Absolute Category Rating (ACR) score [49]. This means that the 4 videos in the second scenario were not to be rated separately, but as a whole; perceived quality was to be averaged. The rationale behind the choice of ACR was that from the perspective of the test participants, DCR ratings are less straightforward to average out.

The study used the same source video sequences as the previous study on the effects of network impairments. However, longer cuts were taken from the same contents in order to satisfy the requirements of the first scenario (i.e., to have 120-second-long videos). Additionally, sequences E and F were not included in the study due to their low variations in visual information. Therefore, we used 4 source contents in total, and thus, the stimuli of the second scenario were always composed of the same 4 videos, in randomized order.

2) Results: A total of 24 test participants completed the subjective tests (14m, 10f, avg. age 23.3). All of them were active university students (12 B.Sc., 9 M.Sc. and 3 Ph.D. students). 23 used wireless Internet connection and only 1 used a wired connection. Regarding devices, 9 students used laptops, 4 of them connected via a desktop computer, 3 participated through a tablet and 8 through a smart phone.

Figures 4 and 5 show the Mean Opinion Score (MOS) and the distribution of the obtained scores, respectively. For the 120-second-long stimuli, the MOS is consistent across all test conditions, signifying adaptation. In fact, all of these values based on the ratings of the 24 test participants are 4.46. Furthermore, there is no deviation at all between the rating distribution of the two test conditions with higher levels of degradation. On the other hand, the results on the 30-second-stimuli indicate the impact of the artificially added delay and packet loss. Although there is no statistically significant difference between the obtained ratings, there is a shift of 0.33 between means, and 0.42 when compared to the results of the 120-second-long stimuli. Furthermore, the rating distribution of the subjective study on the effects of network impairments signifies the perceivable differences for the shorter stimuli, which, in the case of the current study, does not directly translate to ACR ratings. This is due to the fundamental dissimilarities between the scales themselves: the DCR scale serves a dual purpose, while the ACR scale of the same size has uniformly distributed options. Therefore, while differences are, in fact, perceivable, they hardly reach a quality threshold for shorter stimuli, and for longer stimuli, the selected test conditions did not cause any difference whatsoever, accentuating adaptation.

IV. Surveys and Subjective Studies on Online Meetings

A. Survey on Content Sharing in the Private Sector

The survey addressed the personal preference of employees within the private sector regarding the different possible options for content sharing via Teams.

1) Question: Similarly to the survey on content sharing options in the context of online education, this survey included a single multiple-choice question, asking about the preferred method of content sharing via Teams. However, this question was more focused on presentations (i.e., slide sharing), and thus, accordingly, the possible answers were the following: Desktop sharing (fullscreen); Window sharing; Presentation sharing directly via Teams.

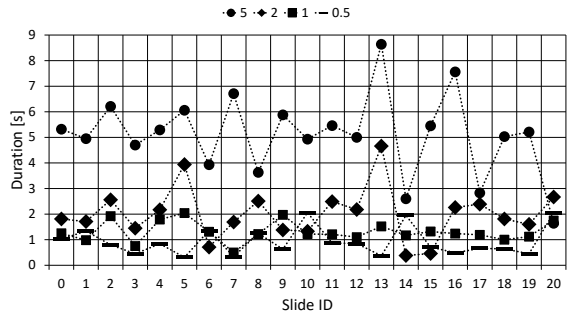


Fig. 6. Results of the study on the rubber band effect of slide sharing. The 4 data series represent the different transition times in seconds.

2) Results: The survey was completed by 61 employees from companies where Teams is the default meeting platform. The dominant preference was window sharing (42), followed by fullscreen desktop sharing (19), while no preference was registered for sharing directly via Teams. These results provided additional support to the experimental configurations; the studies on online meeting utilized application window sharing as well.

B. Preliminary Study on the Rubber Band Effect of Slide Sharing on Online Platforms

Prior to the subjective study on the perceived quality of media (i.e., video) loading times, we carried out an experiment using slides to address the rubber band effect.

1) Experimental Setup: We created 21 slides in total. Every slide had a high-resolution image (equivalent of a 720p video frame) as background – greatly varying with regard to spatial complexity – and a large number in the middle of the slide, going from 0 to 20. The experiment was to share the slides between two clients (i.e., computers) in an online meeting – using Teams as in all of the tests – in order to measure the potential rubber band effect. For this, timed slideshows were used with 5-second, 2-second, 1-second and 500-millisecond transition times.

At first glance, the selected transition times may seem unrealistic in practice. Indeed, changing rapidly between slides during a university lecture or a presentation in the private sector is definitely not the most common practice. However, it does make sense and may serve various purposes. For example, it can be used as a tool to make a point, having only one large image per slide; it is not vital to carefully examine the images themselves. This may be appropriate to emphasize that there are so many examples or use cases to a specific topic; the presenter or lecturer only speaks one word or technical term per slide (1 or 2 seconds per slide is easily realistic in such case). Another example is a "manual" animation, in which the slides behave as frames (even 0.5 seconds per slide may be realistic, depending on the content). Such approach is more than adequate to exhibit the progress related to the content, and the presenter may easily revert the direction of progress (i.e., moving back and forth between slides).

As a first step, both computers were validated in terms of performance. Both passed the validation, as during the local

playbacks, the slides were transitioned with the correct time slots (with negligible deviations).

2) Results: Figure 6 shows the results. The test was repeated and similar data was obtained. Beside the apparent rubber band effect, the total time needed to reach the end of the slideshow was sometimes longer as well: instead of 105, 42, 21 and 10.5 seconds, it needed 107.03, 42.16, 27.57 and 19.43 seconds, respectively. Technically, the more frequent the transition was, the longer the total duration became. Furthermore, certain patterns are detectable among the transition duration trends, which are due to the differences in spatial information between adjacent slide backgrounds. Moreover, less-frequent, higher-duration transitions resulted higher jitter, as exhibited by Figure 6. Yet, it needs to be noted that positive alterations (i.e., more time is needed for the transition) are subsequently balanced out by negative alterations better, which issued the lower deviation regarding the total time. Moreover, while the absolute value of the jitter measured for higher-duration transitions may be greater (e.g., more than 8.5 seconds for a five-second transition), the percentage-wise fluctuation for lower-duration transitions is notably higher (e.g., more than 2 seconds for a half-second transition). A more in-depth analysis of the phenomenon is required to adequately address the open questions related to the topic.

C. Subjective Study on the Overall Perceived Quality of Media Loading Times

The task of the test participants was to report the amount of initial content delay (i.e., how much they missed from the beginning of the video) and to rate the overall quality.

1) Experimental Setup: Following the concept of the previously introduced study, in every single video, a large number in the middle of the frame counted the seconds passed since the video was started. The videos were 21 seconds long, as the counter went from 0 to 20. The task of the test participant was to report the number first visible when the video image became available and to evaluate the overall quality. Again, during the test, the participant had to report 2 numbers: (i) the first perceivable number when the video content appeared on the screen and (ii) the subjective score of the overall quality, rated via a 5-point ACR scale. Hence, 2 numbers were reported per test stimulus. The data was reported verbally, via Teams. The first number was provided by the test participant when the content became visible, and the second one was registered at the end of the stimulus. It needs to be noted that reporting the scores verbally was quite viable as the stimuli contained no audio. Furthermore, the test participant was instructed to take everything into account, including the behavior of counter (i.e., uniformity of counter progression) when assessing the overall quality.

The test conditions of the experiment were the combinations of different video resolutions and content structures. As there were 5 content structures displayed at 3 resolutions, the total number of test conditions was 15. The 3 resolutions were 480p, 720p and 1080p. We considered using 2160p as well, but according to a small one-question survey on 2160p in online education (a binary question whether the individual uses 2160p resolution for content sharing via Teams or not), less than 5% of the respondents use such high resolution. Therefore, we considered it to be unrealistic in the scope of the experiment. Some may argue that 480p might also be deemed irrelevant in this day of age, yet certain older but relevant professional materials may not be available in higher resolution.

TABLE II
FIRST VISIBLE NUMBER IN THE TESTS WITH OVERALL QUALITY

	480p	720p	1080p	a	b	c	d	e
0	11	7	3	7	4	3	2	5
1	85	92	95	52	56	55	56	53
2	3	1	2	1	0	2	1	2
6	1	0	0	0	0	0	1	0

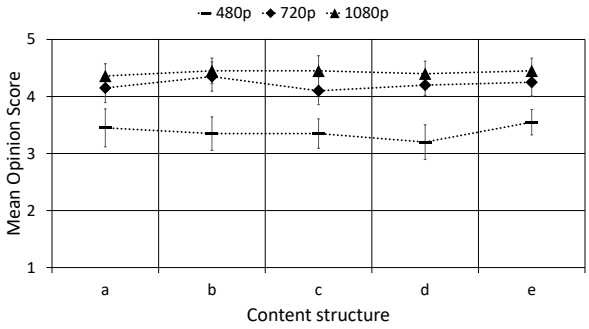


Fig. 7. MOS for overall quality.

The 5 content structures (i.e., the alternations of the background behind the counter) of the experiment were the following: (a) static black screen, (b) a single video sequence without cuts, (c) sequences change every 3 seconds, (d) sequences change every 2 seconds and (e) sequences change every second. The adjacent sequences were selected in a manner to have as much difference as possible, with regards to Spatial Information (SI) and Temporal Information (TI), scene dynamics, camera motions, content types (i.e., camera-captured or rendered), etc.

The source video contents were selected from the Xiph.org Test Media[2] collection. The sequences typically alternated between rendered (e.g., Big Buck Bunny) and camera-captured (e.g., Netflix's El Fuente) scenes, but the aforementioned parameters were taken into consideration as well.

2) Results: The experiment involved 20 test participants (9m, 11f, avg. age 21.6). 15 used wireless connection to access the Internet and 5 connected via Ethernet cable. 14 used the Teams desktop application, 3 used the mobile application and 3 used a web browser. Table II shows the results for the initial loading times (i.e., how many times test participants perceived a given counter number first). At first glance, the results for each category seem pretty much the same, with 1 being dominant. However, Pearson's chi-squared test indicates statistically significant differences: for resolution categories, in the case of 480p and 1080p ($p < 0.01$); for structure categories, in the cases of a and b ($p = 0.04$), a and c ($p = 0.03$), a and d ($p < 0.01$), b and c ($p = 0.03$), b

[2]https://media.xiph.org/

and d ($p = 0.03$), b and e ($p = 0.03$), d and e ($p = 0.03$). A clear conclusion that can be drawn here is that higher resolutions result in higher initial media loading times – which is, of course, expected – and thus, greater differences in resolution result in greater differences between such values. While the adjacent resolutions do not differ on a statistically significant level, the lowest and the highest do. Furthermore, the variations in the spatial and temporal complexities of the transmitted multimedia content may also have a significant impact on the loading times.

Figure 7 shows the MOS values of the tests. Resolutions 720p and 1080p performed similarly across every content structure; although there was a clear preference towards 1080p, the differences were not statistically significant. Regarding the 480p stimuli, the ratings were significantly worse. In order to investigate the cause of the obtained results, the experiment was repeated with the same number of test participants, but with individual quality aspects.

D. Subjective Study on the Separate Quality Aspects of Media Loading Times

The aim of this study was to address the separate quality aspects of the previous experiment.

1) Experimental Setup: In this study, the overall quality was separated into 3 aspects: (i) the visual quality of the video, (ii) the frame rate and (iii) the uniformity, the behavior of the counter. These were all rated via the same ACR scale. Evidently, in the these tests, 4 numbers were reported per test stimulus.

2) Results: The experiment involved 20 test participants (12m, 8f, avg. age 21). 15 used wireless connection to access the Internet, 5 connected via Ethernet cable. 11 used the Teams desktop application, 5 used the mobile application and 4 used a web browser. Table III shows the results for the initial loading times. Although the categories follow a similar pattern, there are, in fact, statistically significant differences: similarly to the results of the previous experiment, for resolution categories, in the case of 480p and 1080p ($p < 0.01$); for structure categories, in the cases of a and b ($p = 0.01$), a and c ($p < 0.01$), a and d ($p < 0.01$), b and c ($p < 0.01$), b and d ($p < 0.01$), b and e ($p < 0.01$), d and e ($p < 0.01$). This is extended by a and e ($p < 0.01$), c and d ($p < 0.01$), c and e ($p < 0.01$). Technically speaking, this means that the results of every single structure category is significantly different from the results of every other structure category. The conclusions that can be drawn from these results – particularly regarding resolution – are analogous to the findings presented earlier.

Figure 8 shows the MOS values of the tests. Compared to the ratings obtained on overall quality, visual quality was assessed in a similar manner, but there were statistically significant differences between 720p and 1080p as well.

In the case of perceived frame rate, content structures a and b were not distinguished with respect to resolution; the plain-black and the single-scene stimuli caused either no degradations in frame rate or applied to every resolution at a similar extent. However, for the other videos with content switches (i.e., cuts), higher resolutions were penalized more, especially in the case of structures d and e.

