Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

September 2022 Volume XIV Number 3 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

Selected ICT topics from quantum communications to personalized speech synthesis

Pal Varga

INFOCOMMUNICATIONS Journal covers a broad area of the scientific and engineering spectrum. The current issue is a demonstrative example of this, where topics of quantum communications, software defined networking, software defined radio, blockchain-supported decentralized authentication, IPv6 transition technologies (and their implementation benchmarking), personalized text-to-speech, interference of nationwide deployments for ICT technologies, the traveling salesman problem or microservice-based engineering process optimization of industrial IoT fields get presented together.

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

First, one of the most interesting topics of recent years, the road to Quantum Communications is visited by Daryus Chandra and his colleagues. Their overviews and discussions provide an easy-reading discourse (as they put it), requiring no deep knowledge of the topic. They clearly describe some of the most important challenges and open problems for the coming years. These include creating low-complexity yet powerful short codes for quantum error-correction, connecting medium-sized quantum computers to solve problems connected to quantum search, factorization, and optimization, and further addressing the various challenges that come with the birth of such Quantum Internet.

Following the Quantum line of thought (if there is such a thing), Suadad S. Mahdi and Alharith A. Abdullah propose a hybrid quantum key distribution protocol to enhance the security of SDN and network slices. The SDN control plane – and particularly the openflow channel – is prone to attacks that could have serious effects on the overall transmission and data security. The authors propose a solution that uses a hybrid key consisting of classical and quantum key distribution protocols to provide double security depending on the computational complexity and physical quantum properties.

Software Defined Radio (SDR) is an essential building block of modern communications. Dimitrie Popescu and Rolland Vida provide a comprehensive introduction to the wolds of SDRs, from their technological advancements through the theoretical background to the current SDR architectures and platforms. Their paper illustrates some typical features and applications of SDRs through two case studies.

Hafida Khalfaoui, Abderrazak Farchane, and Said Safi present a novel, BCT (blockchain technology)-based decentralized authentication mechanism for mobile ad-hoc networks. Their approach builds on the cryptographic characteristics of BCT, and they utilize fog computing to ensure communication between admins by delivering the update of blockchain anytime nodes are added to the network.

Next, Omar D'yab presents a comprehensive survey on IPv4 as a Service transition technologies and the performance of five important, comparable implementations: 464XLAT, Dual-Stack Lite, Lightweight 4over6, MAP-E, and MAP-T. These technologies appear in several different products and tools that the paper also describes, leading to the most comprehensive overview of the domain to date.

A related article by Ahmed Al-hamadani and Gábor Lencse details the benchmarking of MAP-T devices, a distinguished implementation of IPv6 transition technologies, standardized by RFC 8219. The authors built a tester, discussed its capabilities in detail, and showed the results of a testbed for validation.

In their paper, Ali Raheem Mandeel, Mohammed Salah Al-Radhi, and Tamás Gábor Csapó conducted detailed experiments for adaptive speaker text-to-speech (TTS) synthesis in which they have limited training samples regarding the target speaker. Their results presented the experiments for finding the minimum dataset and training period required to construct a TTS model with an unseen target speaker's dataset. Their results provide deep foundations for those who build applications with personalized text-to-speech synthesis.

Following the stream of comprehensive tutorial and survey articles in this issue of the Infocommunication Journal, Hussein Taha, Péter Vári, and Szilvia Nagy discuss the challenges related to the interference of cable television networks and mobile/fixed communication networks. They conclude by proposing measures for reducing or mitigating such mutual interference effects.

A topic of general interest – namely, searching for the optimal solution to the traveling salesman problem – is discussed in the next paper by Ali Jawad Ibada1, Boldizsár Tüű-Szabó, and László T. Kóczy. They investigated the effect of the initial population construction on Discrete Bacterial Memetic Evolutionary Algorithm applied to the problem and found that the Circle Group Heuristic gives better results than other well-known heuristic tour construction methods.

Germar Schneider and his colleagues present their recent results on the sensor integration efficiency of the micro-service-based approach by Eclipse Arrowhead, applied in the semiconductor industry. Besides describing the solution in general, they demonstrate its applicability through use-cases. They show that significantly fewer human resources must be utilized with this micro-service-based process engineering approach.

With this brief overview, we wish you a pleasant read – or deep study – of the current Infocommunications Journal issue.

Pal Varga received his Ph.D. degree from the Budapest University of Technology and Economics, Hungary. He is currently the Head of Department of Telecommunications and Media Informatics at the Budapest University of Technology and Economics. 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.

On the Road to Quantum Communications

Daryus Chandra, Panagiotis Botsinis, Dimitrios Alanis, Zunaira Babar, Soon-Xin Ng and Lajos Hanzo

Abstract—**Moore's Law has prevailed since 1965, predicting that the integration density of chips will be doubled approximately every 18 months or so, which has resulted in nano-scale in- tegration associated with 7 nm technologies at the time of writing. At this scale however we are about to enter the transitory range between classical and quantum physics. Based on the brilliant proposition by Feynman a new breed of information bearers was born, where the quantum bits are mapped for example to the spin of an electron. As a benefit, the alluring properties of the nano-scale quantum world have opened up a whole spate of opportunities in signal processing and communications, as discussed in this easy-reading discourse requiring no background in quantum physics.**

I. INTRODUCTION

The Internet has revolutionized our lives. This revolution was catalyzed by the groundbreaking discoveries of information theory, followed by the evolution of integrated circuit technology, which has broadly speaking followed the predictions of Moore's Law ever since 1965. This trend has gradually led to nano-scale integration, where encountering quantum effects is no longer avoidable. The processing of quantum-domain information has to obey the basic postulates of quantum physics, where a so-called quantum bit or *qubit* may be represented as the *superposition* of a logical zero and a logical one. More explicitly, we could visualize this superposition as a coin spinning in a box, hence being in an equiprobable superposition of 'head' and 'tail', so that we can avoid the somewhat unpalatable reference to the famous Schrödinger cat analogy. Metaphorically speaking, we have to carry out all quantum signal processing operations, while the coin is still spinning in the box, because once it has stopped, we can no longer 'manipulate' or process it in the quantum-domain - it has 'collapsed' back into the classical domain. Therefore upon lifting the lid of the box, we can reveal the resultant classical-domain outcome, which is either 'head' or 'tail'.

Another property of the above-mentioned qubits is that they cannot be copied, because trying to copy them would result again in their collapse to the classical domain, hence precluding their further processing in the quantum domain. Instead, the so-called *entanglement* operation has to be used. Intriguingly, entangled qubits have the property that if we change the spin of the electron representing the qubit, that of its entangled pair is also changed at the same instant. However, it has to be mentioned that at the time of writing entanglement has only been demonstrated in practice by relying on classical-domain preparatory operations carried out before the entanglement is established, which had to obey the speed of light. Upon entangling large vectors of qubits, representing the quantum-domain operands parallel processing becomes feasible, hence it also becomes possible to construct so-called quantum computers capable of solving various classically intractable problems. Having said that, these bespoke quantum computers can still be outperformed in certain tasks by classical computers, but they are eminently suitable for tailor-made tasks, which cannot be efficiently solved by classical computers [1]. In parallel to these alluring developments, next-generation communication systems aim for realizing flawless telepresence. It has also been predicted that the number of devices connected to the Internet has outnumbered the entire human population of planet Earth. In this context, the power of superposition and entanglement may be harnessed for efficiently solving various problems, which have hitherto been deemed to be unsolvable in our lifetime.

A striking example demonstrating the power of quantum computing is Grover Quantum Search Algorithm (QSA), which is capable of finding a single solution in an unstructured database having N elements at a complexity order of $\mathcal{O}(\sqrt{N})$, whilst its classical full-search-based counterpart requires on the order of $\mathcal{O}(N)$ cost-function evaluations (CFEs). As wonderful as it sounds, quantum computers also impose a massive threat on classical security and privacy. The most popular public cryptosystem, known as the Rivest-Shamir-Adleman (RSA) algorithm, heavily relies on the hardness of the so-called integer factorization problem. Although this problem is impractical to solve using the current classical computers, this will no longer be the case when a fully functioning quantum computer is available. For instance, the time required for breaking a 2048-bit public key can be reduced from billions of years required by classical computers to a matter of minutes by employing a quantum computer [2].

Fortunately, quantum information processing also provides a wonderful solution for mitigating this emerging threat. Quantum key distribution (QKD) [3]–[5] constitutes one of the already commercialized quantum technologies. QKD circumvents the problem of the impractical, but absolutely secure one-time pad secret key distribution of classical communication. Therefore, QKD will remain provably secure in the face of the physical security attacks that may be carried out by quantum computers. Another impressive development has suggested that it is also possible to directly transmit classical information totally securely over quantum channels, whilst relying on the so-called quantum secure direct communication (QSDC) protocol [6]. This field of finding a novel scheme for securely transmitting classical information using quantum-domain techniques is widely referred to as *quantum cryptography*.

The authors are with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK
(email: {dc3c18, sxn, lh}@soton.ac.uk).

DOI: 10.36244/ICJ.2022.3.1

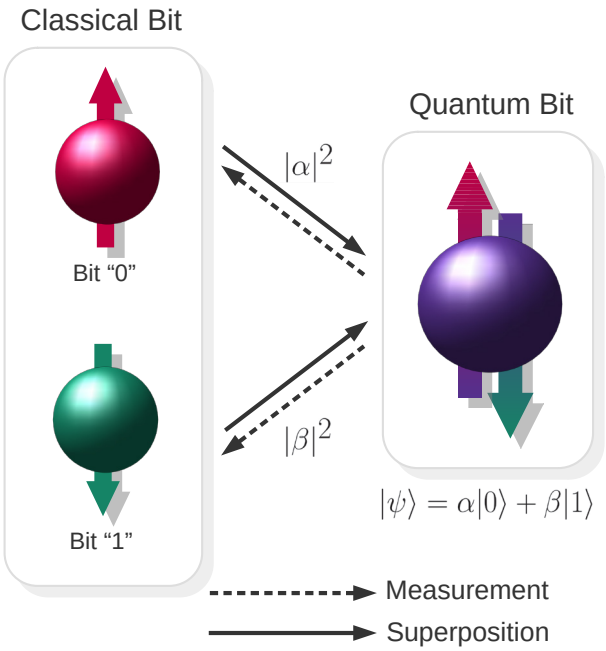


Fig. 1: A quantum bit or *qubit* can be in a *superposition* of two values or states at the same time. However, this superposition will collapse after measurement with a certain probability for each value "0" and "1".

At the time of writing, quantum technologies gradually approach maturity, and hence the exchange of quantum information will become inevitable and eventually ubiquitous. Connecting multiple quantum computers using quantum links potentially offers the capability of outperforming a single quantum computer by creating a larger distributed quantum computer. One of the key requirements for creating such a system is the capability to maintain seamless quantum links among the quantum computers. The vital resource required in this architecture is the so-called maximally-entangled quantum state relying on the Einstein-Podolski-Rosen (EPR) electron-pairs, potentially facilitating an instantaneous action at a distance. This entangled pair is created in a unique superposition state so that any operation applied to one of the particles will immediately affect the other particle, even if they are separated by a great distance - again, provided that the appropriate preparatory entanglement operations have been carried out. Boldly and explicitly, this does not 'violate' the speed of light, because these preparatory communications actions of course have to obey the speed of light, regardless whether they rely on optical fiber or free-space optical links. As quantum technologies become more prevalent in mainstream publications, several questions have emerged concerning what quantum technologies can offer in the realms of communication engineering. Although we have already touched upon them briefly, we would like to elaborate a little further on some promising applications for motivating further research.

II. Quantum-Based Communication

Again, in contrast to classical bits, which can only assume a value of "0" or "1" in any time interval, a qubit can hold both values simultaneously in a form of superposition as shown in Fig. 1. Therefore, N qubits in a state of superposition can be used to hold all the 2^N classical bit combinations simultaneously. Another highly relevant property of quantum information in this context is the *no-cloning theorem*, which we have briefly alluded to above by stating that upon trying to observe the qubits they collapse to the classical domain. In scientific parlance this dictates that no unitary operation can perform a perfect copying operation of a qubit found in an unknown superposition state to another auxiliary qubit. These two properties, in addition to the entanglement, can be exploited for developing several novel communication protocols.