TABLE III
FIRST VISIBLE NUMBER IN THE TESTS WITH QUALITY ASPECTS

	480p	720p	1080p	a	b	c	d	e
0	16	13	9	9	6	13	9	1
1	73	74	76	46	45	42	40	50
2	9	11	10	4	8	2	9	7
3	0	2	4	0	0	2	2	2
5	0	0	1	1	0	0	0	0
6	1	0	0	0	0	1	0	0
10	1	0	0	0	1	0	0	0

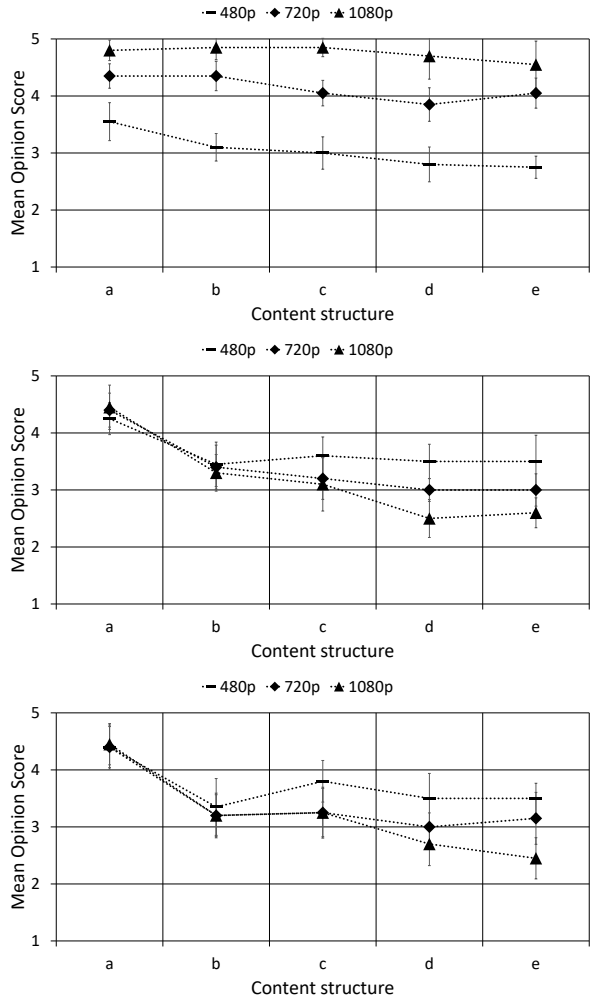


Fig. 8. MOS for visual quality (top), frame rate (middle) and counter assessment (bottom).

The subjective assessment of the counter was analogous to the evaluation of the frame rate, clearly indicating the connection between the two. It is important to note that it is technically possible to have video content sharing on an online meeting platform where the frame rate fluctuates but the behavior of the counter remains mostly uniform. In our experiment, the uniformity of the counter was affected similarly to the frame rate.

The topic of this study is additionally investigated by the following survey. While these results already signify that

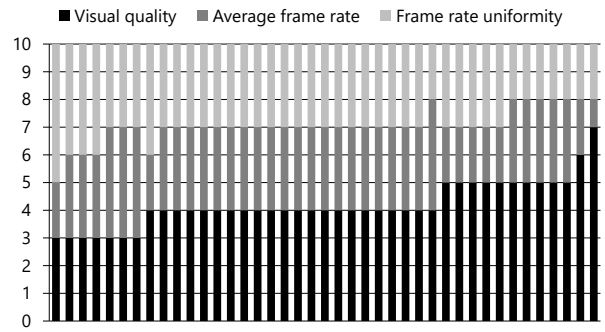


Fig. 9. Preference point distribution between visual quality, average frame rate and frame rate uniformity.

visual quality contributes the most to the overall perceived quality of content sharing in online meetings, these aspects were directly assessed on the level of personal preference.

E. Survey on Visual Quality, Average Frame Rate and Frame Rate Uniformity

The survey addressed the personal priorities and preferences between visual quality, average frame rate and frame rate uniformity.

1) Questions: The first task within the survey was to distribute 10 points among visual quality, average frame rate and frame rate uniformity. Higher points reflected higher personal preferences. Any combination was permitted (including giving 10 points to one aspect and 0 to the others), and only integers were to be used. The two other questions directly addressed the investigated preferences with 3 options each. One asked about visual quality and frame rate (*better visual quality but lower average frame rate and frame rate uniformity*; *worse visual quality but higher average frame rate and frame rate uniformity*; *equal preference*) and the other one asked about average frame rate and frame rate uniformity (*higher average frame rate but lower frame rate uniformity*; *lower average frame rate but higher frame rate uniformity*; *equal preference*).

2) Results: The survey was completed by 41 individuals. The average preference points for visual quality, frame rate and frame rate uniformity was 4.2, 2.85 and 2.95, respectively. The distribution of the points for each individual who answered the survey is shown on Figure 9. These results are analogous to the findings shown on Figure 8; visual quality contributes the most to the overall quality, while average frame rate and frame rate uniformity are equally lower priority. The most common distribution is 4/3/3, which applies to the preference of 20 out of the 41 individuals.

Regarding the two other questions, the results are the following: 21 voted for better visual quality, 6 preferred frame rate and the preference was equal for 14 individuals; 9 voted for higher average frame rate, 19 preferred higher frame rate uniformity and the preference was equal for 13 individuals.

While the results of first question are somewhat analogous to the preference point distribution, the second question seems to contradict the distribution at first glance. In the distribution, points on average frame rate and frame rate uniformity were balanced, yet the results of the second question clearly favor frame rate uniformity over average frame rate. However, the point distribution task covered visual quality as well, and it was not necessarily a simple task to correctly indicate the relations between all 3 aspect at the same time, taking into consideration the assigned priority proportions. On the other hand, the questions did not take the magnitude, the weight of preference into consideration, enabling smaller differences with regard to personal priorities to be indicated.

Additionally, note that in the subjective study on the investigated aspects, the test participants provided ratings based on what they experienced in the scope of the experiment, while in the survey, responses were solely based on existing prior experience. Of course, in the subjective study, the test participants were fundamentally influenced by prior experience as well. The effect of such influence is also worth studying in the future.

V. CONCLUSION

In this paper, we presented our surveys and studies on the QoE of content sharing in online education and on online meeting platforms. The results of the individual surveys and studies support each other in terms of experimental configuration and aid the deeper understanding of the investigated phenomena. Furthermore, the obtained ratings for the different subjective studies on the same topic of interest draw similar conclusion.

The quality ratings in the context of online education indicate an excellent level of adaptation to impairments. However, as the degradation was set to be around the extent of JND, such adaptation may not be applicable to more severe impairments. Nonetheless, the topic of QoE over time is greatly relevant to online education – due to the potentially longer contents of educational multimedia – and further research may benefit the modeling of adaptation.

We conclude that the personal preference related to the visual quality and frame rate on content sharing via online meetings is the opposite of the trends of modern real-time Video-on-Demand (VoD) services. In the recent years, the majority of VoD platforms started utilizing Dynamic Adaptive Streaming over HTTP (DASH) – also known as MPEG-DASH – which may sacrifice visual quality by using lower-quality segments to ensure playback fluency. Our results gathered by both the subjective study and the survey indicate that DASH-like trade-offs (i.e., compromise regarding visual quality) are not necessarily beneficial to content sharing via online meeting platforms.

The results presented in the paper also highlight that there are statistically significant differences between resolutions (480p and 1080p) with regard to initial delay, and this is also applicable to the different content structures. Practically, as expected, contents with lower resolution initiate playback faster while being shared via an online meeting platform. Regarding the video content itself, higher TI values (i.e., greater differences between adjacent video frames) – particularly in the beginning of the video – may result higher initial delays. The obtained ratings also indicate notable differences

with regard to frame rate uniformity on the basis of content structure, some of which are statistically significant.

While the paper did introduce a generous amount of research effort, there is still quite a lot of additional work to be done. First of all, the content characteristics of educational multimedia fundamentally affect the perception of quality impairments – as also indicated by the obtained consistent results. Our studies were limited to 3 archetypes, but there are many more to investigate, such as writing and drawing on a board. The work on adaptation should include wider varieties of content duration, and repeated impairment patters [47] should be addressed as well. Future research efforts should directly consider the SI and TI values of the investigated contents when studying the impacts of the rubber band effect (i.e., the stimuli should be created along a fine-grained SI/TI matrix). Additionally, the phenomenon of frame freezing in the contexts of both online education and online meetings is a relevant, yet underinvestigated issue, the addressing of which could benefit the understanding of both single-event scenarios and QoE over time. Finally, related studies should separately address the various technical options for connecting to the Internet while participating in such experiments, and data clustering based on the capabilities of the user endpoints is also advised, as online meeting platforms may optimize differently for different devices.

A particular limitation of this work is the number of test participants. Although a total of 391 individuals were recruited for the research efforts, this total was spread among 4 subjective studies and 6 surveys. Each of the experiments in the context of online education involved 24 test participants, and this number was 20 for online meetings. The ITU and the VQEG recommend a minimum of 15 [48] and 24 [56] test participants, respectively. Accordingly, many published QoE experiments are of this scale. However, having more test participants may greatly contribute to the statistical strength of the results (e.g., the same rating deviation would result a smaller confidence interval). As Brunnström and Barkowsky conclude [57], going below 24 test participants relies on low rating deviation. In the future, extensive studies of the investigated topics should aim to recruit more test participants.

ACKNOWLEDGMENT

The scientific efforts leading to the results reported in this paper were supported by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme. The authors would like to thank the many individuals who participated in our surveys and studies.

REFERENCES

[1] B. Ferster, Sage on the screen: *Education, media, and how we learn*. JHU Press, 2016.

[2] H. Wehberg, "Some recent developments in the educational film field," *The Journal of Educational Sociology*, vol. 12, no. 3, pp. 163–166, 1938, DOI: 10.2307/2261883.

[3] *Tufts College to Give Radio Lecture Course*. Olympia (WA) Daily Recorder, 1922.

[4] A. Husniyah, "Blended learning in EFL classrooms with slow Internet: Insights from teachers and students," in *The Fourth International Conference on English Across Cultures*, 2018.

[5] C. Coman, L. G. Tîru, L. Meseşan-Schmitz, C. Stanciu, and M. C. Bularca, "Online Teaching and Learning in Higher Education during the Coronavirus Pandemic: Students' Perspective," *Sustainability*, vol. 12, no. 24, 2020, DOI: 10.3390/su122410367.

[6] K. N. Wea and A. D. Kuki, "Students' Perceptions of Using Microsoft Teams Application in Online Learning During the Covid-19 Pandemic," in *Journal of Physics: Conference Series*, vol. 1842, no. 1. IOP Publishing, 2021, DOI: 10.1088/1742-6596/1842/1/012016.

[7] N. Cavus and D. Sekyere-Asiedu, "A comparison of online video conference platforms: Their contributions to education during COVID-19 pandemic," *World Journal on Educational Technology: Current Issues*, vol. 13, no. 4, pp. 1162–1173, 2021, DOI: 10.18844/wjet.v13i4.6329.

[8] S. Dash, S. Samadder, A. Srivastava, R. Meena, and P. Ranjan, "Review of Online Teaching Platforms in the Current Period of COVID-19 Pandemic," *Indian Journal of Surgery*, pp. 1–6, 2021, DOI: 10.1007/s12262-021-02962-4.

[9] J. Gottfried, L. DeLancey, C. Watwood, and A. Hardin, "Virtual conferencing and meeting systems: Resources for online connections," *College & Research Libraries News*, vol. 76, no. 2, pp. 98–101, 2015, DOI: 10.5860/crln.76.2.9265.

[10] R. Jayaraman, V. Jothiswaran et al., "Web-based platforms for virtual learning," *Biotica Research Today*, vol. 2, no. 5, pp. 184–186, 2020.

[11] M. K. Barbour, "Real-time virtual teaching: Lessons learned from a case study in a rural school," *Journal of Interactive Online Learning*, vol. 19, no. 5, pp. 54–68, 2015, DOI: 10.24059/olj.v19i5.705.

[12] M. Jibrin, M. Musa, and T. Shittu, "Effects of internet on the academic performance of tertiary institutions' students in Niger State, Nigeria," *International Journal of Education, Learning, and Training*, vol. 2, no. 2, pp. 57–69, 2017.

[13] A. LoBue, "Low-Bandwidth Teaching Strategies in Response to the COVID-19 Pandemic," EDREDESIGN, Harvard Graduate School on Education, 2020.

[14] K. Mukhtar, K. Javed, M. Arooj, and A. Sethi, "Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era," *Pakistan journal of medical sciences*, vol. 36, no. COVID19-S4, 2020, DOI: 10.12669/pjms.36.covid19-s4.2785.

[15] S. Dhawan, "Online learning: A panacea in the time of COVID-19 crisis," *Journal of Educational Technology Systems*, vol. 49, no. 1, pp. 5–22, 2020, DOI: 10.1177/0047239520934018.

[16] E. Munastiwi, "Colorful Online Learning Problem of Early Childhood Education During the COVID-19 Pandemic," *Al-Ta lim Journal*, vol. 27, no. 3, pp. 227–235, 2020, DOI: 10.15548/jt.v27i3.6630.

[17] M. M. Scarlat, J. Sun, P. M. Fucs, P. Giannoudis, A. F. Mavrogenis, T. Benzakour, A. Quaile, and J. P. Waddell, "Maintaining education, research and innovation in orthopaedic surgery during the COVID-19 pandemic. The role of virtual platforms. From presential to virtual, front and side effects of the pandemic," pp. 1–6, 2020, DOI: 10.1007/s00264-020-04848-8.