Quantum key distribution (QKD) [3], [4] constitutes one of the most well-known quantum communications protocols, albeit in all truth QKD only represents a secret key negotiation protocol. By relying on the no-cloning theorem and on the fact that the action of 'measurement' or observation collapses the superposition of quantum states to the classical domain, sharing the so-called 'one-time pad' secret key now becomes plausible. The seminal QKD proposal is commonly referred to as the Bennett-Brassard protocol (BB84) [3], which is based on the 'prepare-and-measure' protocol, while the E91 protocol [4] is based on pre-shared EPR electron-pairs.

One of the features of a qubit is that it can be used to convey either quantum information or classical information. While the QKD protocol can be used for the exchange of the classical one-time pad secret key, **quantum superdense coding** [7] supports the secure transmission of classical information through pre-shared EPR electron-pair. This was an early demonstration that instead of acting as the medium of exchanging the secret key, the pre-shared maximally-entangled quantum state can also be used for directly transferring confidential classical information. This ingenious concept was then ultimately further developed by the proposal of **quantum-secure direct communication (QSDC)** [6], which constitutes a fully-fledged confidential quantum-based classical communications protocol, rather than being a pure secret key negotiation procedure. Given the increasing number of mobile devices communicating by broadcasting information, the secrecy and the privacy of the information becomes more crucial than ever. Quantum cryptography may pave the way for providing unbreachable physical layer security for next-generation communication. Naturally, there are numerous open challenges in the way of wide-spread QSDC, such as its limited attainable rate and distance, as well as its reliance on quantum memory, which future research has to tackle. It is also important to highlight that the underlying security-proofs of both the QKD and QSDC protocols are based on the classical information-theoretic physical layer security definitions.

To expound a little further, the direct transfer of quantum information over a quantum channel faces the following challenge. Due to the no-cloning theorem, any quantum information that is lost during its transmission cannot be

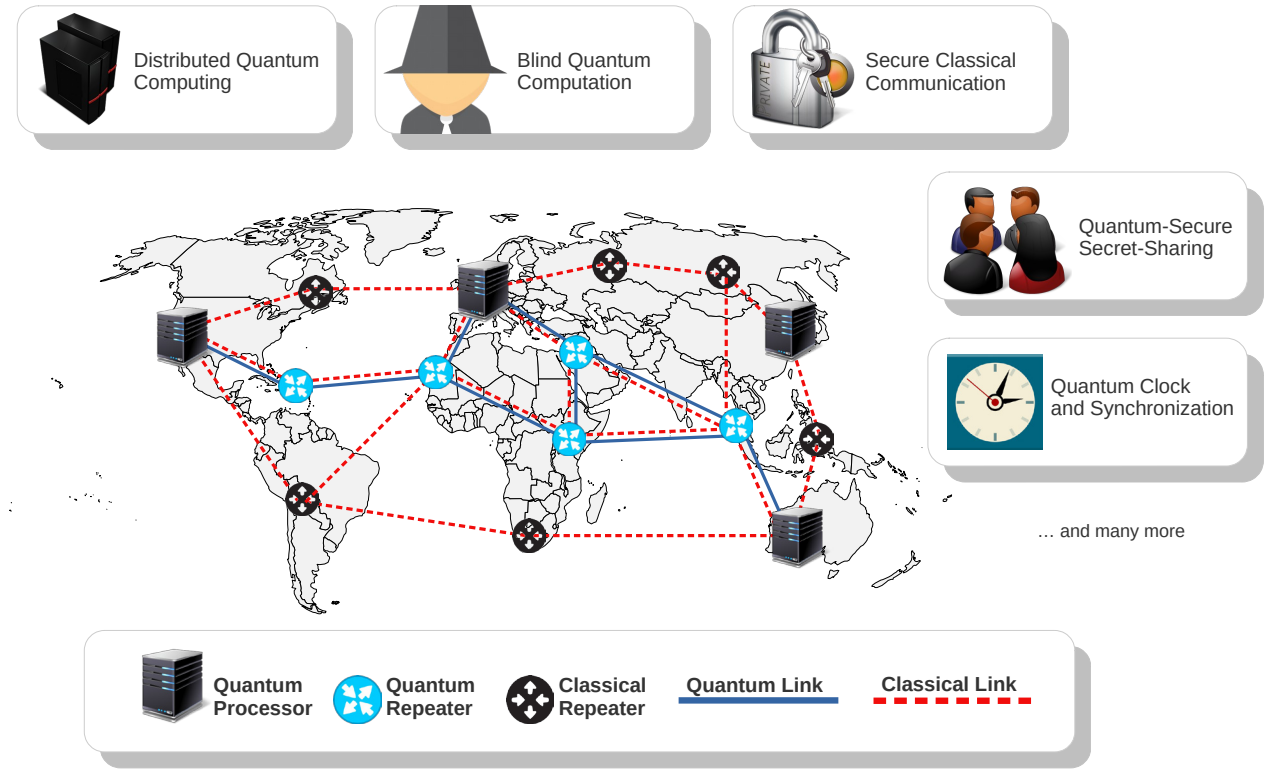


Fig. 2: Stylized vision of the Quantum Internet of the near future, which will rely on a combination of both classical and quantum devices.

readily replaced. Hence the traditional method of ensuring a reliable transmission by sending multiple copies of the same information is no longer feasible. However, the properties of quantum mechanics allow us to transfer quantum information without sending it through the quantum channel with the aid of **quantum teleportation** [8]. The transfer of quantum information can be replaced by the joint action of an EPR electron-pair and classical communication. The employment of quantum teleportation is promising for the following reason. Multiple copies of EPR electron-pairs can be generated, hence an error-control procedure commonly referred to as **entanglement distillation** can be invoked for improving the integrity of quantum communications.

Therefore, a paradigm shift is taking shape concerning the role of repeaters and network coding. For a quantum network, both **quantum repeaters** and **quantum network coding** are indispensable for the reliable distribution of the EPR pairs across multiple nodes in the context of long-distance transmissions. While in classical networks the operation of the repeater is often based on the decode-and-forward mechanism, in the quantum domain the role of the repeater is to maintain connectivity in the form of the seamless generation and sharing of EPR electron-pairs between quantum nodes. To support this functionality, each quantum repeater may rely on the capability of performing **entanglement swapping** and **entanglement distillation**. This, in turn, will hinge on several novel network utilization metrics, which must be considered during the quantum network design of the near future.

The long-term goal in the exploration of quantum computation and communication is to conceive the perfectly secure **Quantum Internet** [9], which is an emerging concept in the landscape of quantum engineering, as portrayed in the stylized illustration of Fig. 2. The concept is reminiscent of that of the classical Internet, interconnecting multiple quantum nodes in the quantum network. The Quantum Internet will facilitate the perfectly secure exchange of quantum information, whilst supporting a plethora of other compelling applications such as distributed quantum computation, blind quantum computation, quantum-secure secret-sharing, and many more [10], [11]. For example, multiple inter-connected quantum computers can jointly act as a distributed quantum computer and can perform more advanced computational tasks than a single quantum computer. However, there are numerous other attractive applications that cannot even be predicted at the time of writing.

III. QUANTUM-SEARCH-AIDED WIRELESS COMMUNICATIONS

The inherent parallelism of quantum information processing equips quantum computers with immense computational power. It has been shown theoretically that there are several classes of problems that can be solved very efficiently by quantum computers, such as integer factorization, finding solutions in large unstructured databases, and large-scale optimization problems, just to name a few. For instance, as we have mentioned briefly earlier, QSAs are capable of finding the correct solution in unstructured databases at a significantly reduced number of CFEs compared to the classical full-search based method. In order to form a clearer picture concerning the taxonomy and the potential applications of various QSAs,

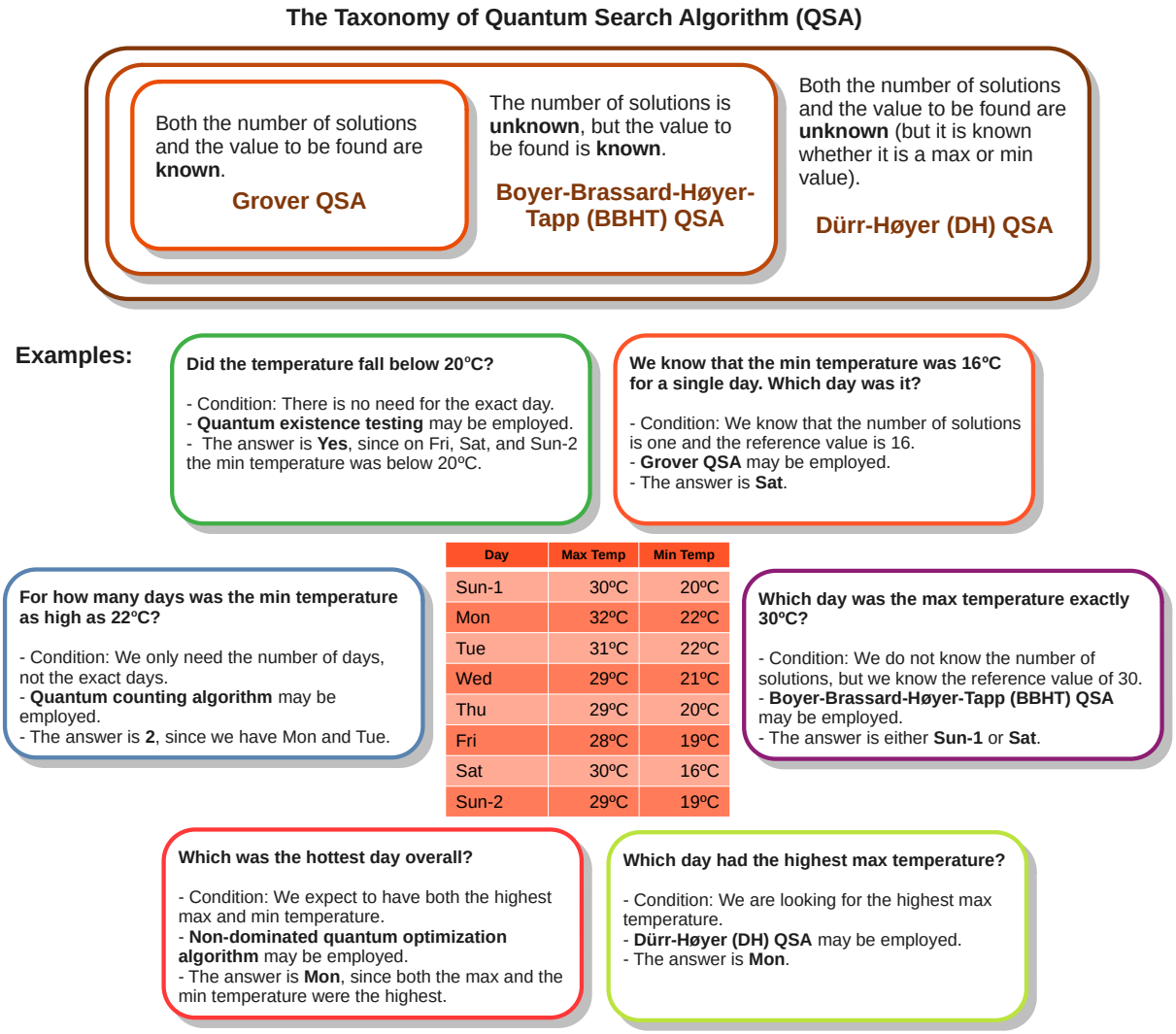


Fig. 3: The taxonomy of QSAs and their hypothetical use cases on weather data.

please refer to Fig. 3. However, the most intriguing question is, how we can exploit the beneficial computational speed-up attained by QSAs in solving the huge gamut of large-scale problems routinely found in classical wireless communications. Hence, this section will be dedicated to a number of applications, which have been shown to be capable of solving diverse problems arising in classical communication [12].

Quantum-Search-Aided Multi-User Detection [14]. The practical realization of the optimal full-search-based solution for classical wireless communication problems, such as the maximum-likelihood multiuser detector (ML-MUD), is hindered by its potentially excessive computational complexity. To circumvent this impediment, Grover's QSA may be invoked by exploiting the inherent parallelism of quantum information processing for approaching the ML-MUD's performance at a significantly lower number of cost function evaluations (CFEs). More explicitly, a derivative of Grover's QSA - namely the Durr-Høyer QSA - is capable of finding the correct solution with almost 100% probability after only evaluating on the order of $\mathcal{O}(\sqrt{N})$ CFEs, as opposed to the classical full-search-based solution requiring on the order of $\mathcal{O}(N)$ CFEs.

Quantum-Search-Aided Multi-Objective Routing [15], [16]. The emergence of the Internet of things (IoT) has motivated the development of the so-called self-organizing networks (SONs). Compared to conventional networks, SONs may act autonomously for achieving the best possible network performance. Thus, the underlying routing protocols should be capable of striking a delicate compromise amongst a range of conflicting quality-of-service (QoS) requirements. However, as the network size increases in terms of the number of nodes, finding the optimal solution typically becomes a non-polynomial-hard (NP-hard) search problem. Moreover, the employment of single-component objective functions relying on pure capacity or sum-rate maximization, on power or energy minimization or alternatively on dely or complexity minimization do not necessarily lead to attractive well-balanced system design. As a remedy, the concept of Pareto optimality comes to rescue in the context of multi-component optimization, which is capable of amalgamating various potentially conflicting design objectives. In this scenario, the Pareto

front represents the collection of all optimal solutions, where none of the parameters involved in the objective function can be improved without degrading at least one of the others, as exemplified by the power versus bit-error ratio (BER) trade-off, just to mention one of them. Although a plethora of bio-inspired algorithms may be tailored specifically for solving multi-objective optimization problems, they often fail to generate all the optimal solutions constituting the Pareto front. As an attractive alternative, a quantum-aided multi-objective optimization algorithm may be invoked, which is capable of finding all Pareto optimal routes at a dramatically reduced number of CFEs. The complexity of finding the best route can be reduced to the order of $\mathcal{O}(N)$ and $\mathcal{O}(N\sqrt{N})$ CFEs in the best- and the worst-case scenarios, respectively, which corresponds to a substantial complexity reduction from the order of $\mathcal{O}(N^2)$ CFEs imposed by the classical full-search-based solution.

Quantum-Search-Aided Non-Coherent Detection [17]. With the proliferation of wireless devices in support of ubiquitous connectivity, solving large-dimensional search problems, such as cooperative multicell processing in areas of high user density - such as airports, train stations, and densely-populated metropolitan areas - imposes a major challenge. In these scenarios, an accurate estimation of every single channel gains is required for performing a coherent detection. However, every time the Doppler frequency is doubled, the pilot overhead used for sampling the channel's complex-valued envelope also has to be doubled. Consequently, both the pilot overhead and the detection complexity escalate rapidly as the Doppler frequency increases. Hence, a differentially encoded modulation scheme relying on non-coherent detection constitutes an attractive design alternative, since it may be invoked for mitigating the pilot overhead required for channel estimation, albeit naturally, at the cost of some performance erosion. As a beneficial solution, quantum-search-aided multiple-symbol differential detection may be employed for matching the performance of the classical full-search-based multiple-symbol differential detectors, despite requiring a significantly reduced number of CFEs.

Joint Quantum-Search-Aided Channel Estimation and Data Detection [18]. Joint channel estimation and multi-user detection (MUD) is capable of approaching the performance of perfect channel estimation by iteratively exchanging soft extrinsic information between these two components of the receiver. A quantum-aided repeated-weighted boosting search (QRWBS) algorithm may be readily combined with a quantum-search-aided MUD for performing iterative channel estimation and data detection in the uplink of MIMO-aided orthogonal frequency-division multiplexing (OFDM) systems. As an additional benefit, this powerful system is capable of operating in rank-deficient scenarios, where the number of receive antenna elements (AEs) at the base station (BS) is lower than the number of users transmitting in the uplink. Furthermore, the QRWBS-based channel estimation is capable of outperforming its classical counterpart, despite requiring a substantially lower number of CFEs, which is an explicit benefit of invoking iterative information exchange between the MUD, the channel estimator, and the channel decoders at the BS's receiver.

Quantum-Search-Aided Localization [19]. For various compelling applications of the next-generation communication technology - as exemplified by assisted living and the assignment of users to radio-frequency (RF) as well as to visible light communication (VLC) and narrow-beam millimeter-wave (mm-Wave) access points - the position of the users has to be accurately estimated. Furthermore, indoor localization may be used for creating new applications, such as personalized marketing and shopping experience. Therefore, there is a mutually beneficial relationship in the development of indoor localization, VLC, as well as mm-Wave and THz systems. However, the computational complexity of carrying out full-search based finger-printing based localization for numerous VLC-based and mm-Wave-based localization may become excessive. This is because it requires the knowledge of the entire room's topology, which has to be partitioned into numerous finely-grained tiles of a virtual grid. In this context, quantum-search-aided localization algorithm may be invoked for achieving the same localization accuracy as the classical full-search-based solution at substantially reduced number of CFEs.

Suffice to say in conclusion of this section that many more attractive applications can be found in the literature and some others are yet to be discovered. Quantum technology has opened new avenues for solving problems that previously were excessively complex to solve. This gives us the perfect timing to revisit the hitherto unsolved problems of the classical signal processing and communications domain - QSAs might provide the long-awaited answers.

IV. QUANTUM DECOHERENCE

The gravest challenge in constructing a large-scale quantum computer is how to mitigate the deleterious effects of quantum decoherence, which inevitably affects the results of quantum computation or communication tasks, just like the Brownian motion of electrons imposes the ubiquitous Gaussian noise in the classical receivers. Completely isolating the qubits from any environmental influence is practically impossible, hence the mitigation of these effects is paramount. The deleterious circuit-impairments imposed by quantum decoherence are typically modeled by the so-called quantum depolarizing 'channel', even though this does not actually entail transmission over a communications channel. After all, the classic Gaussian 'channel' is also a simple abstraction representing the undesired effects of the above-mentioned Brownian motion of electrons. The demonstration of the quantum depolarizing channel effects inflicted on the performance of Grover QSA is portrayed in Fig. 4. Observe from Fig. 4 that as the depolarizing probability increases, the success probability of finding the corrrect solution tends towards $\frac{1}{2^N}$, which is equivalent to the random decisions of the classical full-search-based method

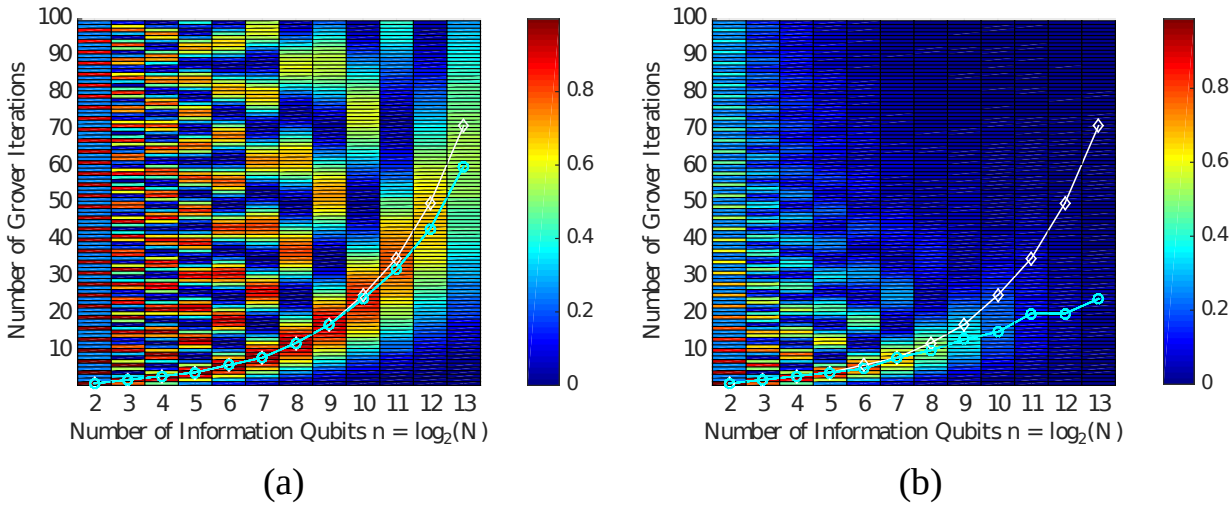


Fig. 4: The color map portraying the success probability of Grover QSA finding the correct solution in the face of quantum depolarizing channel, where the values of the depolarizing probability are given by (a) $p = 10^{-3}$ and (b) $p = 10^{-2}$. The white line represents the scenario of an ideal Grover QSA obtaining the correct solution with a probability of ≈ 1, while the blue line represents the maximum probability values of obtaining the correct solution for imperfect Grover QSA [13].

operating in the face of an excessively hostile channel. Hence as expected, the advantage of QSAs will erode in the face of excessive quantum decoherence [13].

The employment of quantum error-correction codes (QECC) is one of the most potent design alternatives of mitigating the decoherence. Even though error correction has been shown to perform well in the classical domain, implementing the QECCs imposes its own challenges. Indeed, any error correction procedure - both classical and quantum - depends on attaching redundancy to the information, which will be invoked at the decoder for error correction [20]. In the classical domain, the effect of noise in the encoder and decoder circuitry may be deemed negligible in comparison to the noise inflicted by the transmission channel. However, in the quantum domain both the QECC encoder and decoder circuitry impose more substantial imperfections, which simply cannot be ignored. A further challenge is that we additionally have to deal with the specific quantum-domain phenomenon of error proliferation, because a single quantum-gate error encountered by a quantum encoder may in fact precipitate multiple component errors, rather than simply passing on its input errors without proliferating them. This motivates the design of inherently fault-tolerant quantum computation, which is capable of correcting both the self-inflicted errors imposed by its own encoder and decoder as well as the errors caused by the quantum channel.

V. CHALLENGES AND OPEN PROBLEMS

Quantum signal processing relies on delicate quantum particles, such as photons and electrons. Hence, any interaction with the surrounding environment will compromise the integrity of the desired operation. An immeasurable amount of effort has been invested in trying to minimize the presence of decoherence by perfecting the hardware implementation of the qubits as well as by developing sophisticated error correction procedures. Many of the QECC techniques are rooted in their classical counterparts. However, to achieve an excellent error correction performance, long QECC codewords are required, which have to rely on a large number of qubits [20]. The problem with this approach is that at the time of writing most quantum circuits have a shorter coherence time than the time required for carrying out the decoding of long QECCs. Hence, low-complexity yet powerful short codes are required for mitigating the effects of short coherence times.

Another aspect requiring substantial attention is to find meaningful applications, where the unique benefits of quantum computing may be exploited, even if they only have the modest capability of handling just a few hundred qubits [21]. To elaborate a little further, quantum search, factorization, and optimization algorithms tend to require thousands to millions of qubits. Therefore, an intriguing idea is to connect many medium-sized quantum computers with the aid of the Quantum Internet relying on teleportation protocols for creating more powerful quantum computers. Some attractive applications are constituted by the variational quantum eigensolver (VQE) and the quantum approximate optimization algorithm (QAOA) [22].

Finally, to fully realize the Quantum Internet, a whole suite of quantum computers relying on superconducting, trapped ion, nuclear magnetic resonance, optical, and other technologies have to be benchmarked. Furthermore, the entire gamut of quantum links, such as free space terrestrial, satellite, fiber optic, and other connections will have to be further developed. Similarly, sophisticated protocols, such as for example, routing, multiple access, as well as repeat-and-request solutions will require massive standardization efforts. Indeed, the road to the perfecty secure quantum communications era is inevitably a rocky one, which requires the collaboration of the entire IEEE community. This is why about half-a-dozen IEEE Societies have formed a New Initiative in Quantum Engineering (https://qce.quantum.ieee.org) and the new multi-disciplinary open access journal of quantum engineering

(https://quantum.ieee.org/publications). **Valued Colleague, we invite you to join this exhilarating multi-disciplinary journey to solve some of the above-mentioned problems of true frontier-research into Communications v2.0.**

REFERENCES

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, pp. 505–510, 2019; DOI: 10.1038/s41586-019-1666-5

[2] R. Van Meter, K. M. Itoh, and T. D. Ladd, Architecture-Dependent Execution Time of Shor's Algorithm, pp. 183–188. *World Scientific*, 2008, DOI: 10.1142/9789812814623_0029

[3] C. H. Bennett and G. Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing," in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, pp. 175–179, 1984; DOI: 10.48550/arXiv.2003.06557

[4] A. K. Ekert, "Quantum Cryptography Based on Bell's Theorem," *Physical Review Letters*, vol. 67, no. 6, 1991; DOI: 10.1103/PhysRevLett.67.661

[5] L. Gyongyosi, L. Bacsardi, and S. Imre, "A survey on quantum key distribution," *Infocommunications journal 11 (2), 14-21*, vol. 11, no. 2, pp. 14–21, 2019; DOI: 10.36244/ICJ.2019.2.2