[18] T. Chen, L. Peng, B. Jing, C. Wu, J. Yang, and G. Cong, "The impact of the COVID-19 pandemic on user experience with online education platforms in China," *Sustainability*, vol. 12, no. 18, 2020, DOI: 10.3390/su12187329.

[19] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang, and G. Cong, "Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic," in *Healthcare*, vol. 8, no. 3. Multidisciplinary Digital Publishing Institute, 2020, DOI: 10.3390/healthcare8030200.

[20] A. Nurhopipah, I. Nuraida, and J. Suhaman, "Exploring Indirect Aspects in Motivation and Academic Achievement During The Pandemic," *Journal of Education, Teaching and Learning*, vol. 6, no. 2, pp. 163–168, 2021, DOI: 10.26737/jetl.v6i2.2590.

[21] X. Zhang, S. Braley, C. Rubens, T. Merritt, and R. Vertegaal, "LightBee: A self-levitating light field display for hologrammatic telepresence," in *CHI Conference on Human Factors in Computing Systems*, 2019, DOI: 10.1145/3290605.3300242.

[22] A. Cserkaszky, A. Barsi, Z. Nagy, G. Puhr, T. Balogh, and P. A. Kara, "Real-time light-field 3D telepresence," in *7th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2018, DOI: 10.1109/euvip.2018.8611663.

[23] M. Chessa and F. Solari, "The sense of being there during online classes: analysis of usability and presence in web-conferencing systems and virtual reality social platforms," *Behaviour & Information Technology*, vol. 40, no. 12, pp. 1237–1249, 2021, DOI: 10.1080/0144929x.2021.1957017.

[24] P. Shea and T. Bidjerano, "Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education," *Computers & Education*, vol. 52, no. 3, pp. 543–553, 2009, DOI: 10.1016/j.compedu.2008.10.007.

[25] W. Bao, "COVID-19 and online teaching in higher education: A case study of Peking University," *Human Behavior and Emerging Technologies*, vol. 2, no. 2, pp. 113–115, 2020, DOI: 10.1002/hbe2.191.

[26] Y. T. Prasetyo, A. K. S. Ong, G. K. F. Concepcion, F. M. B. Navata, R. A. V. Robles, I. J. T. Tomagos, M. N. Young, J. F. T. Diaz, R. Nadlifatin, and A. A. N. P. Redi, "Determining factors Affecting acceptance of e-learning platforms during the COVID-19 pandemic: Integrating Extended technology Acceptance model and DeLone & Mclean is success model," *Sustainability*, vol. 13, no. 15, 2021, DOI: 10.3390/su13158365.

[27] Z. W. Kassahun, "Exploring post-COVID-19 Lockdown Students' Satisfaction in Ethiopian Higher Education Context in case of some Selected Universities," *Journal of Pharmaceutical Research International*, pp. 293–301, 2021, DOI: 10.9734/jpri/2021/v33i45b32807.

[28] M. M. Archibald, R. C. Ambagtsheer, M. G. Casey, and M. Lawless, "Using Zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants," *International Journal of Qualitative Methods*, vol. 18, 2019, DOI: 10.1177/1609406919874596.

[29] M. Shameem, C. Kumar, and B. Chandra, "Challenges of management in the operation of virtual software development teams: A systematic literature review," in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2017, pp. 1–8, DOI: 10.1109/icaccs.2017.8014695.

[30] A.-M. Suduc, M. Bizoi, and F. G. Filip, "Exploring multimedia Web conferencing," *Informatica Economica*, vol. 13, no. 3, 2009.

[31] C. Topand B. J. Ali, "Customer satisfaction in online meeting platforms: Impact of efficiency, fulfillment, system availability, and privacy," *Amazonia Investiga*, vol. 10, no. 38, pp. 70–81, 2021, DOI: 10.34069/ai/2021.38.02.7.

[32] H. Al-Samarraie, "A scoping review of videoconferencing systems in higher education: Learning paradigms, opportunities, and challenges," *International Review of Research in Open and Distributed Learning*, vol. 20, no. 3, 2019, DOI: 10.19173/irrodl.v20i4.4037.

[33] Y. Lu, Y. Zhao, F. Kuipers, and P. Van Mieghem, "Measurement study of multi-party video conferencing," in *International Conference on Research in Networking*. Springer, 2010, pp. 96–108, DOI: 10.1007/978-3-642-12963-6_8.

[34] B. J. Ali, P. F. Saleh, S. Akoi, A. A. Abdulrahman, A. S. Muhamed, H. N. Noori, and G. Anwar, "Impact of Service Quality on the Customer Satisfaction: Case study at Online Meeting Platforms," in *International journal of Engineering, Business and Management*, vol. 5, no. 2, 2021, pp. 65–77, DOI: 10.22161/ijebm.5.2.6.

[35] N. Berente and J. Howison, "Strategies for success in virtual collaboration: structures and norms for meetings, workflow, and technological platforms," in *Strategies for Team Science Success*. Springer, 2019, pp. 563–574, DOI: 10.1007/978-3-030-20992-6_43.

[36] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hoßfeld, "Impact of frame rate and resolution on objective QoE metrics," in *2010 second international workshop on quality of multimedia experience (QoMEX)*. IEEE, 2010, pp. 29–34, DOI: 10.1109/qomex.2010.5518277.

[37] L. Janowski and P. Romaniak, "QoE as a function of frame rate and resolution changes," in *International Workshop on Future Multimedia Networking*. Springer, 2010, pp. 34–45, DOI: 10.1007/978-3-642-13789-1_4.

[38] L. Janowski, P. Romaniak, and Z. Papir, "Content driven QoE assessment for video frame rate and frame resolution reduction," *Multimedia tools and applications*, vol. 61, no. 3, pp. 769–786, 2012, DOI: 10.1007/s11042-011-0932-9.

[39] T. Hoßfeld, S. Biedermann, R. Schatz, A. Platzer, S. Egger, and M. Fiedler, "The memory effect and its implications on Web QoE modeling," in *2011 23rd international teletraffic congress (ITC)*. IEEE, 2011, pp. 103–110.

[40] M. Schmitt, D. C. Bulterman, and P. S. Cesar, "The contrast effect: QoE of mixed video-qualities at the same time," *Quality and User Experience*, vol. 3, no. 1, pp. 1–17, 2018, DOI: 10.1007/s41233-018-0020-2.

[41] P. A. Kara, A. Cserkaszky, M. G. Martini, L. Bokor, and A. Simon, "The effect of labeling on the perceived quality of HDR video transmission," *Cognition, Technology & Work*, pp. 1–17, 2019, DOI: 10.1007/s10111-019-00582-3.

[42] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: Between the devil and the deep blue sea," in *2012 Fourth International Workshop on Quality of Multimedia Experience*. IEEE, 2012, pp. 1–6, DOI: 10.1109/qomex.2012.6263849.

[43] H. T. Tran, N. P. Ngoc, A. T. Pham, and T. C. Thang, "A multi-factor QoE model for adaptive streaming over mobile networks," in *2016 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2016, pp. 1–6, DOI: 10.1109/glocomw.2016.7848818.

[44] M. Seufert, R. Schatz, N. Wehner, and P. Casas, "QUICker or not? - an Empirical Analysis of QUIC vs TCP for Video Streaming QoE Provisioning," in *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. IEEE, 2019, pp. 7–12, DOI: 10.1109/icin.2019.8685913.

[45] J. Shaikh, M. Fiedler, P. Paul, S. Egger, and F. Guyard, "Back to normal? Impact of temporally increasing network disturbances on QoE," in *2013 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2013, pp. 1186–1191, DOI: 10.1109/glocomw.2013.6825154.

[46] P. Reichl, S. Egger, S. Möller, K. Kilkki, M. Fiedler, T. Hoßfeld, C. Tsiaras, and A. Asrese, "Towards a comprehensive framework for QoE and user behavior modelling," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE,2015, pp. 1–6, DOI: 10.1109/qomex.2015.7148138.

[47] P. A. Kara, W. Robitza, M. G. Martini, C. T. Hewage, and F. M. Felisberti, "Getting used to or growing annoyed: How perception thresholds and acceptance of frame freezing vary over time in 3D video streaming," in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2016, DOI: 10.1109/icmew.2016.7574686.

[48] ITU-T Rec. BT.500: Methodologies for the subjective assessment of the quality of television images. Version in force: BT.500-14 (10/19).

[49] ITU-T Rec. P.910: Subjective video quality assessment methods for multimedia applications. Version in force: P.910 (11/21).

[50] K. Mebarkia and Z. Zsoka, "QoS Impacts of Slice Traffic Limitation," I*nfocommunications Journal*, vol. 8, no. 3, pp. 24–32, 2021, DOI: 10.36244/ICJ.2021.3.3.

[51] J. Xuand, B. W. Wah, "Exploiting just-noticeable difference of delays for improving quality of experience in video conferencing," in *Proceedings of the 4th ACM Multimedia Systems Conference*, 2013, pp. 238–248, DOI: 10.1145/2483977.2484006.

[52] J. Frnda, M. Voznak, and L. Sevcik, "Impact of packet loss and delay variation on the quality of real-time video streaming," *Telecommunication Systems*, vol. 62, no. 2, pp. 265–275, 2016, DOI: 10.1007/s11235-015-0037-2.

[53] T. N. Minhas and M. Fiedler, "Impact of disturbance locations on video quality of experience," in *EuroITV 2011 Workshop: Quality of Experience for Multimedia Content Sharing*. EuroITV 2011 Workshop: Quality of Experience for Multimedia Content Sharing, 2011.

[54] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the Quality of Experience of HTTP video streaming," in *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*. IEEE, 2011, pp. 485–492, DOI: 10.1109/inm.2011.5990550.

[55] P. Perez, J. Macias, J. J. Ruiz, and N. Garcia, "Effect of packet loss in video quality of experience," *Bell Labs Technical Journal*, vol. 16, no. 1, pp. 91–104, 2011, DOI: 10.1002/bltj.20488.

[56] VQEG: Report on the validation of video quality models for high definition video content, 2010.

[57] K. Brunnström and M. Barkowsky, "Statistical quality of experience analysis: on planning the sample size and statistical significance testing," *Journal of Electronic Imaging*, vol. 27, no. 5, 2018, DOI: 10.1117/1.jei.27.5.053013.

Tushig Bat-Erdene is a student of the Budapest University of Technology and Economics. He studies computer engineering in the B.Sc. course of the Faculty of Electrical Engineering and Informatics. His research interests include Internet-of-Things (IoT), data science, quantum computing and cybersecurity.

Yazan N. H. Zayed is a student of the Budapest University of Technology and Economics. He studies electrical engineering in the M.Sc. course of the Faculty of Electrical Engineering and Informatics, Department of Electric Power Engineering. His research interests include electric mobility, smart city applications and renewable energy.

Xinyu Qiu is a student of the Budapest University of Technology and Economics. She studies computer engineering in the B.Sc. course of the Faculty of Electrical Engineering and Informatics, Department of Automation and Applied Informatics. Her research interests include project management and service quality.

Ibrar Shakoor is a student of the Budapest University of Technology and Economics. He studies computer engineering in the B.Sc. course of the Faculty of Electrical Engineering and Informatics, Department of Control Engineering and Information Technology. His research interests include Artificial Intelligence (AI), data science, software development and telecommunication systems.

Achref Mekni is a student of the Budapest University of Technology and Economics. He studies computer engineering in the B.Sc. course of the Faculty of Electrical Engineering and Informatics, Department of Automation and Applied Informatics. His research interests include telecommunication systems and software development.

Peter A. Kara (M'14) received his M.Sc. degree in Computer Engineering from the Department of Networked Systems and Services (HIT) at the Budapest University of Technology and Economics (BME) in 2013, and he was awarded the Ph.D. title by Kingston University (KU) in 2020. He participated in the EU FP7 ICT CONCERTO and EU H2020 QoE-NET projects, worked as a research associate of the Wireless Multimedia and Networking (WMN) Research Group at KU, and he is currently a research fellow of the Multimedia Networks and Services Laboratory (MEDIANETS) at BME-HIT, an invited fellow of KU and vice-chair of IEEE P3333.1.4.

Maria G. Martini (SRM'07) is (full) Professor in the Faculty of Science, Engineering and Computing in Kingston University, London, where she also leads the Wireless Multimedia Networking Research Group and she is Course Director for the M.Sc. in "Networking and Data Communications". She received the Laurea in electronic engineering (summa cum laude) from the University of Perugia (Italy) in 1998 and the Ph.D. in Electronics and Computer Science from the University of Bologna (Italy) in 2002. She is a Fellow of The Higher Education Academy (HEA). She has led the KU team in a number of national and international research projects, funded by the European Commission (e.g., OPTIMIX, CONCERTO, QoE-NET, Qualinet), UK research councils (EPSRC, British Council, Royal Society), UK Technology Strategy Board / InnovateUK, and international industries.