[6] Z. Sun, L. Song, Q. Huang, L. Yin, G. L. Long, J. Lu, and L. Hanzo, "Toward Practical Quantum Secure Direct Communication: A Quantum-Memory-Free Protocol and Code Design," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5778–5792, 2020; DOI: 10.1109/TCOMM.2020.3006201

[7] C. H. Bennett and S. J. Wiesner, "Communication via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States," *Physical Review Letters*, vol. 69, no. 20, 1992; DOI: 10.1103/PhysRevLett.69.2881

[8] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels," *Physical Review Letters*, vol. 70, no. 13, 1993; DOI: 10.1103/PhysRevLett.70.1895

[9] M. Caleffi, D. Chandra, D. Cuomo, S. Hassanpour, and A. S. Cacciapuoti, "The Rise of the Quantum Internet," *Computer*, vol. 53, no. 6, pp. 67–72, 2020; DOI: 10.1109/MC.2020.2984871

[10] R. Van Meter and S. J. Devitt, "The Path to Scalable Distributed Quantum Computing," *Computer*, vol. 49, no. 9, pp. 31–42, 2016; DOI: 10.1109/MC.2016.291

[11] J. F. Fitzsimons, "Private Quantum Computation: An Introduction to Blind Quantum Computing and Related Protocols," *npj Quantum Information*, vol. 3, no. 1, pp. 1–11, 2017; DOI: 10.1038/s41534- 017-0025-3

[12] P. Botsinis, D. Alanis, Z. Babar, H. V. Nguyen, D. Chandra, S. X. Ng, and L. Hanzo, "Quantum Search Algorithms for Wireless Communications," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1209–1242, 2018; DOI: 10.1109/COMST.2018.2882385

[13] P. Botsinis, Z. Babar, D. Alanis, D. Chandra, H. V. Nguyen, S. X. Ng, and L. Hanzo, "Quantum Error Correction Protects Quantum Search Algorithms Against Decoherence," *Scientific Reports*, vol. 6, no. 1, pp. 1–13, 2016; DOI: 10.1038/srep38095

[14] P Botsinis, S-X. Ng and L. Hanzo: Quantum Search Algorithms, Quan- tum Wireless, and a Low-Complexity Maximum Likelihood Iterative Quantum Multi-User Detector Design, IEEE Access 2013, Vol. 1; DOI: 10.1109/ACCESS.2013.2259536

[15] D Alanis, P Botsinis, Z Babar, H.V. Nguyen, D. Chandra, S-X. Ng. L. Hanzo: A Quantum-Search-Aided Dynamic Programming Framework for Pareto Optimal Routing in Wireless Multihop Networks, IEEE Transactions on Communications, 2018, Vol. 66, Issue 8; DOI: 10.1109/TCOMM.2018.2803068

[16] D. Alanis, P. Botsinis, Z. Babar, H.V. Nguyen, D. Chandra, S-X. Ng. L. Hanzo: Quantum-Aided Multi-Objective Routing Optimization Using Back-Tracing-Aided Dynamic Programming Dimitrios Alanis; Panagiotis Botsinis; Zunaira Babar; Hung Viet Nguyen; Daryus Chandra; Soon Xin Ng; Lajos Hanzo, IEEE Transactions on Vehicular Technology, 2018, Vol. 67, Issue 8; DOI: 10.1109/TVT.2018.2822626

[17] D. Alanis, P. Botsinis, Z. Babar, S-X. Ng. L. Hanzo: Noncoherent Quantum Multiple Symbol Differential Detection for Wireless Systems IEEE Access, 2015, Vol 3; DOI: 10.1109/ACCESS.2015.2432015

[18] D. Alanis, P. Botsinis, Z. Babar, S-X. Ng. L. Hanzo: Joint Quantum-Assisted Channel Estimation and Data Detection IEEE Access, 2016, Vol 4, DOI: 10.1109/ACCESS.2016.2591903

[19] P. Botsinis, D. Alanis, S Feng, Z. Babar, H. V. Nguyen, D. Chandra, S. X. Ng, R Zhang and L. Hanzo, Quantum-Assisted Indoor Localization for Uplink mm-Wave and Downlink Visible Light Communication Systems, IEEE Access, 2017, Vol 5 DOI: 10.1109/ACCESS.2017.2733557

[20] Z. Babar, D. Chandra, H. V. Nguyen, P. Botsinis, D. Alanis, S. X. Ng, and L. Hanzo, "Duality of Quantum and Classical Error Correction Codes: Design Principles and Examples," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 970–1010, 2018; DOI: 10.5258/SOTON/D0616

[21] J. Preskill, "Quantum Computing in the NISQ Era and Beyond," Quantum, vol. 2, no. 79, pp. 1–20, 2018; DOI: 10.22331/q-2018-08-06-79

[22] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. Obrien, "A Variational Eigenvalue Solver on A Photonic Quantum Processor," *Nature Communications*, vol. 5, 2014; DOI: 10.1038/ncomms5213

Daryus Chandra received the B.Eng. and M.Eng. degrees from Universitas Gadjah Mada (UGM), Indonesia, in 2013 and 2014, respectively, and the Ph.D. degree from the University of Southampton, UK, in 2020. Currently, he is a research fellow at the University of Southampton.

Panagiotis Botsinis received the M.Eng. degree from the National Technical University of Athens (NTUA), Greece, in 2010 and the M.Sc. (Hons) and Ph.D. degrees from the University of Southampton, UK, in 2011 and 2015, respectively. Currently, he is a wireless system engineer at Apple Inc., Germany. He is a member of the Technical Chamber of Greece.

Dimitrios Alanis received the M.Eng. degree from the Aristotle University of Thessaloniki, Greece, in 2011 and the M.Sc. and Ph.D. degrees from the University of Southampton in 2012 and 2017, respectively. Currently, he is a wireless system en- gineer at Apple Inc., Germany.

Zunaira Babar received B.Eng. degree from the National University of Science and Technology (NUST), Pakistan, in 2008, and the M.Sc. (Hons) and Ph.D. degrees from the University of Southampton, UK, in 2011 and 2015, respectively. Currently, she is a senior algorithm engineer at VIAVI Solutions Inc., UK.

Soon Xin Ng (SMIEEE) received the B.Eng. degree (First Class) in electronic engineering and the Ph.D. degree in telecommunications from the University of Southampton, UK, in 1999 and 2002, respectively. Since 2003, he has been an academic staff at the same university and he is currently a Professor of Next Generation Communications.

Lajos Hanzo (FREng, FIEEE) received his doctorate in 1983 and since 1986 he has been with the University of Southampton. He holds honorary doctorates from the Technical University of Budapest (2009) and the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences.

Enhanced Security of Software-defined Network and Network Slice Through Hybrid Quantum Key Distribution Protocol

Suadad S. Mahdi[1,2] and Alharith A. Abdullah[1]

***Abstract*—Software-defined networking (SDN) has revolutionized the world of technology as networks have become more flexible, dynamic and programmable. The ability to conduct network slicing in 5G networks is one of the most crucial features of SDN implementation. Although network programming provides new security solutions of traditional networks, SDN and network slicing also have security issues, an important one being the weaknesses related to openflow channel between the data plane and controller as the network can be attacked via the openflow channel and exploit communications with the control plane. Our work proposes a solution to provide adequate security for openflow messages through using a hybrid key consisting of classical and quantum key distribution protocols to provide double security depending on the computational complexity and physical properties of quantum. To achieve this goal, the hybrid key used with transport layer security protocol to provide confidentiality, integrity and quantum authentication to secure openflow channel. We experimentally based on the SDN-testbed and network slicing to show the workflow of exchanging quantum and classical keys between the control plane and data plane and our results showed the effectiveness of the hybrid key to enhance the security of the transport layer security protocol. Thereby achieving adequate security for openflow channel against classical and quantum computer attacks.**

***Index Terms*—hybrid key, openflow protocol, quantum key distribution, software-defined networking, network slicing, transport layer security.**

I. INTRODUTION

Software-defined networking (SDN) is an emerging and rapidly growing technology that separates the control plane from the network devices in order to give more flexibility to control the network, based on specific policies and security enforcements [1]. The advantages of SDN and virtualization have inspired the creation of network slicing (NS) and contributed to build the infrastructure of 5G networks [2].

NS came up to address the problem of growing network services [3]. Previously, the prevailing concept in networks was "one size fits all", but this concept does not apply to the fifth-generation networks and beyond, the reason is due to different network requirements of heterogeneous applications.

[1] University of Babylon, Babil, Iraq;
E-mail: {suadadsafaa, alharith}@itnet.uobabylon.edu.iq
[2] Al-Mustaqbal University College, Babil, Iraq;
E-mail: suadad.safaa@mustaqbal-college.edu.iq

DOI: 10.36244/ICJ.2022.3.2

Today, with network virtualization technology and software-defined networks, and the ability to abstract resources, the concept of network slicing is ready to create programmable network slices isolated from each other and release them to the real world.

The concept of network slicing is depicted in Figure 1, which allows for the establishment of logical networks for various types of services.

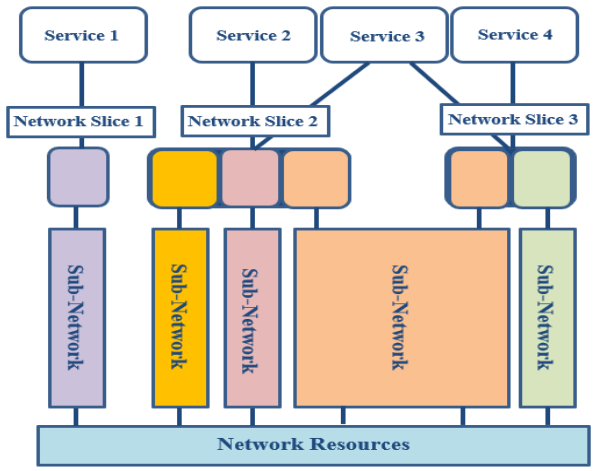


Fig. 1. Network slicing concept.

Despite the flexibility of SDNs, it has properties that can be considered as traps for network attackers such as the network information's centralization in the controller [4]. As a result, anyone can reach the servers hosting the control software, also, centralization means the SDN architecture has a one point of failure that makes the attackers direct their attacks to the controller. For instance, a malicious application (or controller) can be employed to reprogram the whole network to purposes of data stealing from the data center.

The controller has been able to enhance the network security by taking advantage of the global network view feature by running security applications in the controller in order to detect attacks and thus address them in addition to helping to understand the nature of the network in various threats, incidents and security vulnerabilities [5]. The controller uses the information collected and analyzed to enforce the appropriate security policies and thereby improve data plane security. But the idea of the SDN that stems from the decouple of control part from network devices has led to

the need to add new components to the network, namely the controller and communication channels between the planes, which include many challenges that pose security problems that deserve attention.

The Fig. 2 illustrates some critical security threats in SDN. Some of them are popular in the present networks and some other threats are more specific in SDN [6]. But the most dangerous attack is the one which exploits any vulnerability to access the controller and thus destroys the entire network.

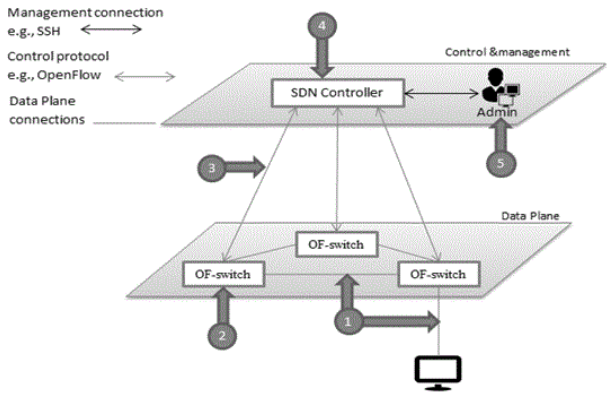


Fig. 2. Security threats in SDN architectures.

As shown in the figure, the threat 1, 2 and 5 are already present in traditional networks. On the contrary, from the threats 3 and 4 are specific to SDN and that stem from decoupling the controller from the network devices and making them centralized. Third threat is seen as the most dangerous, because the network process can be compromised. This type of threat is specific to SDN networks. The attacks focus on control plane communications (such as openflow channel) through which denial of service attacks or data theft are generated.

The danger of this attack depends on the access to the control plane, then it is able to collect enough power (in terms of the number of SDN switches[1] under its control) to launch distributed denial-of-service (DDoS) attacks, in additional to Man-in-the-Middle (MitM) attacks. This is due to a lack of authentication between the controller and SDN switches that makes it easier to create a virtual black hole network allowing data to leak during normal production flows. In essence, Transport Layer Security (TLS) protocol is necessary to protect connections in the SDN. However, relying on TLS use alone is not enough and as is well known in the security community, the use of SSL / TLS does not guarantee a secure connection especially with the advent of quantum computers [7][8], which can perform calculations very quickly, and have significantly affected classic security protocols.

The progress in quantum physics has led to thinking of new ways to ensure security in communication [9]. Quantum key distribution (QKD) is a suitable technology for securing network communication channels, where a single or entangled quantum state is transferred between two parties [10]. Each of parties has two channels: the quantum channel for the exchange of photons and the classic public channel to check for eavesdropping [11]. If a third party makes measurement of the transferred quantum, both of party will discover an eavesdropper presence on the communication channel based on the rules of the mechanics of quantum and the no-cloning theorem [12]. Several researchers have presented work on using quantum protocols to achieve key distribution instead of using traditional cryptographic methods (e.g., RSA), where Czermann, Márton et al. [13] demonstrated the successful distribution of quantum keys using the BB84 protocol in practice on a fiber-optic system.

This work explained how to use hybrid key [14] for key exchange and thus secure openflow channel via quantum TLS (QTLS) in SDN and NS. This paper is an extension of the work in [15] but in this paper the methodology is proposed on the NS environment. In particular, the hybrid key is considered the best solution to achieve authentication between the two parties based on quantum properties in addition to providing a strong key to supply double security based on mathematical complexity of classical method and physical properties of quantum protocol.

This paper is organized as follows. Section II clarifies related work and reveals its boundaries. Section III explains the steps and workflow used for securing openflow channel. The implementation, along with some of our test findings and evaluations, are presented in Section IV. Section V discusses security analysis, while this paper concludes in Section VI.

II. RELATED WORK

In this section, previous studies concerning openflow security in SDN and NS were shown. In [16] authors proposed a quantum key distribution (QKD) and encryption algorithm one time pad (OTP) to encrypt openflow protocol messages, which is named QKDFlow. This scheme is considered a solution that aims to block the MitM attack and thus secure openflow messages in SDN. While the researchers at [17] suggested an identity-based cryptography protocol (IBC) to secure software-defined networks connections, especially for controller communications with network devices within the data plane. Where they suggested that the role of private key generator (PKGs) be transferred to the controller to create the private keys for the network devices and thus reduce the load of PKGs managing the controller.

The authors in [18] discussed security for openflow communication protocol in SDNs by using the security protocol TLS and discussing the security loopholes in TLS. As a result, they proposed a change in TLS handshake protocol to achieve authentication between the parties by adding messages containing the random number, timestamp and hello message ID to revalidation of client and server status before sending finished messages. Therefore, based on this, the MitM attacker is prevented because of the timestamp of the response since the time taken by the attacker to decrypt a random number would certainly exceed the timeframe of the client's response to the server request. While, authors explain in [19] a method for detecting DDoS attacks in the SDN depended on the quantum parameters of QKD system, such as the secret key rate (SKR) and quantum bit error rate (QBER). Where the controller monitors the QBER, if the rate exceeds the

[1] Term SDN switch is example of data plane devices.

threshold limit, the controller makes a decision to change the path and thus mitigate DDoS.

The authors in [20] offered a solution for achieving effective and secure service-oriented authentication for 5G IoT applications, including network slicing and fog computing, to assure anonymity, user credibility, and service data confidentiality. Users are authenticated by utilizing access credentials produced by the IoT server, which allow them to access the IoT service. Otherwise, the attacker would be unable to do so without a legitimate access credential. While authors in [21] developed a hybrid strategy to protect communications between 5G network slices in distinct public cryptosystems, and two heterogeneous cipher schemes to achieve reciprocal communications between the public key infrastructure (PKI) and Certificate Less Public Key Cryptography (CLC) environments.

The authors in [22] introduced a security solution to address security problems related to data exchange in software-defined networks. Where proposed that the TLS use of protocol between the SDN nodes to provide adequate security for communication channel. In addition to use an integrated security module to enhance the security of communications through the application of the access control list (ACL), Strengthening of the TLS protocol configuration and contribute to minimizing the impact of private key hijacking.

Authors in [23] propose a key-distribution scheme suitable for the network slicing architecture when the slices are accessed by third-party applications. The proposed scheme consists of two technologies, the first is Shamir's secret sharing to distribute and rebuild private key shares, and the second technique is ElGamal cryptosystem to encrypt and decrypt the separator keys.

The authors in [24] focused on exploring the concerns of a distributed denial of service attack on a network slicing and presented a model based on deep learning to create a robust network slicing framework to proactively combat DDoS attacks and eliminate overburdened connections before they impact and invade 5G networks.

While the researchers in [15] presented a mechanism for implementing a quantum hybrid protocol with the classic protocol to achieve security for the openflow channel by encrypting messages between the controller and network devices in the software-defined networks.

III. Proposed Methodology for Securing OpenFlow Channel

The use of authentication and encryption is the most important security measures to protect communications. So, in the proposed methodology, the hybrid key [14] is used in the TLS protocol to add new way of authentication based on physical properties of quantum as well as improving encryption process depending on the hybrid key produced by two systems, the first one depends on the computational complexity and the other depends on the quantum properties of the QKD.

The main goal is to secure the communication between the controller and network devices in SDN and NS, thus secure openflow messages. So, there are many messages between controller and SDN switches before exchanging encrypted openflow messages as shown in Figure 3.

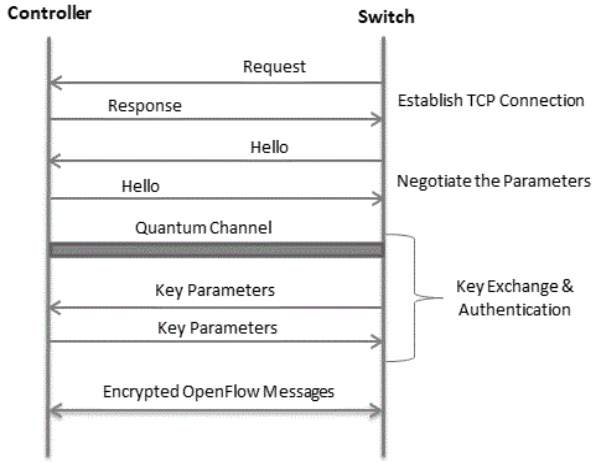


Fig. 3. Controller-switch channel.

In the first step, TCP connection was established, the network device in data plane initiates a request to connect the controller based on the IP and port of controller, and then it responds to this request.

In the second step, session parameters were negotiated to open a secure channel between controller and SDN switch (in data plane) over TLS, Hello message contains the main security parameters.

While in the third step, the key exchange begins between the controller and SDN switch, which is used in the encryption. This step begins by opening a quantum channel and achieving the authentication based on the quantum properties. Then the quantum and classical parameters are exchanged to establish a hybrid key. These steps are known as TLS handshake protocol.

Finally, the hybrid key is passed to the TLS Record protocol to encrypt the communication channel between the controller and SDN switch by using the AES-256 encryption algorithm, thus confidentiality for openflow messages is achieved.

In Figure 4 we summarized the operation of the proposed methodology where a switch initiates a connection to the controller. When the controller receives the connection request, it checks whether the switch supports QKD protocol. If not, the connection will be established by use standard TLS protocol. Otherwise, controller will open the quantum channel and exchange the parameters of the hybrid key. Then, the controller checks if quantum bit error rate (QBER) is less than the threshold (threshold limit of 0.5 has been used) then pass otherwise the key exchange phase is repeated. After this the hybrid key is generated and the quantum TLS handshake protocol is completed, and communication based on openflow between the controller and the switch begins via the QTLS channel.

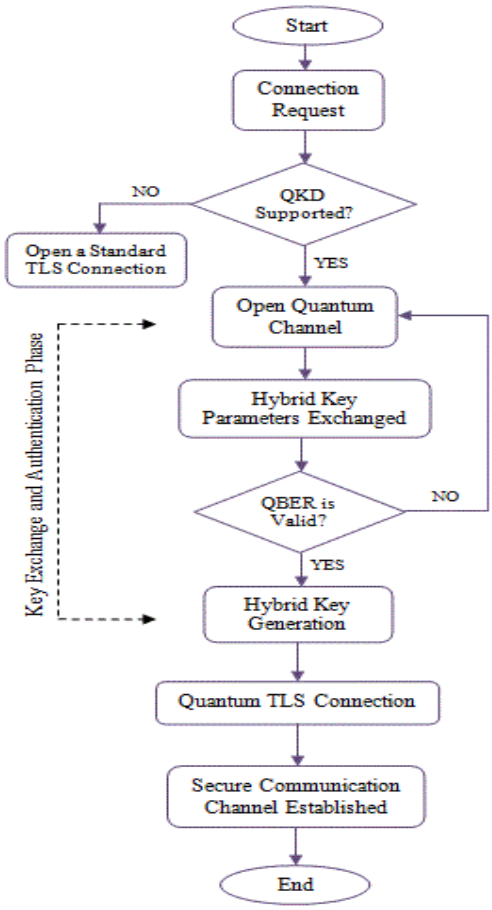


Fig. 4. Operation of the proposed methodology

IV. Implementation Results and Evaluation

A. SDN-Testbed

To do our work, the network environment is initialized depended on the SDN testbed as explain in Figure 5.

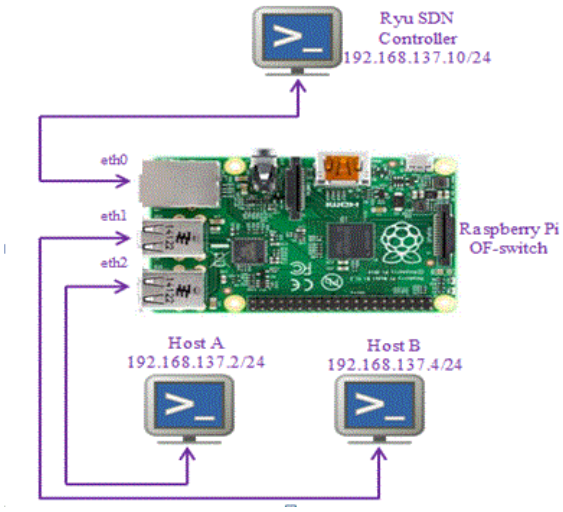


Fig. 5. Environment of SDN-testbed.

In our work, raspberry pi board was converted to openflow switch (SDN switch) to use in our SDN-testbed. We can install and configuring the open vswitch to convert Raspberry Pi as an openflow switch based on the following steps:

- The "ovs-vsctl" was used to create a bridge and add three interfaces (eth0, eth1, eth2) to it through using the add-port command.
- Interfaces were configured and they enable the links connection with three Laptops via Ethernet interface through by using USB-to-Ethernet adapters. One Laptop was used as a ryu controller to manage the open vswitch remotely and two laptops as hosts for test.
- The openflow protocol 1.3 was enabled on the bridge and used the -O option to enable the openflow version in ovs-ofctl.
- Then, the bridge was connected to a remote controller by using IP address and port number of SDN controller.
- Next, the fallback mode of OVS was set to secure mode.

Based on the previous steps, the raspberry pi was converted to openflow switch in low cost [25]. While ryu controller was used depending on Ubuntu 18.04 to configuration of the openflow protocol to allow the controller to program the openflow switch (raspberry pi).

B. Network Slicing Implementation

The scenario for implementing network slicing based on SDN consists of several controllers and a single owner. SDN proxy (Network Hypervisor) plays a key role in dividing the network infrastructure into many virtual networks, and the infrastructure owner is usually the person who controls the SDN proxy [26].

Multiple virtual tenants can use this scenario to deploy their SDN controllers on the network slices management infrastructure and maintain isolation between them. Figure 6 shows the structure of the SDN proxy in the slicing environment.

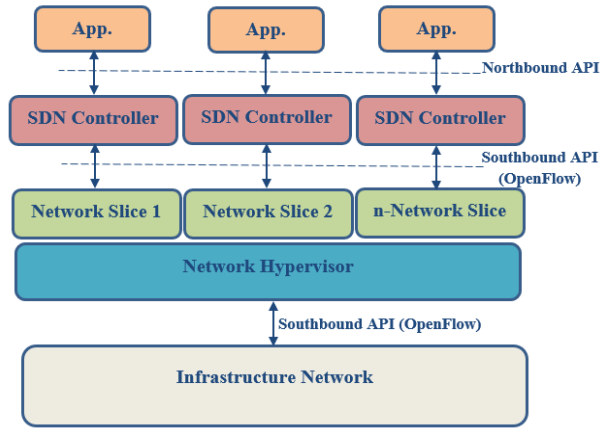


Fig. 6. Architecture of network slicing based on SDN.