Laszlo Bokor (M'04) graduated in 2004 with M.Sc. degree in computer engineering from the Budapest University of Technology and Economics (BME) at the Department of Telecommunicatons. In 2006 he got an M.Sc.+ degree in bank informatics from the same university's Faculty of Economic and So- cial Sciences. He received his Ph.D. degree from the BME Doctoral School of Informatics in 2014. Currently, he is with the Department of Networked Systems and Services (HIT) as assistant professor and leads the Commsignia – BME HIT Automotive Communications Research Group. He is a member of HTE, the Hungarian Standards Institution's Technical Committee for Intelligent Transport Systems (MSZT/MB 911), and the Multimedia Networks and Services Laboratory (MEDIANETS). He received the UNKP-16-4-I. Postdoctoral Fellowship from the New National Excellence Program of the Ministry of Human Capacities of Hungary in 2016.

Aniko Simon received her M.Sc. degree in Computer Engineering from the Budapest University of Technology and Economics (BME) in 2015. During her studies at BME, she participated in the research and development efforts of the Mobile Innovation Centre (MIK). She is currently an information engineer at Sigma Technology.

Evaluation of different extractors of features at the level of sentiment analysis

Fatima Es-sabery[1], Khadija Es-sabery[1], Hamid Garmani[1], Junaid Qadir[2], and Abdellatif Hair[1]

Abstract—**Sentiment analysis is the process of recognizing and categorizing the emotions being expressed in a textual source. Tweets are commonly used to generate a large amount of sentiment data after they are analyzed. These feelings data help to learn about people's thoughts on a various range of topics. People are typically attracted for researching positive and negative reviews, which contain dislikes and likes, shared by the consumers concerning the features of a certain service or product. Therefore, the aspects or features of the product/service play an important role in opinion mining. Furthermore to enough work being carried out in text mining, feature extraction in opinion mining is presently becoming a hot research field. In this paper, we focus on the study of feature extractors because of their importance in classification performance. The feature extraction is the most critical aspect of opinion classification since classification efficiency can be degraded if features are not properly chosen. A few scientific researchers have addressed the issue of feature extraction. And we found in the literature that almost every article deals with one or two feature extractors. For that, we decided in this paper to cover all the most popular feature extractors which are BOW, N-grams, TF-IDF, Word2vec, GloVe and FastText. In general, this paper will discuss the existing feature extractors in the opinion mining domain. Also, it will present the advantages and the inconveniences of each extractor. Moreover, a comparative study is performed for determining the most efficient combination CNN/extractor in terms of accuracy, precision, recall, and F1 measure.**

Index Terms—**Opinion mining, Extractors of features, Big-Data, Sentiment analysis, text analysis.**

I. INTRODUCTION

With the emergence of the internet and the social networking revolution, a large number of individuals can express freely their views and feelings about entities, products, people, etc. [1, 2]. This growth is accompanied by a huge volume of opinion data available on the web. Indeed, 2.5 billion bytes of data are created every day. In recent years, 90% of the world's data has been generated.

Opinion analysis, in the computer domain, is concerned with the automatic processing of opinions, feelings, and subjectivity expressed or conveyed in textual and audiovisual statements [3]. Opinions concern entities that can be products, services, themes, public persons, organizations, etc. Textual statements can be presented in different formats/types: article in a newspaper, comment/critique in a website post/comment in social networks (Facebook, Twitter, etc.). Oral statements, presented in audiovisual documents, are also presented in different formats: news, radio programs, YouTube videos, etc. This paper focuses on textual statements on Twitter [4].

Twitter is a microblogging service that allows its users to send and read short messages of up to 140 characters [5]. These messages, called "tweets" can be received and sent from your computer or mobile phone. Twitter has only been in existence for five years but has already become a major actor in the social media industry. It is a way of expression of internauts because it permits to exchange in real-time, on all subjects, points of view or needs. These tweets are well suited to the dissemination and propagation of information because they can be republished and also contain hash-tags, that is, tags assigned by the authors of the tweets to briefly characterize the subject of the tweet [6, 7]. Tweets are provided with meta-data as well as information about their location, language, keyword, sentiments expressed, etc.

Several works have been carried out in order to solve the problem of opinion analysis with different methods (linguistic and/or numerical). These works can therefore be classified according to three approaches. The first is symbolic, using lexicons and linguistic rules [8]. The second is a numerical approach based on machine learning methods. Finally, there is a hybrid approach that is a combination of the two previous ones: it uses both lexicons and machine learning algorithms. All these approaches consist in training a classifier based on descriptors, also called features, specific to the opinion analysis task. These features allow us to infer the polarity of a new tweet. Thus, the good performances of the classifiers are conditioned on the one hand by the quantity of training data and on the other hand by the quality of the features. Indeed, the size of the training corpus must be sufficient for training the classifier, and the features must be specific to the task [9, 10].

In general, the opinion process consists of several phases which are the pre-processing stage, the feature extraction stage, the feature selection stage and the classification stage. Feature extraction is considered the most critical step because the performance of the classification depends on the set of extracted features. The choice of features is very important in the performance of the learning model. Generally, the identification of relevant features is done by feature extraction and selection algorithms. Classifier performance varies from one set of

[1]Department of Computer Science, Faculty of Sciences and Technology, Sultan Moulay Slimane University, Beni Mellal 23000, Morocco
(e-mail: {fatima.essabery, khadija.essabery, garmani.hamid}@gmail.com; abd_hair@yahoo.com
[2]Department of Electronics, Quaid-i-Azam University, Islamabad 45320, Pakistan (e-mail: junaidqadirqau@gmail.com).

DOI: 10.36244/ICJ.2022.2.9

features to another. These features require good conception and real thinking to define or guess the right features for the classification task.

Therefore, feature extraction addresses the issue of identifying the most distinguishing, informational, and minimized set of features to enhance the effectiveness of the data treatment. Relevant feature vectors are still the most popular and suitable way of representing the sample for classification issues. Many scientists from various fields, who are focused on data analysis and classification, are working together to address feature extraction challenges. Today's developments in both sentiment analysis and feature extraction algorithms have allowed us to design the identification tools that can accomplish tasks that were previously impossible to do. Feature extraction is at the core of these advancements with applications in sentiment analysis, as well as numerous other developing applications.

For an efficient classification, it is essential to employ an accurate feature extraction approach to retrieve a set of distinguishing and informational features from the input data. In essence, if the retrieved features do not accurately identify the employed signals and are not meaningful, a classification technique employing such set of features may have issues in finding the feature classes labels. Therefore, the classification accuracy may be reduced. Due to the importance of feature extractors, in this paper, we will detail the principle of the most commonly used extractors. The main points of this paper can be summarized as follows:

- The discussion of the existing feature extractors in the opinion mining domain.
- The description of the advantages and the inconveniences of each extractor
- The used dataset is the Sentiment140 dataset contains approximately 1.6 million tweets that were automatically retrieved with the Twitter API.
- Application of multiple feature extractors and determination of the most effective extractor in the case of Sentiment140.
- Implementation of the convolutional neural network, NB, SVM, ID3 and C4.5 as a classifier.
- Setting up the Hadoop framework for the parallel implementation of our proposal

II. RELATED WORKS

Most conventional research papers on sentiment analysis has employed supervised machine learning approaches as the primary module for classification or clustering [11]. These approaches typically exploit the Bag-Of-Words, Word2vec, GloVe, FastText, N-Gram and TF-IDF models to extract the essential features of the text containing user-generated sentiments [12].

A. Baseline feature extraction methods

The authors of the paper [13] evaluated the performance of the feature extractor N-gram in opinion mining field. They proposed to combine the approach based lexicon with the N-gram method for performing the sentiment classification. And their proposed Senti-N-Gram lexicon based approach outperforms well-known unigram-lexicon based method employing the VADER lexicon and an n-gram opinion mining method SO-CAL.

The paper [14] provides an introduction to BoW, its importance, how it operates, its implementations, and the challenges of utilizing it. This review is helpful in terms of introducing the BoW methodology to new researchers and providing a good context with related work to researchers working on this model.

In [15], the authors have analyzed the effect of TF-IDF feature level on the SS-Tweet dataset for opinion extraction. They found that by employing the TF-IDF feature extractor, the sentiment analysis performance is 3-4% higher than by employing the N-gram feature.

The authors of the paper [16], introduce a Word2vec pattern that provides additional linguistic features to accommodate short Chinese dataset. It is compared with the Internet content-based pattern for long dataset. The empirical findings demonstrate that our pattern can effectively improve the performance of opinion classification using six different classes on Weibo.

In the work [17] a hybrid pattern of embedding glove words, contextual and string similarity measures are applied on the large dataset for key sentence retrieval and classification. The empirical findings demonstrate that the GloVe extraction pattern is better than existing metrics for key sentence and string similarity in large datasets.

In [18], the authors have studied the fastText feature extractor and the experimental results show that the FastText achieves 0.97 area under the ROC curve, 94.2% F-measure, and 74.8 ms inference times for CPU.

B. The state-of-the-art feature extractors

The authors of the paper [19] proposed an attention-bidirectional algorithm based on the both deep neural networks CNN and RNN as feature extractor for opinion mining. Their approaches extracted past and future features by taking in consideration the temporal data stream in both senses. In addition, the attention layer mechanism is implemented on the outputs of the bidirectional LSTM layers to emphasize different extracted features to a greater or lesser extent. They also applied the convolution and pooling layers of the CNN in order to reduce the dimensionality of the extracted features. The results of the comparison of their approach and six more recently suggested DNNs for opinion mining indicate that their approach accomplishes the best performance on both short tweet and long review polarities detection.

In the paper [20] a novel efficient method for sentiment classification employing machine learning techniques is suggested. The process of this novel approach is carried out in three phases. In the first phase, the dataset is gathered and

pretreated, in the second phase the dataset is tuned by extracting the relevant characteristics, and in the third phase the trained dataset is classified into three classes (negative, neutral, and positive) by implementing several machine learning techniques. Every machine learning algorithms yield distinct results. It is observed that the suggested approach i.e., selective algorithm combined with decision tree provides a high accuracy of 89.47% in comparison to other machine learning techniques

The authors of the paper [21] evaluate different combinations of features in Twitter opinion mining. In addition, they assess and study the effect of combining these separate kinds of characteristics to detect of which aggregation yield crucial insights in the polarity classification task in Twitter opinion mining.

In the paper [22], a comparative study of two extractors (TF-IDF, and Doc2vec) is carried out. The authors of this paper implement these two extractors on three datasets such as Stanford movie review, UCI sentiment, and Cornell movie review datasets. Also, they applied several preprocessing tasks such as removing stop words, eliminating the special characters, stemming and tokenization which increases the accuracy of sentiment classification and reduce the execution of time of used classifier. The pertinent features extracted after the extraction step are tested and trained using various machine learning algorithms like support vector machine, Bernoulli naïve bayes, k-nearest neighbors, decision tree, and logistic regression.

The authors of the paper [23], carried out an experimental analyze of different techniques of feature extraction in Twitter sentiments analysis. Their comparative study is performed in four steps, the first one is the data gathering task which has been carried out from readily available sources. The second phase is the application of several preprocessing tasks utilizing the tool POS. In the third step, various feature selector and extractor are implemented over the collected tweets. Finally, the experimental study is performed for detecting the opinion polarity with different extractors.

Zainuddin et al. [24] proposed a hybrid model for classifying the tweets aspect-based opinion mining. They carried out a comparative analyze in terms of classification rate of three features selectors such as latent semantic analysis, principal component analysis, and random projection. In addition the hybrid model was evaluated employing Twitter datasets to represent various areas, and the evaluation with several machine learning algorithms also proved that the novel hybrid model achieved goods results. Their experimental results showed that the proposed hybrid opinion classification model was capable to increase the classification rate from the existing conventional opinion mining approaches by 76.55%, 71.62% and 74.24 %, respectively.

Pandian suggested in its paper [25], a comparative study of sentiment classification by employing various deep learning models. Its proposed paper has incorporated a feature-extraction with a deep learning model. Furthermore, its research work has three major phases: The first phase is the design of opinion classifiers based on deep learning models. This step is succeeded by the utilization of ensemble techniques and merging of information to get the final ensemble of data sources. As the third phase, an aggregation of ensembles the information is proposed to classify several algorithms along with the suggested algorithm.

III. EXTRACTORS OF FEATURES

Concerning machine learning approaches, many efforts have been performed in the literature on Twitter opinion mining to obtain an efficient vectorization of tweets. In this context, various kinds of features extractors have been suggested already, ranging from simple n-gram based vectorization to meta-level features to word embeddings.

A. N-gram extractor

The N-gram feature extractor is commonly being employed in text based-classification [26]. After applied this extractor, the sentence can be broken down into features of character n-grams and word n-grams. So, an N-gram is a series of “characters or words " picked up, in order, from a body of sentence. N-gram may be unigram (n-gram = 1), bigram (n-gram = 2), trigram (n-gram = 3), and so on.

Usually we pick every word in a sentence to compute the sentiment of the sentence, but there can be a case in which the word is formerly employed in a positive sense, but now it is employed in a negative sense; for example, “what an awesome product, totally waste of money,” if we take only the word “awesome” the sentence will be positive but if we take in consideration the whole sentence, it is indicating the negative. It is because of these types of problems that the N-gram is being developed.

B. TF_IDF extractor

TF_IDF means term frequency - inverse document frequency, which is widely recognized and it is utilized as a weighting procedure and its performance is also still very comparable with new approaches [27]. It is a statistical value that is meant to reflect how much weight a given word has to a certain document in a corpus or a collection. The TF-IDF rate boosts proportionally to the number of occurrences of each term in the document, but is compensated by the occurrence of the term in the corpus, which aids to adapt for the fact that a certain terms occur more frequently in overall. The standardization TF-IDF rates for any document in the corpus via the Euclidean measure are used. The calculations of TF-IDF are presented in the following equation:

$$(\text{TF\_IDF})_{ij} = (\text{TF})_{ij} * \log(\text{IDF})_i \quad (1)$$

Where $\text{TF} = \frac{k_{ij}}{Number\ of\ the\ words\ in\ the\ sentence}$ with $\boldsymbol{k}$ is the number of times the word $\boldsymbol{i}$ appears in the sentence $\boldsymbol{j}$.

And $\text{IDF} = \frac{Number\ of\ sentences}{Number\ of\ sentences\ with\ the\ word\ i}$

However, the equation 1 is only applied in cases where (TF) ≥ 1. If it does not, TF_IDF = 0, the equation (2) is used.

$$(\text{TF\_IDF})_{ij} \begin{cases} (\text{TF})_{ij} * \log(\text{IDF})_i & \text{if } (\text{TF})_{ij} \geq 1 \\ \text{Otherwise} \quad (\text{TF\_IDF})_{ij} = 0 \end{cases} \quad (2)$$

Where TF denotes the weight standard. It is the weight, which indicates the frequency or relative frequency of the word ***i,*** in a given sentence ***j.*** And IDF denotes the weight global. It indicates the support of the word ***i*** in respect to ***jth*** belonging to the corpus. In summary:

$(\mathrm{TF})_{ij}$: Number of occurrences of word *i* in sentence *j*.

$(\mathrm{IDF})_{i}$: Number of sentences containing the word *i*.

C. Bag-of-words extractor

The bag-of-words is an approach that has been suggested for the first time in the text retrieval area issue for analysis of documents based-text, and it was later induced for computer vision implementations [28]. In general, this approach associates a text with a vector indicating the number of occurrences of each chosen word in the training corpus, For example, we have the three book reviews as presented below:

- **Review A**: This book is very long and boring
- **Review B**: This book is not boring and is shortened
- **Review C**: This book is good and enjoyable

The vocabulary of this three movie reviews consists of eleven words which are: 'This', 'book', 'is', 'very', 'boring', 'and', 'long', 'not', 'shortened', 'good', 'enjoyable'. Therefore the numerical vector of each review is created by the bag-of-word method as follows:

- **Vector of Review A**: [This:1, book:1, is:1, very:1, boring:1, and:1, long:1, not:0, shortened:0, good:0, enjoyable:0]
- **Vector of Review B**: [This:1, book:1, is:1, very:0, boring:1, and:1, long:0, not:1, shortened:1, good:0, enjoyable:0]
- **Vector of Review C:** [This:1, book:1, is:1, very:0, boring:0, and:1, long:0, not:0, shortened:0, good:1, enjoyable:1]

D. Word2Vec extractor

Word integration with word2vec [29] identifies the syntactic characteristics of terms and attributes a sentiment score to every term in the vector space. Terms that appear in the identical context are deemed more similar than the terms that appear in the dissimilar contexts. For example, we have a corpus **C** which is composed of a set of tweets, $\boldsymbol{C} = \{\boldsymbol{t_1}, \boldsymbol{t_2}, \boldsymbol{t_3}, \ldots, \boldsymbol{t_n}\}$ and a vocabulary $\boldsymbol{V} = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{w_3}, \ldots, \boldsymbol{w_m}\}$ is composed of a set of unique words retrieved from **C**. Therefore, the vectorization of the words wi are identified by applying one of the both models Skip-gram or Continuous bag-of-words of Word2Vec in order to compute the probability distribution of the rest words of the set $\boldsymbol{V}\backslash\{\boldsymbol{w_i}\}$ in the context provided by the words wi. In addition, $\boldsymbol{w_i}$ is expressed as a vector space $\boldsymbol{s_i}$ which includes the probabilistic rates of all the other words in the lexicon. The Word2Vec approach extract semantic linked among words in the vocabulary. Furthermore, the obtained set of vectors spaces for all words in the lexicon is high-dimensional and is efficacy for sentiment classification.

E. GloVe extractor

The GloVe pattern [30] attempts to create a vector space representation of a term by employing the similarities between the terms as an invariant. The GloVe combines techniques provided by two different patterns, which are the Continuous Bag of Words and Skip-gram pattern. Problem with the former pattern is the low classification rate but its computational time is very efficient, while latter had computational time is inefficient but its classification rate is very high. What the GloVe attempts to do is to integrate the techniques introduced by two patterns and it has demonstrated to be more efficient and accurate than those two patterns.

F. FastText extractor

In recent years, Facebook researchers have launched a new word embedding system called FastText [31], which is a quick and effective way to represent each term with vector space and to classify text-based sentiments. The primary goal of fastText term embeddings is to consider the inner structure of terms rather than to learn term representations. FastText operates by Dragging a window over the entry text and either learning the central term from the remainder of the context (by employing the BOW approach), or all the terms in the remainder of the context from the central term by using the Skip-gram approach. The FastText approach is very identical to Word2Vec approach, the only difference is that the FastText learn the vector representation of sub-parts of a term so-called character n-grams.

IV. ADVANTAGES AND DISADVANTAGES OF EACH EXTRACTOR

In order to implement machine learning approaches to natural language issues, it is necessary to convert the text-based data into digital data. The methods used to carry out this conversion are the extractors described above. Each extractor has the advantages and disadvantages as presented in the tables below:

TABLE I
ADVANTAGES AND DISADVANTAGES OF EACH EXTRACTOR

Extractor	Advantages	Inconveniences
N-gram	-It pick up the representation of the out-of-lexicon terms because it divide the word into N-gram characters [27]. -Simplicity and scalability which means that this approach can efficiently scale small experiments. It can also stock more background with a good understanding of the space-time compromise [28]. -It is efficient in handling textual mistakes and character identification issues that is because of the N-gram structure [27].	- When its parameter N is very large, its parameters space is much too large [28]. - There is also a text smoothness issue because of text sparsity. That's means that we used an approximating function that tries to detect significant features in the data [29]. -It does not take into account the semantics.
Bag-of-words	- It encrypts each term in the lexicon as one-hot vector that renders our training data more meaningful and more expressible, and can be easily rescaled [30]. - It is very simple to comprehend and to implement because is based on one-hot vector representation. - It generates a simplified word representation because it is easier to calculate a likelihood for values by utilizing numerical values. [30].	-It does not take into account the semantics of the term because it does not compute the semantic similarity of each terms [31]. -It does not take into account the semantic connection between the terms because it does not compute the semantic similarity between the terms [31]. -It is suffering from the curse of dimensionality because it represents each term by one-hot vector [30].
TF-IDF	-Short extraction time, simple and easy to calculate because it merges only two notions, term frequency and document frequency [32]. -It provides a certain basic metric to retrieve the most descriptive words in the corpus because it calculates easily the similarities between two sentences or two documents in the corpus [32].	- It does not detect semantics, status in text, co-occurrences in diverse sentences in the corpus because it cannot contribute to convey a semantical sense. [33]. - It is only used as a lexical level characteristic because it gives importance to the terms by the way it weights them and it cannot adequately infer the meaning of the terms and understand their significance in this way [34]. -It cannot pick up the semantics with respect to thematic patterns, term embeddings [34]. - A further drawback is that it may suffer from a lack of memory as TF-IDF may suffer from the curse of dimensionality [33].
Word2Vec	-It identifies the syntactic characteristics of terms because it utilizes a neural network pattern so that once a pattern is trained it can recognize antonymic and synonymous words or can propose a new word to complete an incomplete partial phrase [35].	-It does not learn vectors space of the character n-grams because Word2Vec employs the same vector of numbers to represent any unseen word [35].

TABLE II
ADVANTAGES AND DISADVANTAGES OF EACH EXTRACTOR

Extractor	Advantages	Inconveniences
Word2Vec	-It attributes a sentiment score to every term. For instance, certain negative terms that are adjectives will be more closely related to each other and inversely for positive adjectives. It picks up the semantic and syntactic data of the words [36]. - Its embedding vector size is very small which avoids both drawback of the lack of memory and the curse of dimensionality [36]. -Its context data is never lost because it employs the continuous bag of words method and skip-gram method for predicting the word or the context any word [35].	- It is not very efficacy with term analogy tasks compared to word similarity task [36]. - Word2Vec cannot deal well with out-of-vocabulary terms. It attributes a random vectorial mapping for out-of-vocabulary words, which may be suboptimal [35]. And it is incapable of taking advantage of the statistics of the corpus. - Long extraction time because Word2vec train either continuous bag of words method or skip-gram method and these both methods train the neural network model which trains huge number of instructions and that takes long execution time [36].
GloVe	-It forces term vectors to pick up sub-linear relations in the vector space since vector spaces being by nature linear structures, the easiest way to proceed is to use vector differences [37]. -It outperforms Word2vec in the tasks of terms analogies because it is based on leveraging global word to word co-occurrence counts leveraging the entire corpus [38]. -It adds a more convenient meaning to term vectors by considering the relations between terms pair to pair rather than term to term [37]. -It assigns a smaller weight to very frequent term pairs in order to avoid meaningless terms such as “the”, “a” [38].	-Its pattern is learned on the terms co-occurrence matrix, which requires a lot of storage space because the co-occurrence matrix expands so rapidly and is high-dimensional [38]. -It consumes very time, because the change in level of hyper-parameters requires the reconstruct of the co-occurrence matrix [37]. -It cannot pick up the representation of the out-of-lexicon terms because Glove processes each term in the corpus as an atomic entity and produces a vector for each term [38]. -It is difficult to detach several opposite term pairs using GloVe unlike the Word2Vec [36, 37].
FastText	-It learns usually often the numeric vector of terms in the sentiment analyses process because it is based on the combination of the concept of Word2Vec approach and N-gram method [39]. -It requires a few preprocessing tasks, and little hyper parameter tuning thus needs small memory spaces because it is based on character N-gram [40]. -It learn the vector spaces of character n-grams that make it very efficacy to deal with out-of-vocabulary terms [40].	- Sublinear connections are not explicitly identified because FastText cannot compute the semantic similarity of each term [40]. - As the size of the corpora increases, the memory space used by the FastText word embeddings needs to be increase which takes long execution time for extracting the pertinences features [39]. - It could be very hard to be trained if the Softmax function is used, since the size of the vocabulary is much too big [40].

V. SENTIMENT ANALYSIS METHODOLOGY

In the current work, we train a convolutional neural network as classifier on Sentiment140 dataset [26] of sentiment sentences in order to evaluate each extractor (N-gram, Bag-of-word, TF-IDF, Word2Vec, GloVe, and FastText) for identifying the most efficient one. In general our sentiment analysis methodology consists of four steps which are *data collection phase* in which we used the Sentiment140 dataset, *data pre-processing phase* in which we applied several techniques for improving the data quality and eliminate the data noisy, *feature extraction phase* in which we implemented six extractors in order to determine the most efficient one among them, et finally the *data classification phase*, in which we applied the convolutional neural network (CNN) as classifier as shown in the fig.1.

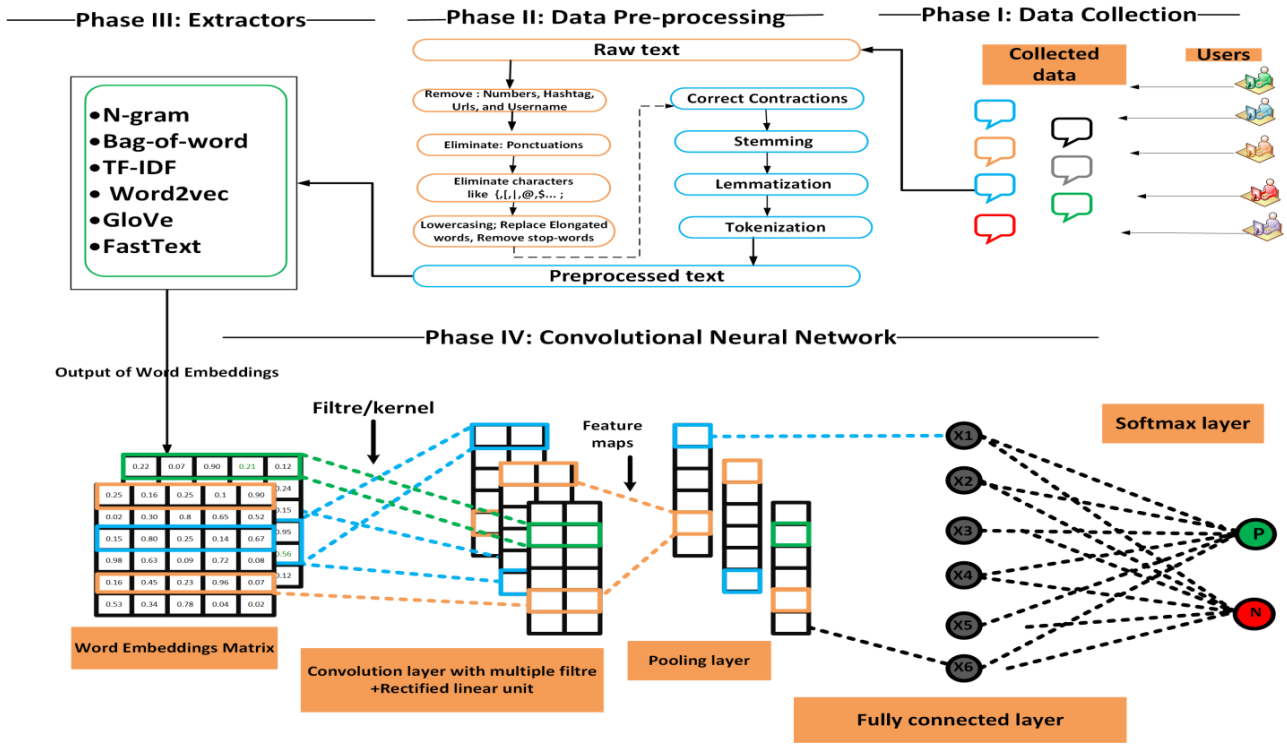


Fig. 1. Architecture global of the sentiment analysis methodology

A. Data collection phase

To implement our contribution we have used the Sentiment140 dataset. This dataset contains approximately 1.6 million tweets that were automatically retrieved with the Twitter API. These tweets were automatically annotated assuming that those containing the ":)" emoticon were positive and those containing the ":(" emoticon were negative. Those containing neither of these emoticons, and those containing both, were not kept. The training set is annotated in two classes (positive and negative) while the test set is annotated by hand on three different classes (positive, negative and neutral). For our experiments, we use only the positive and negative classes of the test set. Table 3 gives the details of the data set.