As indicated in Figure 7, one of the most important hypervisors used to achieve this scenario is FlowVisor, which works as an SDN proxy intercepting messages between the data layer and the control layer [27].

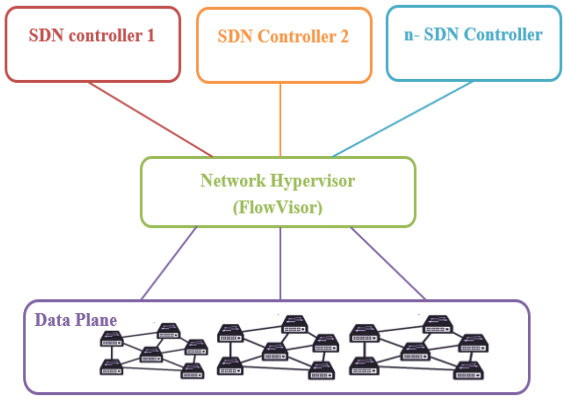


Fig. 7. Network slicing with SDN proxy.

FlowVisor is an infrastructure resource virtualization layer that enables the creation of multiple network slices, and each slice includes a dedicated SDN controller. With FlowVisor, network slices are conceptually separated from each other, and communication between the infrastructure and FlowVisor takes place through the OpenFlow protocol, as well as between FlowVisor and SDN controllers.

Therefore, securing the openflow channel is important in NS, as it represents the communication channel between the infrastructure layer and the virtualization layer, as well as between the virtualization layer and the control layer.

C. Experimental Results

To explain the proposed hybrid solution, we have implemented the hybrid keys to secure openflow channel, through incorporating the hybrid security into TLS protocol.

Figure 8 shows the initial messages to exchange hybrid key parameters, also TLS version that is used to openflow channel.

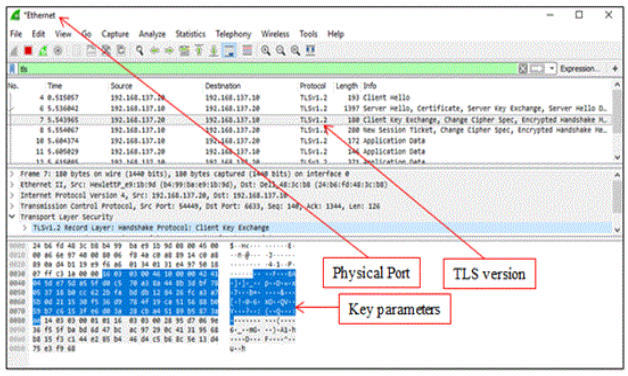


Fig. 8. Initial messages for key agreement.

Our experimental results indicated that the time required to implement the proposed QTLS protocol is acceptable to establish a secure connection compared to the standard TLS protocol. Figure 9 shows the time difference of implementation the standard TLS protocol and proposed QTLS.

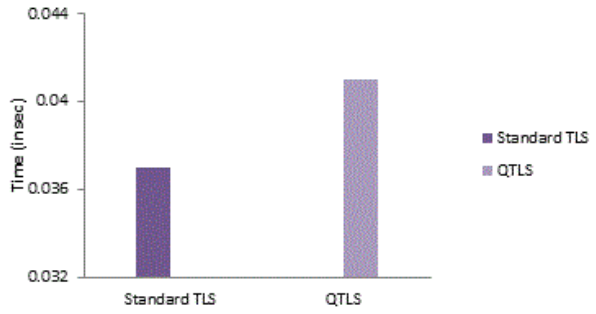


Fig. 9. Required time for implementing the standard TLS and QTLS protocols.

The reason for the time difference is due to the increased complexity within handshake in QTLS protocol, but this is at the expense of increasing security and authentication for key exchange between the two parties.

On the other hand, the randomness of the obtained hybrid key was measured by depended on six randomized NIST tests. The p-value was observed ≥ 0.01, and thus the binary sequence of hybrid key is more randomness, as explained in Table 1.

TABLE I
NIST (RANDOMNESS TESTS) RESULTS

	Key1	Key 2	Key 3	Key 4	Key 5
Frequency Test	0.725	0.508	0.536	0.965	0.595
Block Frequency (n = 128)	0.764	0.952	0.684	0.986	0.724
Runs	0.382	0.165	0.257	0.809	0.732
Longest Block Run	0.253	0.265	0.745	0.498	0.530
Approximate Entropy	1	1	1	1	1
Cumulative Sums	0.895	0.698	0.778	0.972	0.856

While we calculated the key space of the hybrid key and the results showed that the length of the key is large enough to make it impossible for the brute-force attacks to search for all possible keys using classic and quantitative computing. It has been shown [7] that a brute-force key search on a quantum computing cannot be faster than about $2^{n/2}$ when compared with about 2^{n} in the classical computing. Therefore, the hybrid key can be considered safe against quantum brute force attack, as shown in Table 2 the quantum and classical security levels for hybrid keys.

TABLE II
QUANTUM AND CLASSICAL SECURITY LEVELS FOR HYBRID KEYS

Keys	Key Length	Key Space	Security Level (in bits)	
			Classical Computing	Quantum Computing
Key 1	512	2^{512}	512	256
Key 2	1024	2^{1024}	1024	512
Key 3	2048	2^{2048}	12048	1024

D. Performance Evaluation and Comparison

As introduced in section IV.A, SDN-testbed was relied upon to implement our work. So in this section we review the

evaluation of network performance depending on throughput using iperf tests.

Analytically, throughput can be defined as the rate of maximum receiver bandwidth (Max BW) to round trip time (RTT) between hosts, where host A sends the number of packets to host B using the iperf tool.

Figure 10 show throughput according to different sizes of the segment size (64, 128, 512, 1024, and 1400) bytes.

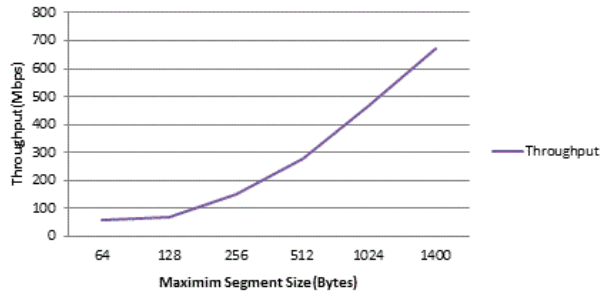


Fig. 10. Throughput of data transfer in the secure SDN-testbed.

Additionally, the results of SDN-testbed was compared with net-FPGA results after being converted to openflow switch [28], as explained in Figure 11. The result concluded that the result of our SDN-testbed is approximately similar to the performance of openflow switch based on net-FPGA hardware.

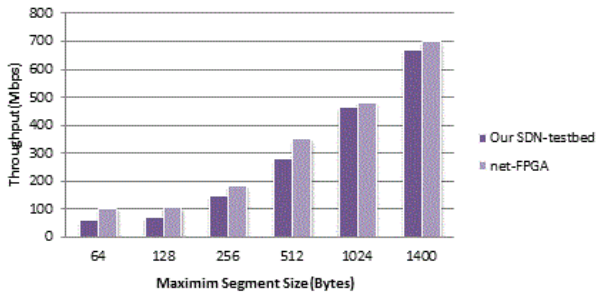


Fig. 11. Throughput in our SDN-testbed and net-FPGA.

Most of the previous research stresses the increase in the security of the openflow communication channel through the use of the standard TLS protocol, but there are security gaps in phase of exchanging the keys within TLS handshake protocol. Also, the emergence of quantum computers and the use of grover's and shor's algorithms have made it easy to break most classic cryptography protocols [29], and thus classic solutions have become inadequate for the purpose of securing communications.

In our work, the focus was on using the quantum keys distribution protocol with classical protocol to add a new layer of security based on quantum laws to increase authentication and security for openflow communication channel between the controller and network devices. Our experimental results showed that the hybrid key enhances the authentication and security of key exchange between the two parties in the QTLS protocol compared to the classic methods used for key exchange in the standard TLS protocol.

Our results were based on an analysis of the effect of classical and quantum computers on hybrid keys. The results showed that the hybrid key has the physical properties of the quantum in addition to the mathematical complexities, which make hybrid key difficult to break using the quantum or classical computer.

V. SECURITY ANALYSIS

In our work, the hybrid key was used to achieve authentication between the two parties at two levels in addition to using it to encrypt the channel. First Level: through the classical methods using exchange of certificates between the two parties while the second level: physical authentication through the quantum channel of QKD protocol and exploiting of the physical properties of quantum [12]. Therefore, It can be said that this study has achieved more secure authentication between the control plane and the data plane in SDN as well as between the virtualization layer and control layer in NS addition to secure the communication channel between the data layer and the virtualization layer. Therefore, it can be said that our proposal helps to avoid MitM attacks and achieve authentication mechanism on the openflow messages flowing through the channel.

On the other hand, in our work we provided a strong and reliable key for encrypting the communication channel, thus preventing data modification and providing a high level of confidentiality to openflow messages. In this way, the openflow communication channel was protected. As well, TLS protocol has been used with the hybrid key for securing communication and, as known, the TLS handshake protocol uses the nonce value and timestamp to prevent replay attacks.

VI. CONCLUSION

In recent times, SDN has been developed significantly as a result of high flexibility and programmability and SDN technology had seen as one of the most promising enablers of network development, which will play an essential role in the design of 5G networks through network slicing technology. Although it is promising in terms of cost reduction, it contains some security vulnerabilities that need solutions to address them.

In our work, we relied on enhancing the security of TLS protocol by using a hybrid key based on the mathematical complexities and the physical properties of the quantum. We have achieved sufficient security of openflow communication channel which is the basis of communication between layers of SDN networks as well as in NS. This security came about by fending off classical and quantum computer attacks by adding a new quantum security layer to TLS, as well as enhancing authentication between layers based on quantum properties.

The current work is effective to reduce the risk of attacks that threaten the security of the openflow communication channel.

REFERENCES

[1] Masoudi, R., & Ghaffari, A., 2016. Software defined networks: A survey. Journal of Network and computer Applications, 67, pp. 1–25. DOI: 10.1016/j.jnca.2016.03.016

[2] Alotaibi, D., 2021. Survey on network slice isolation in 5G networks: fundamental challenges. Procedia Computer Science, 182, pp. 38–45. DOI: 10.1016/j.procs.2021.02.006

[3] Zhang, S., 2019. An overview of network slicing for 5G. IEEE Wireless Communications, 26(3), pp. 111–117. DOI: 10.1109/MWC.2019.1800234

[4] Lee, S., Kim, J., Woo, S., Yoon, C., Scott-Hayward, S., Yegneswaran, V., Porras, P. and Shin, S., 2020. A comprehensive security assessment framework for software-defined networks. Computers & Security, 91, p. 101720. DOI: 10.1016/j.cose.2020.101720

[5] Nisar, K., Welch, I., Hassan, R., Sodhro, A.H. and Pirbhulal, S., 2020. A survey on the architecture, application, and security of software defined networking. Internet of Things, p. 100289. DOI: 10.1016/j.iot.2020.100289

[6] Yurekten, O. and Demirci, M., 2021. Citadel: Cyber threat intelligence assisted defense system for software-defined networks. Computer Networks, 191, p. 108013. DOI: 10.1016/j.comnet.2021.108013

[7] Kumar, M., 2021. Quantum Computing and Post Quantum Cryptography. International Journal of Innovative Research in Physics, 2(4), pp.37-51. DOI: 10.15864/ijiip.2405

[8] Sadkhan, S. B., Abbas, M. S., Mahdi, S. S., & Hussein, S. A., 2022, March. Software-Defined Network Security-Status, Challenges, and Future trends. In 2022 Muthanna International Conference on Engineering Science and Technology (MICEST), pp. 10–15. IEEE. DOI: 10.1109/MICEST54286.2022.9790219

[9] Portmann, C. and Renner, R., 2021. Security in quantum cryptography. arXiv preprint arXiv:2102.00021. DOI: 10.1103/RevModPhys.94.025008

[10] Bennett, C.H. and Brassard, G., 2020. Quantum cryptography: Public key distribution and coin tossing. arXiv preprint arXiv:2003.06557. DOI: 10.1016/j.tcs.2014.05.025