TABLE 3
DETAILS OF THE USED DATA SET

Training set		Testing set	
Positive	720,000	Positive	80,000
Negative	720,000	Negative	80,000

Each line of the file contains a single tweet with a maximum of 140 characters and can contain several sentences (depending on the length). Because the tweets have been collected directly on the twitter API, they can therefore contain HTML addresses, # hashtags and user names (preceded by @). Finally the structure of each line is as follows:

1. The polarity of the tweet (e.g., 0 = negative, 2 = neutral, 4 = positive).
2. The id of the tweet (e.g., 6532).
3. The date of the tweet (e.g., FAR Sep 18 15:45:31 UTC 2021).
4. The name of the user who posted the tweet (e.g., Essabery).
5. The text of the tweet.

B. Data pre-processing phase

After looking at the data, we saw that the sentences contained HTML tags, empty words and all punctuation. So we started by removing the noise to normalize our sentences. We remove HTML tags with the BeautifulSoup2 module. We also remove

all the characters that are not letters and therefore, remove all punctuation from texts. Because stop words, by definition, do not bring any information to the text, we eliminate them too. All letters are also changed to lower case. Finally, we root all the words to process each inflection of a word into a single word. Below we detail some important pre-processing steps.

Nicknames: Since nicknames (e.g., @username) are useless for sentiment analysis, we replace all @usernames with the text AT_USER so that we can delete them later.

Repeated letters: The language used on Twitter is mostly familiar. It is therefore not uncommon for words to be written with a letter (or several) that is repeated when it should not be. For example the word "dog" can be found as "dooooooog" on Twitter. As soon as a word contains identical letters that are repeated more than more than twice, they are replaced by only two occurrences of the same letter ("dooooooog" becomes “doog").

Hashtags: Twitter hashtags are used to create an instant connection with other users. The word that follows the # is usually a word that provides a lot of information about the sentiment of the sentence. We keep this word, but the hashtag character is removed.

Lemmatization: We transform all inflections into their root. The objective is to reduce the derived forms of a word to a common base form in order to facilitate the correspondence between the different terms.

1) EFFECT OF PRETREATMENT

Table 4 shows the effect of these preprocessing on the number of useful words in the text.

TABLE 4
EFFECT OF NOISE REDUCTION

Reduction	Number of features	% of the original
None	1 569 914	100%
Username	65 993	96.88%
URLs	609 692	54.22%
Repeated letters	298 673	78.35%
All	984 139	39.43%

All these text noise removals lead to a reduction of the corpus set to 39.43% of the original corpus size.

TABLE 5
ACCURACY AND ERROR RATE WITHOUT AND WITH PREPROCESSING

Criteria	Without preprocessing	With preprocessing
Accuracy	50.19%	82.35%
Error rate	49.81%	17.65%

As shown in the table below (5), the preprocessing tasks reduce the error rate from 49.81% to 17.65 and increase the accuracy from 50.19% to 82.35%. So, it is necessary to apply the preprocessing process before the application of machine learning algorithm.

C. Feature extraction phase

In order to obtain a reliable system based on a numerical approach, the design of good features is the most important step for classification. Bags of words, n-grams, TF-IDF, Word2vec, GloVe and FastText are the most common extractor of features in sentiment analysis. The main purpose of this work is to test different sets of extractors and pre-trained word embeddings by applying the CNN classifier.

D. Data classification phase

After the feature extraction step, the next step is the data classification in which we have used the CNN. The CNN is a specialized type of multi-layer neural network generally used when the input is structured according to a grid (e.g. an image). These networks were inspired by the visual cortex of animals, and more particularly on its properties: local receptive fields and weight sharing. Figure 2 shows the different layers of a convolutional neural network. The latter is composed of one or more convolution and pooling blocks, one or more hidden layers and an output layer. The CNN takes as input a multi-dimensional grid representing a learning or inference instance, and provides as output the corresponding class.

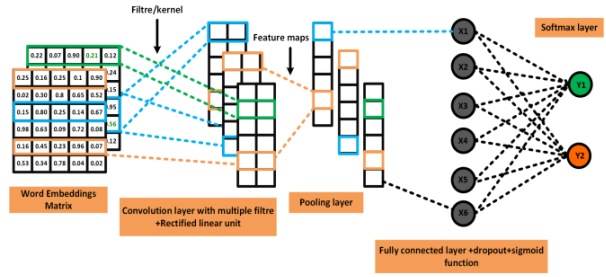


Fig. 2. Simple version of the convolutional neural network

VI. EXPERIMENTAL RESULTS

As mentioned earlier, this approach consists of evaluating the set of extractors described below using the CNN as a classifier. In this section we will examine the performance of each extractors by applying them on the corpus Sentiment10 and by computing four evaluations criteria [41] which are the following:

Precision (P): represents the average of the precisions of the k classes. It is calculated according to equation (5).

$$P = \frac{\sum_{j=1}^{k} P_j}{k} \tag{5}$$

With:

$$P_j = \frac{number\ of\ sentences\ correctly\ assigned\ to\ the\ class\ y_i}{number\ of\ sentences\ assigned\ to\ the\ class\ y_i} \tag{6}$$

Recall (R): represents the average of the recalls of the k classes. It is calculated according to the following equation

$$R = \frac{\sum_{j=1}^{k} R_j}{k} \tag{7}$$

With:

$$R_j = \frac{number\ of\ sentences\ correctly\ assigned\ to\ the\ class\ y_i}{number\ of\ sentences\ belong\ to\ the\ class\ y_i} \quad (8)$$

F1 measure (F1): represents the harmonic mean of precision and recall. It measures the performance of the system and is calculated according to equation (9).

$$F1 = \frac{2 \times P \times R}{P + R} \quad (9)$$

Accuracy A: evaluates our approach in an overall [23]. It is calculated according to the equation (10).

$$A = \frac{Number\ of\ sentences\ correctly\ classified}{Total\ number\ of\ sentences} \quad (10)$$

We have also implemented our approach on Hadoop framework with a cluster of five machine: four slave nodes and one master node. The configuration of the cluster is presented in the following table 6:

TABLE 6
PARAMETERS SETTING OF THE HADOOP CLUSTER

Configuration	Parameters
N. of nodes	5
apache hadoop	version 2.7.2
OS	Ubuntu 20.04.4
Memory	16 GB
CPU	Intel(R) Core(TM) i7-4810MQ CPU @ 2.80GHz 2.80 GHz

The Parameters settings of CNN used in our implementation are shown in the following table 7:

TABLE 7
PARAMETERS SETTING OF THE CNN

Parameter	Value
Vocabulary size	45,000
Padding	0
Regularizer	L2
Size of filter	4,7
Number of filter	15
Activation function	ReLU
Function of Pooling layer	Max-pooling
N. of pooling layer	3
N. of convolutional layer	3
Input embedding matrix	500x500

A. Impact of the choice of input embeddings on the CNN

The CNN architecture was trained on the Sentiment140 corpus with the different existing pre-trained word embeddings (Bag-of-Word, N-grams, TF-IDF, Word2vec, GloVe, and FastText). The table 3 reports the performance of our approach in terms of precision, recall, F1 measure and accuracy.

TABLE 8
P, R F1, A OF THE ALL COMBINATIONS OF OUR APPROACH

Combination\criteria (%)	P	R	F1	A
CNN+BOW	64.08	63.51	63.79	66.92
CNN+N-grams	49.97	60.23	54.62	58.49
CNN+TF-IDF	68.88	71.59	70.20	70.02
CNN+Word2vec	86.13	83.55	84.82	85.06
CNN+GloVe	79.49	80.05	79.76	79.54
CNN+FastText	**93.43**	**90.89**	**92.14**	**91.32**

From the table 8, we remark that the combination **CNN+FastText** performs better than other combinations in terms of P(93.43%), R(90.89%), F1(92.14%) and A(91.32%).

Pattern complexity is a metric for the time and space consumption used by a pattern. In these following tables, we assessed the time and space complexity of all combinations of our approach.

TABLE 9
SPACE COMPLEXITY OF THE ALL COMBINATIONS OF OUR APPROACH

Combination\ complexity	N. operations	N. parameters
CNN+BOW	65.5M	40M
CNN+N-grams	77.25M	36M
CNN+TF-IDF	89M	46M
CNN+Word2vec	53.5M	29M
CNN+GloVe	45M	27M
CNN+FastText	**39M**	**22M**

As the experimental findings indicate in the table 9 above, the combination **CNN+FastText** requires computational complexity in space much lower than others combinations. Since it carried out numerous operations with a size equal to 39M and its size of parameters is equal to 22M.

The following table 10 illustrates the experimental findings after gauging the computational time complexity of all combinations of our approach in terms of both training time consumed and testing time consumed.

TABLE 10
TIME COMPLEXITY OF THE ALL COMBINATIONS OF OUR APPROACH

Combination\ complexity	Training time	Testing time
CNN+BOW	46.23s	17.5s
CNN+N-grams	51s	19s
CNN+TF-IDF	69s	21s
CNN+Word2vec	38.25s	14.5s
CNN+GloVe	28.65s	15.25s
CNN+FastText	**15.46s**	**10.98s**

As the experimental findings indicate in the table 10 above, the combination **CNN+FastText** requires computational complexity in time much lower than others combinations. Since it consumed a training time equal to 15.46s and a testing time equal to 10.98s.

B. A comparison of our approach with other machine learning algorithms.

This experiment makes a comparison of the combinations of four machines-learning algorithms which are Naive Bayes (NB), Support Vector Machine **(**SVM), ID3 and C4.5 decision tree algorithm with six feature extractors which are BOW, N-grams, TF-IDF, Word2vec, GloVe and FastText in terms of precision (P), Recall (R), and F1 measure (F1) and Accuracy (A). Its empirical findings are displayed in the Table 11.

TABLE 11
P, R, F1 AND A OF THE COMBINATION OF FOUR MACHINES- LEARNING ALGORITHMS AND SIX FEATURE EXTRACTORS

Combination\criteria	P	R	F1	A
NB+BOW	48.31	45.64	46.93	47.12
NB+N-grams	35.84	34.15	34.97	36.02
NB+TF-IDF	50.97	51.08	51.02	52.68
NB+Word2vec	49.64	50.09	49.86	50.32
NB+GloVe	53.26	55.42	54.31	55.29
NB+FastText	**56.74**	**57.49**	**57.11**	**58.18**
SVM+BOW	45.61	44.88	45.24	46.07
SVM+N-grams	40.33	39.50	39.91	40.26
SVM+TF-IDF	51.64	50.37	50.99	51.63
SVM+Word2vec	45.87	46.25	46.05	45.91
SVM+GloVe	50.89	49.68	50.27	49.82
SVM+FastText	**60.43**	**61.27**	**60.48**	**61.39**
ID3+BOW	62.54	63.19	62.86	63.42
ID3+N-grams	53.48	54.02	53.74	54.13
ID3+TF-IDF	64.85	63.94	64.39	65.07
ID3+Word2vec	58.46	60.32	59.37	61.53
ID3+GloVe	64.15	63.24	63.69	64.18
ID3+FastText	**70.65**	**72.39**	**71.50**	**72.87**
C4.5+BOW	60.58	59.67	60.12	59.93
C4.5+N-grams	58.34	60.59	59.44	61.48
C4.5+TF-IDF	70.49	71.64	71.06	70.97
C4.5+Word2vec	68.31	69.25	68.77	69.51
C4.5+GloVe	71.58	73.82	72.68	73.96
C4.5+FastText	**77.65**	**76.92**	**77.28**	**76.64**

From the table 11, we remark that the combination **machine-learning algorithm+FastText** performs better than other combinations in terms of P, R, F1 and A. Therefore, we notice that the feature extractor FastText outperforms all others feature extractors (BOW, N-grams, TF-IDF, Word2vec, and GloVe). And from the tables 10 and 11, we remark that **CNN+FastText** performs better than others machine learning algorithms (NB,SVM,ID3 and C4.5) in terms of P(93.43%), R(90.89%), F1(92.14%) and A(91.32%).