[11] Sun, S., & Huang, A., 2022. A review of security evaluation of practical quantum key distribution system. Entropy, 24(2), 260. DOI: 10.3390/e24020260

[12] Chen, Y., Gong, M., Xue, P., Yuan, H. and Zhang, C., 2021. Quantum deleting and cloning in a pseudo-unitary system. arXiv preprint arXiv:2103.15353. DOI: 10.1007/s11467-021-1063-z

[13] Czermann, M., Trócsányi, P., Kis, Z., Kovács, B., & Bacsárdi, L., 2021. Demonstrating BB84 quantum key distribution in the physical layer of an optical fiber based system. Infocommunications Journal, 13(3), pp. 45–55. DOI: 10.36244/ICJ.2021.3.5

[14] Abdullah, A.A. and Mahdi, S.S., 2019. Hybrid Quantum-Classical Key Distribution. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(12), pp. 4786–4791. DOI: 10.35940/ijitee.L3682.1081219

[15] Mahdi, S. S., & Abdullah, A. A., 2022, March. Improved Security of SDN based on Hybrid Quantum Key Distribution Protocol. In 2022 International Conference on Computer Science and Software Engineering (CSASE), pp. 36–40. IEEE. DOI: 10.1109/CSASE51777.2022.9759635

[16] Peng, Y., Wu, C., Zhao, B., Yu, W., Liu, B. and Qiao, S., 2016, November. QKDFlow: QKD based secure communication towards the openflow interface in SDN. In International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystem (pp. 410–415). Springer, Singapore. DOI: 10.1007/978-981-10-3969-0_45

[17] Lam, J., Lee, S.G., Lee, H.J. and Oktian, Y.E., 2016. Securing SDN southbound and data plane communication with IBC. Mobile Information Systems, 2016. DOI: 10.1155/2016/1708970

[18] Agborubere, B. and Sanchez-Velazquez, E., 2017, June. Openflow communications and tls security in software-defined networks. In 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (pp. 560–566). IEEE. DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.88

[19] Hugues-Salas, E., Ntavou, F., Ou, Y., Kennard, J.E., White, C., Gkounis, D., Nikolovgenis, K., Kanellos, G., Erven, C., Lord, A. and Nejabati, R., 2018, March. Experimental demonstration of DDoS mitigation over a quantum key distribution (QKD) network using software defined networking (SDN). In 2018 Optical Fiber Communications Conference and Exposition (OFC) (pp. 1–3). IEEE. DOI: 10.48550/arXiv.1802.05679

[20] Ni, J., Lin, X., & Shen, X. S., 2018. Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT. IEEE Journal on Selected Areas in Communications, 36(3), pp. 644–657. DOI: 10.1109/JSAC.2018.2815418

[21] Liu, J., Zhang, L., Sun, R., Du, X., & Guizani, M., 2018. Mutual heterogeneous signcryption schemes for 5G network slicings. IEEE Access, 6, p.p.7854-7863. DOI: 10.1109/ACCESS.2018.2797102

[22] Yigit, B., Gur, G., Tellenbach, B. and Alagoz, F., 2019. Secured communication channels in software-defined networks. IEEE Communications Magazine, 57(10), pp. 63–69. DOI: 10.1109/MCOM.001.1900060

[23] Porambage, P., Miche, Y., Kalliola, A., Liyanage, M., & Ylianttila, M., 2019. Secure keying scheme for network slicing in 5G architecture. In 2019 IEEE Conference on Standards for Communications and Networking (CSCN) (pp. 1-6). IEEE. DOI: 10.1109/CSCN.2019.8931330

[24] Thantharate, A., Paropkari, R., Walunj, V., Beard, C., & Kankariya, P., 2020. Secure5G: A deep learning framework towards a secure network slicing in 5G and beyond. In 2020 10th annual computing and communication workshop and conference (CCWC) (pp. 0852-0857). IEEE. DOI: 10.1109/CCWC47524.2020.9031158

[25] Gupta, V., Kaur, K. and Kaur, S., 2018. Developing small size low-cost software-defined networking switch using raspberry pi. In Next-generation networks (pp. 147–152). Springer, Singapore. DOI: 10.1007/978-981-10-6005-2_16

[26] Blenk, A., Basta, A., Reisslein, M., & Kellerer, W., 2015. Survey on network virtualization hypervisors for software defined networking. IEEE Communications Surveys & Tutorials, 18(1), pp. 655–685. DOI: 10.1109/COMST.2015.2489183

[27] Kurniawan, M. T., Moszardo, I., & Almaarif, A., 2022, June. Network Slicing On Software Defined Network Using Flowvisor and POX Controller To Flowspace Isolation Enforcement. In 2022 10th International Conference on Smart Grid (icSmartGrid), pp. 29–34. IEEE. DOI: 10.1109/icSmartGrid55722.2022.9848585

[28] Naous, J., Erickson, D., Covington, G. A., Appenzeller, G., & McKeown, N., 2008, November. Implementing an OpenFlow switch on the NetFPGA platform. In Proceedings of the 4th ACM/IEEE Symposium on Architectures for Networking and Communications Systems, pp. 1–9. DOI: 10.1145/1477942.1477944

[29] Alagic, G., Alperin-Sheriff, J., Apon, D., Cooper, D., Dang, Q., Kelsey, J., Liu, Y.K., Miller, C., Moody, D., Peralta, R. and Perlner, R., 2020. Status report on the second round of the NIST post-quantum cryptography standardization process. US Department of Commerce, NIST. DOI: 10.6028/NIST.IR.8309

Suadad S. Mahdi received her B.S. degree in 2016 from the College of Information Technology (IT), University of Babylon, Iraq, in 2020, her MS degree in Information Networks from the College of Information Technology (IT). Her main interests include Software-Defined Networks, Cryptography, Steganography and Quantum Cryptography.

Alharith A. Abdullah received his BS degree in Electrical Engineering from Military of Engineering College, Iraq, in 2000, his MS degree in Computer Engineering from University of Technology, Iraq, in 2005, and his PhD degree in Computer Engineering from Eastern Mediterranean University, Turkey, in 2015. His research interests include Security, Network Security, Cryptography, Quantum Computation and Quantum Cryptography.

A Primer on Software Defined Radios

Dimitrie C. Popescu, *Senior Member, IEEE* and Rolland Vida, *Senior Member, IEEE*

***Abstract*—The commercial success of cellular phone systems during the late 1980s and early 1990 years heralded the wireless revolution that became apparent at the turn of the 21st century and has led the modern society to a highly interconnected world where ubiquitous connectivity and mobility are enabled by powerful wireless terminals. Software defined radio (SDR) technology has played a major role in accelerating the pace at which wireless capabilities have advanced, in particular over the past 15 years, and SDRs are now at the core of modern wireless communication systems. In this paper we give an overview of SDRs that includes a discussion of drivers and technologies that have contributed to their continuous advancement, and presents the theory needed to understand the architecture and operation of current SDRs. We also review the choices for SDR platforms and the programming options that are currently available for SDR research, development, and teaching, and present case studies illustrating SDR use. Our hope is that the paper will be useful as a reference to wireless researchers and developers working in the industry or in academic settings on further advancing and refining the capabilities of wireless systems.**

***Index Terms*—Software defined radio, field programmable gate array, digital signal processing, wireless communication networks.**

I. INTRODUCTION

OVER the past three decades wireless communication systems have revolutionized the modern society, becoming essential components of our daily lives. Today's wireless devices provide much more than the mobile phone service enabled by the first generation of cellular phones available during the 1980s. They make extensive use of the Internet with capabilities that include accessing business and financial data, providing email, text messaging, and videoconference capabilities, enabling online shopping and entertainment with augmented reality features, assisting drivers with navigation and up-to-the-minute traffic information, and many more. Consumers preference of a wireless device and design has even become a personal statement about their status and social identity.

This unprecedented revolution in wireless communication systems occurred over multiple generations of wireless technologies that succeeded since the late 1980s and has been fueled by two main factors that have acted in synergy:

- Advances in hardware, starting from the clumsy, brick-like mobile phone terminals in the first generation to the sleek smartphones of the current generation that bring the Internet to our finger tips.
- Demand from consumers and society for applications that evolved from providing basic voice service using mobile phones and enabling wireless networking over short distances, and have advanced to supporting ubiquitous connectivity and edge computing through a vast heterogeneous infrastructure of interconnected wired and wireless networks.

A significant shift in the design paradigm of wireless systems occurred during the mid 1990s with the transition between second and third generations, when the SDR concept was formally introduced by visionary engineer and wireless pioneer Joseph Mitola [1], [2]. According to the Wireless Innovation Forum [3], a SDR is defined as "a radio in which some or all of the physical layer functions are software defined". We note that the physical layer of a communication system has been traditionally associated with the hardware, and any changes to physical layer functions such as modifying the modulation scheme or changing the frequency band associated with a particular system for example, would require hardware changes. Thus, in order to support multiple wireless standards on a conventional radio, all the corresponding hardware blocks would have to be built in, which would increase the manufacturing cost and limit flexibility to a predefined set of choices. By contrast, SDRs have a minimal set of hardware components and can change their operating parameters as needed through programming, providing a cost-effective alternative to multi-functional wireless devices.

In the three decades that have passed since the introduction of the SDR concept, SDRs have facilitated major advances in wireless communication systems through low-cost rapid prototyping, becoming the building blocks of modern communication systems. We note that, despite the fact that three decades of existence is expected to be a significant life time in the realm of modern electrical and electronic technologies, SDRs continue to thrive and are an ubiquitous presence in all aspects of research, development, and teaching of wireless communication systems and networks.

Motivated by the vitality of SDR technologies, in this paper we provide an overview of their salient aspects that can be used as a self-guided introduction to SDRs. We start by reviewing, in Section II, the drivers and enabling technologies that have shaped the SDR evolution over the past three decades, highlighting the current trends that maintain SDRs in the focus of the wireless communications research and development communities. We continue with a brief theoretical background, in Section III, that is indispensable to understanding SDR operation. This includes representation of bandpass signals in terms of in-phase and quadrature components along with heterodyning for frequency up- and down-conversions, and is

This work was completed while the first author was on a Fulbright US Scholar fellowship at the Budapest University of Technology and Economics.

D. C. Popescu is with the Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529, USA. (e-mail: dpopescu@odu.edu)

R. Vida is with the Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Magyar tudósok körútja 2, 1117 Budapest, Hungary. (e-mail: vida@tmit.bme.hu)

DOI: 10.36244/ICJ.2022.3.3

followed by presentation of SDR architectures in Section IV. SDR choices that are currently available on the commercial market along with programming alternatives are reviewed in Section V. Two case studies illustrating the use of SDR platforms in academic projects are also reviewed, in Section VI, before concluding the paper with final remarks in Section VII.

We hope that this SDR primer will become a useful reference to wireless researchers and developers working in the industry or in academic settings on future generations of wireless communication systems.

II. SDR Drivers and Enabling Technologies

Similar to cellular wireless systems, which have matured over multiple generations, SDRs have also seen the succession of multiple generations over which they have developed and have been refined. The timeline of SDR generations, however, does not align with that of the cellular wireless systems that have succeeded in the commercial/consumer market. Rather, SDR generations started in the late 1990s and are defined in terms of their increasing volume and presence in the overall wireless industry as outlined in [4] and illustrated in Fig. 1.

A. First SDR Generation

In the early days of SDR, during the late 1990s, the main driver was the defense industry with its efforts aimed at replacing existing radios used by the US military with a single one that was dubbed the Joint Tactical Radio System (JTRS) [5]. The idea behind the JTRS was that the new system could be programmed for multimode radio operation to eliminate the need for multiple radio units in a single military vehicle, and system upgrades would also be performed through software updates rather than through hardware changes. Besides the JTRS, other drivers of the initial development of SDRs include public-safety communications [6] along with spectrum monitoring and signal intelligence (SIGINT) [7], [8].

In terms of enabling technologies, the late 1990s and early 2000 years witnessed significant advances in integrated circuits (ICs) for radio frequency (RF) applications (also referred to as RFICs) [9] as well as in field programmable gate array (FPGA) technology [10]. These advances supported the needs of the defense-related SDR applications while also impacting the commercial market. Specifically, RFIC manufacturers were able to overcome important design challenges related to practical implementations of highly-integrated RF transceivers using CMOS technology, and RFICs advanced towards system-on-chip (SoC) solutions that combined complex RF analog and digital functionality, making possible the "ultimate transmission" [11]. At the same time, implementations of digital signal processing (DSP) algorithms using FPGA-based hardware also advanced to the point where they would be able to compete with application-specific integrated circuits (ASICs) and application-specific standard products (ASSPs) used in the current wireless communication systems [12], [13].