In Table 11, we see that some values are lower than 0.5. In the case of the NB classifier. This is because the input values to the NB classifier are numerical values in this contribution. And, as we know from the machine learning literature, NB performs well for categorical versus numerical input variables.

C. A comparison of our approach with other approaches selected from the existing literature.

For further testing of our proposed approach, we carried out another experiment aimed at comparing our method with the other approaches taken from the literature, namely Naresh et al. [14], Carvalho et al. [15], Avinash et al. [16], Kumar et al. [17] and Zainuddin et al. [18]. However, in this experiment, the evaluation measures used will be precision (P), Recall (R), and F1 measure (F1) and Accuracy (A). Its empirical findings are displayed in the Table 12.

TABLE 12
P, R, F1 AND A OF OUR APPROACH WITH OTHER APPROACHES SELECTED FROM THE EXISTING LITERATURE

Approach	P	R	F1	A
Naresh et al. [14]	65.48	67.12	66.28	66.87
Carvalho et al. [15]	79.56	80.04	79.79	78.95
Avinash et al. [16]	70.19	69.38	69.78	68.42
Kumar et al. [17]	83.21	82.40	82.80	81.94
Zainuddin et al. [18]	69.34	70.67	69.99	71.68
CNN+FastText	93.43	90.89	92.14	91.32

From the results shown in the table 12, we remark that our approach (CNN+FastText) obtained the strongest performances in terms of accuracy (91.32%), precision (93.43%), recall (90.89%), and F1 measures (92.14%) compared to other chosen classifiers from the literature which are Naresh et al. [14], Carvalho et al. [15], Avinash et al. [16], Kumar et al. [17] and Zainuddin et al. [18].

VII. CONCLUSION

Feature extraction is needed to get good performance in sentiment classification. The purpose of feature extraction is to identify the strongest and most informational set of features to enhance the effectiveness of the classifier. Moreover, Feature extraction is the most critical aspect of opinion classification since classification efficiency can be negatively affected if features are not properly chosen. For that, in this paper, we presented a preliminary study of the most popular feature extractors. And, we combined a CNN, NB, SVM, ID3, and C4.5 with several word embedding methods in order to identify the most efficient extractor of features that positively affected the classifier performances. Accordingly to the experimental results, the performance of the used classifiers varies a little with the nature of the word embedding sets. In general we found the combination **CNN+FastText** outperforms all other combinations in terms of accuracy, precision, recall, and F1 measure.

REFERENCES

[1] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," Proceedings of the 2019 International Conference on Pervasive Computing Advances and Applications, Jaipur, India, (2019) January 8-10.

[2] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies, Tiruchengode, India, (2013) July 4-6.

[3] H. Saif, Y. He, and H. Alani, "Semantic Sentiment Analysis of Twitter," Proceedings of the International Conference on Semantic Web, Boston, MA, USA, (2012) November 11-15.

[4] N. Al-Twairesh and H. Al-Negheimish, "Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets," Journal of IEEE Access, vol. 7, (2019), pp. 84122–84131.

[5] M. Venugopalan and D. Gupta, "Exploring sentiment analysis on twitter data," Proceedings of the 2015 Eighth International Conference on Contemporary Computing, Noida, India, (2015) August 20-22.

[6] H. Kaur, V. Mangat, and Nidhi, "A survey of sentiment analysis techniques," Proceedings of the 2017 International Conference on IoT in Social, Mobile, Analytics and Cloud, Palladam, India, (2017) February 10-11.

[7] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," Proceedings of the 8th International Conference on Advances in Information Technology, Macau, China, (2016) December 19-22.

[8] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python," International Journal of Computer Application, vol. 165, no. 9, pp. 29–34.

[9] H. Hamdan, P. Bellot, and F. Bechet, "Lsislif: Feature Extraction and Label Weighting for Sentiment Analysis in Twitter," Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, Colorado, (2015) June 4-5.

[10] A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," Proceedings of the 2nd International Conference on Applied and Theoretical Computing and Communication Technology, Bangalore, India, (2016) July 21-23.

[11] F. Es-sabery, K. Es-sabery, and A. Hair, "A MapReduce Improved ID3 Decision Tree for Classifying Twitter Data," Edited M. Fakir, M. Baslam, and R. El Ayachi, Springer, Cham, (2021), pp. 160–182.

[12] F. Es-Sabery et al., "A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier," International Journal of the IEEE Access, vol. 9, (2021), pp. 58706–58739.

[13] A. Dey, M. Jenamani, and J. J. Thakkar, "Senti-N-Gram: An n-gram lexicon for sentiment analysis," Expert Systems with Applications, vol. 103, pp. 92–105, Aug. 2018, DOI: 10.1016/j.eswa.2018.03.004.

[14] W. A. Qader, M. M. Ameen, and B. I. Ahmed, "An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges," in 2019 International Engineering Conference (IEC), Jun. 2019, pp. 200–204. DOI: 10.1109/IEC47844.2019.8950616.

[15] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," Procedia Computer Science, vol. 152, pp. 341–348, Jan. 2019, DOI: 10.1016/j.procs.2019.05.008.

[16] B. Shi, J. Zhao, and K. Xu, "A Word2vec Model for Sentiment Analysis of Weibo," in 2019 16th International Conference on Service Systems and Service Management (ICSSSM), Jul. 2019, pp. 1–6. DOI: 10.1109/ICSSSM.2019.8887652.

[17] S. Anjali Devi and S. Sivakumar, "An efficient contextual glove feature extraction model on large textual databases," Int J Speech Technol, Sep. 2021, https://doi.org/10.1007/s10772-021-09884-2.

[18] J. Kralicek and J. Matas, "Fast Text vs. Non-text Classification of Images," in Document Analysis and Recognition – ICDAR 2021, Cham, 2021, pp. 18–32. DOI: 10.1007/978-3-030-86337-1_2.

[19] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis," International Journal of the Future Generation Computer Systems, vol. 115, (2021), pp. 279–294.

[20] A. Naresh and P. Venkata Krishna, "An efficient approach for sentiment analysis using machine learning algorithm," International Journal of the Evolutionary Intelligence, vol. 14, no. 2, (2021), pp. 725–731.

[21] J. Carvalho and A. Plastino, "On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis," International Journal of the Artificial Intelligence Review, vol. 54, no. 3, (2021), pp. 1887–1936.

[22] M. Avinash and E. Sivasankar, "A Study of Feature Extraction Techniques for Sentiment Analysis," Edited M. Ajith, M. Paramartha, M. Jyotsna, M. Abhishek, M. Soumi, Springer, Cham, (2019), pp. 475–486.

[23] J. A. Kumar and S. Abirami, "An Experimental Study of Feature Extraction Techniques in Opinion Mining," International Journal on Soft Computing, Artificial Intelligence and Applications, vol. 4, no. 1, (2015), pp. 15–21.

[24] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," International Journal of the Applied Intelligence, vol. 48, no. 5, (2018), pp. 1218–1232.

[25] A. P. Pandian, "Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis," International Journal of Soft Computing Paradigm, vol. 3, no. 2, (2021), pp. 123–134.

[26] F. Es-Sabery et al., "Sentence-Level Classification Using Parallel Fuzzy Deep Learning Classifier," International Journal of IEEE Access, vol. 9, (2021), pp. 58706–58739.

[27] F. Es-sabery, K. Es-sabery, H. Garmani, and A. Hair, "Sentiment Analysis of Covid19 Tweets Using A MapReduce Fuzzified Hybrid Classifier Based On C4.5 Decision Tree and Convolutional Neural Network," E3S Web Conf., vol. 297, p. 01052, 2021, DOI: 10.1051/e3sconf/202129701052.

[28] F. Es-sabery and A. Hair, "A MapReduce C4.5 Decision Tree Algorithm Based on Fuzzy Rule-Based System," Fuzzy Information and Engineering, vol. 11, no. 4, (2019), pp. 446–473, DOI: 10.1080/16168658.2020.1756099.

[29] F. Es-sabery and A. Hair, An Improved ID3 Classification Algorithm Based On Correlation Function and Weighted Attribute *. 2019, p. 8. DOI: 10.1109/ISACS48493.2019.9068891.

[30] H. Choi, K. Cho, and Y. Bengio, "Context-dependent word representation for neural machine translation," Computer Speech & Language, vol. 45, pp. 149–160, Sep. 2017, DOI: 10.1016/j.csl.2017.01.007.

[31] L. Wu, S. C. H. Hoi, and N. Yu, "Semantics-Preserving Bag-of-Words Models and Applications," IEEE Transactions on Image Processing, vol. 19, no. 7, pp. 1908–1920, Jul. 2010, DOI: 10.1109/TIP.2010.2045169.

[32] S.-W. Kim and J.-M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," Hum. Cent. Comput. Inf. Sci., vol. 9, no. 1, p. 30, Aug. 2019, DOI: 10.1186/s13673-019-0192-7.

[33] M. Karthiga, S. Sountharrajan, A. Bazila Banu, S. Sankarananth, E. Suganya, and B. Sathish Kumar, "Similarity Analytics for Semantic Text Using Natural Language Processing," in 3rd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing, Cham, 2022, pp. 239–248. DOI: 10.1007/978-3-030-78750-9_17.

[34] V. Kalra, I. Kashyap, and H. Kaur, "Improving document classification using domain-specific vocabulary: hybridization of deep learning approach with TFIDF," Int. j. inf. tecnol., Mar. 2022, DOI: 10.1007/s41870-022-00889-x.

[35] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews," Procedia Computer Science, vol. 179, pp. 728–735, Jan. 2021, DOI: 10.1016/j.procs.2021.01.061.

[36] B. Li, A. Drozd, Y. Guo, T. Liu, S. Matsuoka, and X. Du, "Scaling Word2Vec on Big Corpus," Data Sci. Eng., vol. 4, no. 2, pp. 157–175, Jun. 2019, DOI: 10.1007/s41019-019-0096-6.

[37] J. Bernabé-Moreno, A. Tejeda-Lorente, J. Herce-Zelaya, C. Porcel, and E. Herrera-Viedma, "A context-aware embeddings supported method to extract a fuzzy sentiment polarity dictionary," Knowledge-Based Systems, vol. 190, p. 105236, Feb. 2020, DOI: 10.1016/j.knosys.2019.105236.

[38] M. I. Prabha and G. Umarani Srikanth, "Survey of Sentiment Analysis Using Deep Learning Techniques," in 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Apr. 2019, pp. 1–9. DOI: 10.1109/ICIICT1.2019.8741438.

[39] P. Mojumder, M. Hasan, Md. F. Hossain, and K. M. A. Hasan, "A Study of fastText Word Embedding Effects in Document Classification in Bangla Language," in Cyber Security and Computer Science, Cham, 2020, pp. 441–453. DOI: 0.1007/978-3-030-52856-0_35.

[40] A. G. D'Sa, I. Illina, and D. Fohr, "BERT and fastText Embeddings for Automatic Detection of Toxic Speech," in 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA), Feb. 2020, pp. 1–5. DOI: 10.1109/OCTA49274.2020.9151853.

[41] F. Es-sabery, K. Es-sabery, B. El Akraoui, and A. Hair, "Optimization Focused on Parallel Fuzzy Deep Belief Neural Network for Opinion Mining," in Business Intelligence, Cham, 2022, pp. 3–28. DOI: 10.1007/978-3-031-06458-6_1.

Fatima Es-Sabery received PhD degree in BigData Mining from the Department of Computer Sciences, Sultan Moulay Sliman University, Beni Mellal, Morocco, in 2022. Her general research interests include data mining area, big data field, wireless sensor networks, fuzzy systems, machine learning, deep learning, and the Internet of Things.

Khadija Es-Sabery received the Engineering degree from the Department of Computer Science, National School of Applied Sciences, Cadi Ayyad University, Marrakech, Morocco, in 2021. Her general interests include data mining area, big data field, wireless sensor networks, fuzzy systems, machine learning, deep learning, and the Internet Things.

Hamid Garmani received the Ph.D. degrees from University Sultan Moulay Slimane, Morocco, in 2020. His research interests include network economics, network security, applications of game theory in wireless networks, and radio resource management.

Junaid Qadir is currently pursuing the Ph.D. degree with the Department of Electrical, Electronic and Telecommunications Engineering, and Naval Architecture (DITEN), University of Genova, Italy. He is also a Research Collaborator with the Department of Signal Theory, Communications and Telematics Engineering, University of Valladolid, Spain. He has published many research articles in highly reputed international journals and conferences.

Abdellatif Hair currently works as a Full Professor with the Department of Computer, FST Beni Mellal, Morocco, and a member of the LAMSC Laboratory. His research interests include object-oriented analysis/ design, security of mobile agents, wireless sensor network (WSN), data warehousing, and machine learning (ML).

2023 IEEE Wireless Communications and Networking Conference

Wireless Communications for Social Innovation

26 – 29 March 2023 // Glasgow, Scotland

ORGANISING COMMITTEE

General Co-chairs
James Irvine, University of Strathclyde
Julie Snell, Scotland 5G Centre

TPC Co-Chairs
Muhammad Imran, University of Glasgow
Muhammad Zeeshan Shakir, University of West of Scotland

Workshop Co-chairs
Klaus Moessner, Technische Universität Chemnitz
Christos Tachtatzis, University of Strathclyde

Tutorial Chair
George Goussetis, Heriot-Watt University

Industry Co-Chairs
Majid Butt, Nokia
Doug Carson, Illuminate Technologies

Panel Chair
Lajos Hanzo, University of Southampton

Publicity Chair
Siân Williams, CENSIS

Outreach Chair
Ciara Mitchell, Scotland IS

Publications Chair
Robert Atkinson, University of Strathclyde

Local Arrangements Chair
Qammer Abbasi, University of Glasgow

Standards Chair
David Law, HPE

Diversity & Inclusion Chair
Carol Marsh, Leonardo

Innovation Chair
Greig Paul, University of Strathclyde

Questions Regarding Technical Paper Submissions?
Technical-co-chair-WCNC-2023@ComSoc.org

Call for Papers

The IEEE Wireless Communications and Networking Conference (WCNC) is one of the premier annual events of IEEE in the wireless research arena bringing together researchers, academics, industry, and government. WCNC 2023 will be held in Glasgow, home of four universities as well as the Scotland 5G Centre, stimulating the deployment of new wireless technologies across urban and rural areas and transforming society, economy and industry. WCNC 2023 will include technical sessions, tutorials, workshops, and technology and business panels. You are invited to submit papers, and proposals for panels, tutorials, and workshops, in all areas of wireless communications, networks, services, and applications. Information on how to submit proposals for panels, tutorials, and workshops can be found on the WCNC 2023 conference website. The submissions of technical papers should be made on EDAS in the following four tracks.

IMPORTANT DATES

Paper Submissions Deadline:	**12 September 2022**
Notification of Acceptance:	1 December 2022
Camera-Ready Paper:	15 January 2023

For more information visit:
wcnc2023.ieee-wcnc.org

IEEE/IFIP Network Operations and Management Symposium

8-12 May 2023 // Miami, FL // USA

Integrated Management Towards Resilient Networks and Services

• • • Call for Papers • • •

The 19th IEEE/IFIP Network Operations and Management Symposium (NOMS 2023) will be held 8-12 May 2023 in Miami FL, USA. First organized in 1988, NOMS 2023 follows the 35 years tradition of NOMS and IM as the primary IEEE Communications Society's forum for technical exchange on network and service operations and management, focusing on research, development, integration, standards, service provisioning, and users communities. The theme of NOMS 2023 is "Integrated Management Towards Resilient Networks and Services." NOMS 2023 will offer various types of sessions, including technical, keynote, experience, poster, panel, and dissertation. High quality is assured through a well-qualified Technical Program Committee (TPC) and stringent peer review of paper submissions. Papers can be submitted as full and short technical session papers, experience session papers, and dissertation papers. In addition, we invite proposals for demonstrations, exhibits, panels, tutorials, and workshops.

Topics of Interest

Authors are invited to submit papers that fall into or are related to the following topics of interest:

Network Management
- IP Networks
- Wireless and Cellular Networks
- Optical Networks
- Virtual Networks
- Home Networks
- Access Networks
- Fog and Edge Networks
- Wide Area Networks
- Enterprise and Campus Networks
- Data Center Networks
- Industrial Networks
- Vehicular Networks
- Internet of Things and Sensor Networks
- Information-Centric Networks

Service Management
- Multimedia Services
- Content Delivery Services
- Cloud Computing Services
- Internet Connectivity and Internet Access Services
- Internet of Things Services
- Security Services
- Context-Aware Services
- Information Technology Services
- Service Assurance

Business Management
- Economic Aspects
- Multi-Stakeholder Aspects
- Service Level Agreements
- Lifecycle Aspects
- Process and Workflow Aspects
- Legal Perspective
- Regulatory Perspective
- Privacy Aspects
- Organizational Aspects

Functional Areas
- Fault Management
- Configuration Management
- Accounting Management
- Performance Management
- Security Management

Management Paradigms
- Centralized Management
- Hierarchical Management
- Distributed Management
- Federated Management
- Autonomic and Cognitive Management
- Policy- and Intent-Based Management
- Model-Driven Management
- Pro-active Management
- Energy-aware Management
- QoE-Centric Management

Technologies
- Communication Protocols
- Middleware
- Overlay Networks
- Peer-to-Peer Networks
- Cloud Computing and Cloud Storage
- Data, Information, and Semantic Models
- Information Visualization
- Software-Defined Networking
- Network Function Virtualization
- Orchestration
- Operations and Business Support Systems
- Control and Data Plane Programmability
- Distributed Ledger Technology

Methods
- Mathematical Logic and Automated Reasoning
- Optimization Theories
- Control Theory
- Probability Theory, Stochastic Processes, and Queuing Theory
- Artificial Intelligence and Machine Learning
- Evolutionary Algorithms
- Economic Theory and Game Theory
- Monitoring and Measurements
- Data Mining and (Big) Data Analysis
- Computer Simulation Experiments
- Testbed Experimentation and Field Trials
- Software Engineering Methodologies

Important Dates:

- **Paper Submission Deadline**: September 12, 2022
- **Notification of Acceptance**: December 16, 2022
- **Final Camera Ready**: January 30, 2023

General Co-chairs:

- **Kemal Akkaya**, *FIU*, USA
- **Olivier Festor**, *Université de Lorraine*, France

TPC Co-chairs:

- **Mohammad Ashiqur Rahman**, *FIU*, USA
- **Carol Fung**, *Concordia University*, Canada
- **Lisandro Zambenedetti Granville**, *UFRGS*, Brazil

Sustainable Communications for Renaissance

The 2023 IEEE International Conference on Communications (ICC) will be held in the Roma Convention Center La Nuvola in Rome, Italy, from 28 May – 1 June 2023. Themed "Sustainable Communications for Renaissance," this flagship conference of the IEEE Communications Society will feature a comprehensive high-quality technical program including 13 symposia and a variety of tutorials and workshops. IEEE ICC 2023 will also include an attractive industry program aimed at practitioners, with keynotes and panels from prominent research, industry and government leaders, business and industry panels, and technological exhibits.

TECHNICAL SYMPOSIA

- IoT & Sensor Networks
- Cognitive Radio & AI-Enabled Networks
- Communication & Information System Security
- Communication QoS, Reliability, & Modeling
- Communication Software & Multimedia
- Communication Theory
- Green Communication Systems & Networks
- Mobile & Wireless Networks
- Next-Generation Networking & Internet
- Optical Networks & Systems
- Signal Processing for Communications
- Wireless Communications
- Selected Areas in Communications :
  - Big Data
  - Cloud Computing, Networking and Storage
  - e-Health- Molecular, Biological and Multi-Scale Communications
  - Satellite & Space Communications
  - Smart Grid Communications
  - Social Networks
  - Machine Learning for Communications
  - Backhaul/Fronthaul Networking and Communications
  - Aerial Communications
  - Quantum Communications & Computing
  - Reconfigurable Intelligent Surfaces

INDUSTRY FORUMS AND EXHIBITION PROGRAM
Proposals are sought for forums, panels, presentations and demos, specifically related to issues facing the broader communications and networking industries.

TUTORIALS
Proposals are invited for half- or full-day tutorials in all communication and networking topics.

WORKSHOPS
Proposals are invited for half- or full-day workshops in all communication and networking topics.

icc2023.ieee-icc.org

IMPORTANT DATES

Paper Submission	**Tutorial Proposals**
11 October 2022	4 October 2022
Acceptance Notification	**Workshop Proposals**
18 January 2023	2 August 2022
Camera-Ready	**Technical Panel Proposals**
15 February 2023	13 December 2022

Submission procedures are available at icc2023.ieee-icc.org

ORGANIZING COMMITTEE

General Chairs
Marco Luise, University of Pisa, Italy
Marco Ajmone Marsan, IMDEA Networks Institute, Spain

Honorary Chairs
Sergio Benedetto, Politecnico di Torino, Italy
Marina Ruggieri, University of Rome "Tor Vergata", Italy

Executive Chair
Roberto Verdone, University of Bologna, Italy

Executive Vice Chair
Luca Sanguinetti, University of Pisa, Italy

Senior Advisor
Stefano Bregni, Politecnico di Milano, Italy

Technical Program Co-Chairs
Meixia Tao, Shanghai Jiao Tong University
Michele Zorzi, University of Padua, Italy

Conference Operations Chairs
Veronica Trastulli, Marco Gazzarrini, First Class srl, Livorno, Italy

Technical Program Vice-Chair
Walid Saad, Virginia Tech, USA

Finance Chair
Stefano Giordano, University of Pisa, Italy

Tutorial Program Co-Chairs
Marco di Renzo, CNRS Paris, France
Neelesh Mehta, Indian Institute of Science

Treasurer
Bruce Wortman, IEEE ComSoc

Workshop Program Co-Chairs
Davide Dardari, University of Bologna, Italy
Angela Yingjun Zhang, Chinese Univ. HK, PRC
Chengjun Sun, Beijing Samsung Telecom R&D Center, PRC

Industry Forums & Exhibition Co-Chairs
Roberto Sabella, Ericsson Research, Italy
Tomoo Takahara, Fujitsu, Japan
Mario Di Dio, CableLabs, USA
Petar Popovski, Aalborg Univ., Denmark
Sanyogita Shamsunder, Verizon, USA

Publications Chairs
Kaibin Huang, HKU, Hong Kong
Michele Segata, Free University of Bozen-Bolzano, Italy

Publicity Co-Chairs
Paweł Kryszkiewicz, Uni Poznan, Poland
Rui Wang, Tongji University, PRC
Omid Semiari, Uni Colorado, USA

Marketing Chair
Venkatesha Prasad Ranga Rao, TU Delft, The Netherlands

Awards Chair
Rui Zhang, NUS, Singapore

ComSoc Project Manager
Melissa Torres, IEEE ComSoc

Web Co-Chairs
Rosario Garroppo, University of Pisa, Italy
Cristina Rottondi, Politecnico di Torino, Italy

Student Travel Grant Co-Chairs
Margot Deruyck, Univ. Ghent, Belgium
Filippo Malandra, University of Buffalo, USA

GIMS Advisor
Luis Correia

GITC Advisor
Christos Verikoukis

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on

https://journals.ieeeauthorcenter.ieee.org/
Then click: "IEEE Author Tools for Journals"
- "Article Templates"
- "Templates for Transactions".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *Index Terms (Keywords)*. For the final version of accepted papers, please send the short cvs and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information. When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue
g) Document Object Identifier (DOI)

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984. DOI: 10.1109/TEI.1984.298778

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to submit their papers electronically via the following portal address:

https://www.ojs.hte.hu/infocommunications_journal/about/submissions

If you have any question about the journal or the submission process, please do not hesitate to contact us via e-mail:

Editor-in-Chief: Pál Varga – pvarga@tmit.bme.hu

Associate Editor-in-Chief:
Rolland Vida – vida@tmit.bme.hu
László Bacsárdi – bacsardi@hit.bme.hu

Special Issue

of the **Infocommunication Journal**

Call for Papers

Internet of Digital Reality: Applications and Key Challenges

A Digital Reality (DR) is a high-level integration of virtual reality (including augmented reality, virtual and digital simulations and twins), artificial intelligence and 2D digital environments which creates a highly contextual reality for humans in which previously disparate realms of human experience are brought together. DR encompasses not only industrial applications but also helps increase productivity in all corners of life (both physical and digital), thereby enabling the development of new social entities and structures, such as 3D digital universities, 3D businesses, 3D governance, 3D web-based digital entertainment, 3D collaborative sites and marketplaces. The Internet of Digital Reality (IoD) is a set of technologies that enables digital realities to be managed, transmitted and harmonized in networked environments (both public and private), focusing on a higher level of user accessibility, immersiveness and experience with the help of virtual reality and artificial intelligence.

This special issue collects the latest results emerging on the field of Cognitive Infocommunications.

Chief Editor:
Prof. Peter Baranyi
Széchenyi István University

Guest Editors:
Prof. György Wersényi
Széchenyi István University
Dr. Ádám Csapó
Széchenyi István University
Prof. Anna Esposito
Università degli Studi della Campania "Luigi Vanvitelli"
Prof. Atsushi Ito
Utsunomiya University, Japan

Important dates:

Submission paper deadline: **30th of September, 2022**
Notification first review: **15th of December, 2022**
Deadline for revised paper: **10th of February, 2023**
Camera Ready: **15th of March, 2022**

Regarding manuscript submission information, please visit:
https://www.infocommunications.hu/for-our-authors

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its 1000 individual members, the Scientific Association for Infocommunications (in Hungarian: Híradástechnikai Tudományos Egyesület, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society.

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **FERENC VÁGUJHELYI** • *elnok@hte.hu*

Secretary-General: **ISTVÁN MARADI** • *istvan.maradi@gmail.com*

Operations Director: **PÉTER NAGY** • *nagy.peter@hte.hu*

International Affairs: **ROLLAND VIDA, PhD** • *vida@tmit.bme.hu*

Address: H-1051 Budapest, Bajcsy-Zsilinszky str. 12, HUNGARY, Room: 502

Phone: +36 1 353 1027

E-mail: *info@hte.hu*, Web: *www.hte.hu*