B. Subsequent SDR Generations

The advances made in RFIC and FPGA technologies led to the emergence of a commercial ecosystem of providers supporting SDR applications and prompted a second generation of SDRs in the early 2000 years. Equipment providers

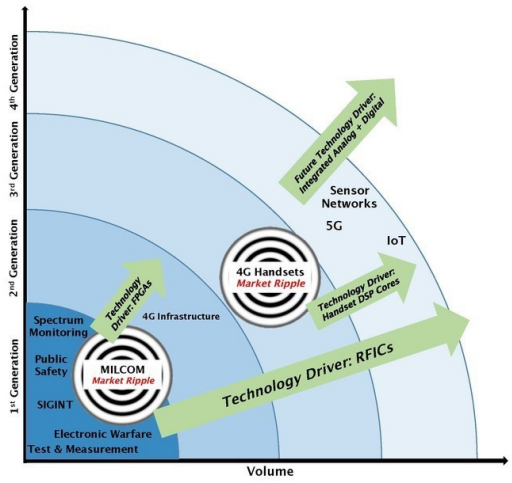


Fig. 1. The succession of SDR generations along with the drivers and enabling technologies for each generation [4].

came to the realization that SDR architectures are also beneficial to the development of cellular wireless systems [14] and SDRs made their way into the fourth generation (4G) equipment that was aligning with the long-term evolution (LTE) and LTE-advanced (LTE-A) standards [15], [16]. As a result, the LTE base station infrastructure was developed using SDR RFIC and FPGA technology, and new concepts such as software defined networking (SDN) and network function virtualization (NFV) were introduced as an approach to decouple the various network functions and services from the underlying hardware components of the network in order to support legacy services as well as future evolutions [17]. In addition, a software communications architecture (SCA) was established as a distributed systems architecture that allows the distinct components of a SDR application to run on different processors, which communicate with each other based on the Common Object Request Broker Architecture (CORBA) middleware [18].

Further advances in low-power high-performance ICs prompted the move of SDR technology to the handset segment of the 4G LTE networks starting in the early 2010 years. Specifically, low-power RFICs [19] in conjunction with high-performance FPGAs optimized to function as DSP cores [20] have started to be used in consumer handsets, significantly increasing the volume of SDRs on the commercial market. This marked the third generation of SDRs that also resulted in the SDR technology becoming a de facto industry standard for radios.

C. Future Trends

Currently, emerging systems such as the fifth generation (5G) of cellular systems and the Internet of Things (IoT), provide impetus for further development of SDR technology that will include advances on both sides of SDR platforms, hardware and software.

In terms of technology drivers, advances are expected to occur on the hardware side of SDRs that will bring the analog and digital sides closer together [21], by combining them in

a single monolithic chip that will result in integration of the FPGAs or of the ASICs with the analog-to-digital converters (ADC) and digital-to-analog converters (DAC) [22], which will likely lower the overall size and cost of the SDR platforms, making them even more affordable and widespread in practice. At the same time, on the software side of the SDRs, the programming tools used by developers and researchers will evolve to enable the implementation of more complex tasks and novel DSP algorithms on increasingly more powerful FPGAs and ASICs.

The wireless industry will continue to rely on SDRs in the development of 5G systems, using them for various purposes that include practical experimentation and prototyping [23], as well as for enabling reconfigurable wireless networks with efficient spectrum utilization where the SDRs provide the programmable RF front-end needed for adjusting modulation schemes for operation in different frequency bands [24]. In this direction we note the performance evaluation of the non-orthogonal multiple access (NOMA) approach using SDR platforms in [25] and the 5G radio prototypes that are based on SDRs [26], [27].

5G systems are also expected to support the IoT with its specific requirements implied by the need to interconnect a multitude of sensors operating on strict energy and latency constraints [28], and SDRs will also be beneficial to the development of IoT networks by enabling rapid prototyping and experimentation. In this direction we note the SDR implementations of time-sensitive IoT networks in [29] and of RF identification (RFID) readers in [30], [31]. In addition, SDR implementations of receivers for the proprietary long range low power (LoRa) modulation technique [32] have recently been presented [33], [34].

Other emerging applications that have started to influence SDR development and evolution in recent years include satellite communications, where the SDR cost and versatility makes them attractive for implementing reconfigurable radio links that can deliver high data rate with low power consumption in small satellite systems [35]. In addition, satellite communications are also envisioned to support the Internet of Remote Things (IoRT), where sensors or other smart devices are located in remote areas or they are dispersed over a wide geographical area such that they are inaccessible to terrestrial networks [36] and SDR-based gateways are used to connect them to a satellite network [37].

III. THEORETICAL BACKGROUND

Like any other type of radio system, a SDR is used to transmit and receive bandpass signals that carry information. In order to have a complete picture of SDR operation, a good understanding of the canonical representation of bandpass signals [38, Appendix 2] is a necessary prerequisite. This need is also emphasized in references that discuss the more general concept of "software defined electronics" [39], [40], which includes SDRs as well as other types of modern measurement systems that rely on converting bandpass RF signals to lowpass baseband equivalent ones and then using software approaches for further processing.

Bandpass signals are formally defined as signals with spectrum concentrated in a band of frequencies that is centered at some frequency f_c, usually much larger than the bandpass signal bandwidth, which is denoted by $2W$ and extends over the frequency interval $[f_c - W, f_c + W]$. They are frequently encountered in communication systems and are obtained from a baseband information-bearing signal by applying a specific modulation scheme to a sinusoidal carrier signal with frequency f_c. To illustrate bandpass signals with a practical example, Fig. 2 shows the instantaneous spectrum along with the waterfall plot corresponding to a RF scan of the FM broadcast band displaying three active FM stations. We note that, the three distinct stations that are active display different patterns of frequency use in time as seen in the waterfall plot part of Fig. 2, which correspond to the distinct music and/or talk shows broadcast at scan time on the three stations. Their instantaneous spectra, however, look similar as they correspond to bandpass signals obtained by applying the same type of modulation (frequency modulation – FM) to baseband signals that contain similar information (music and speech signals).

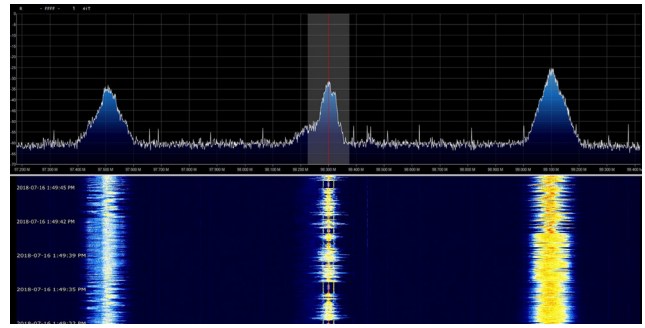


Fig. 2. Example of bandpass signal spectra collected in a RF scan of the FM broadcast bands [41]. The RF spectrum monitoring application displays only the spectral images that correspond to positive frequencies.

A. Pre-Envelope and Complex Lowpass Equivalent Signals

We consider an arbitrary bandpass signal $s(t)$ with a generic amplitude spectrum $|S(f)|$ shown in Fig. 3(a)[1], and we note that the fist step in obtaining the canonical representation of bandpass signals is to construct the *pre-envelope signal*, which is a complex-valued signal whose real part consists of the original bandpass signal $s(t)$, while its imaginary part consists of the Hilbert transform $\hat{s}(t)$ of the bandpass signal $s(t)$:

$$s_+(t) = s(t) + j\hat{s}(t) \tag{1}$$

We note that the Hilbert transform performs a phase shift of $\pm\pi/2$ on all components of $s(t)$ and may be obtained by passing $s(t)$ through a linear filter[2] with impulse response $h(t) = 1/(\pi t)$ and transfer function $H(f) = -j\text{sgn}(f)$, where $\text{sgn}(\cdot)$ denotes the signum function [38, Appendix 2]. We also note some properties of the Hilbert transform that are relevant to the canonical representation of bandpass signals:

- A signal $s(t)$ and its Hilbert transform $\hat{s}(t)$ are orthogonal, that is

$$\int_{-\infty}^{\infty} s(t)\hat{s}(t)dt = 0 \tag{2}$$

[1]The bandpass signal $s(t)$ is assumed to be real-valued, hence the symmetry of its amplitude spectrum $|S(f)|$ with respect to the vertical axis that can be noticed in Fig. 3(a).

[2]This linear filter is referred to as a *Hilbert transformer* [38, Appendix 2].

The orthogonality property of the Hilbert transform goes along with the intuition that the real and imaginary parts of a complex-valued quantity are real-valued quantities corresponding to two orthogonal dimensions that are represented by the horizontal and vertical axes of the Cartesian representation of the complex plane, and support the construction of the pre-envelope signal (1) having

$$s(t) = \mathfrak{Re}\{s_+(t)\} \quad \text{and} \quad \hat{s}(t) = \mathfrak{Im}\{s_+(t)\}. \tag{3}$$

- A signal $s(t)$ and its Hilbert transform $\hat{s}(t)$ have the same amplitude spectrum

$$|S(f)| = |\hat{S}(f)|, \tag{4}$$

where $|S(f)|$ and $|\hat{S}(f)|$ denote the Fourier transforms of $s(t)$ and $\hat{s}(t)$, respectively.

Using simple algebra one can easily show that the Fourier transform $S_+(f)$ of the pre-envelope signal $s_+(t)$ can be expressed in terms of the Fourier transform $S(f)$ of the bandpass signal $s(t)$ as

$$S_+(f) = \begin{cases} 0 & \text{for } f < 0 \\ S(0) & \text{for } f = 0 \\ 2S(f) & \text{for } f > 0, \end{cases} \tag{5}$$

which shows that $S_+(f)$ has no components with negative frequencies. Thus, for the bandpass signal $s(t)$ with generic amplitude spectrum shown in Fig. 3(a), the amplitude spectrum $|S_+(f)|$ of its corresponding pre-envelope signal $s_+(t)$ looks like the one shown in Fig. 3(b) and can be obtained through a shift in frequency by f_c of the amplitude spectrum $|\tilde{S}(f)|$ shown in Fig. 3(c) that corresponds to signal $\tilde{s}(t)$.

The signal $\tilde{s}(t)$ with amplitude spectrum shown in Fig. 3(c) is referred to as the *complex lowpass equivalent signal* of the bandpass signal $s(t)$, and the frequency shifting relationship between the $|S_+(f)|$ and $|\tilde{S}(f)|$ amplitude spectra,

$$S_+(f) = \tilde{S}(f - f_c) \tag{6}$$

translates into multiplication by a complex exponential of the complex lowpass equivalent signal in time domain, that is

$$s_+(t) = \tilde{s}(t)e^{j2\pi f_c t}. \tag{7}$$

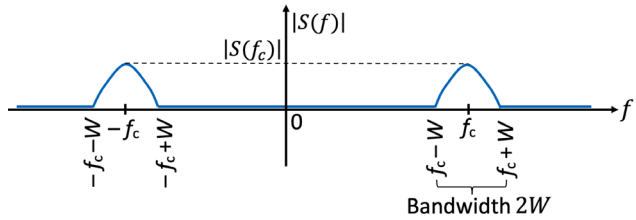
Noting that, by construction (3), the original bandpass signal $s(t)$ corresponds to the real part of the pre-envelope signal, we can now write the relationship between $s(t)$ and $\tilde{s}(t)$ as

$$s(t) = \mathfrak{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\}. \tag{8}$$

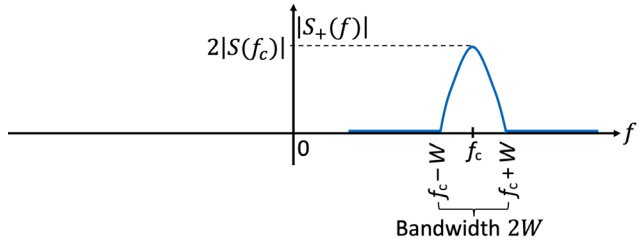
Expression (8) highlights the two components of the canonical representation of bandpass signals:

- The information content of bandpass signal $s(t)$, which is implied by the spectrum of its complex lowpass equivalent signal $\tilde{s}(t)$ with bandwidth $2W$, and
- The frequency band where the bandpass signal occurs, which is centered at f_c, the frequency of the complex exponential term.

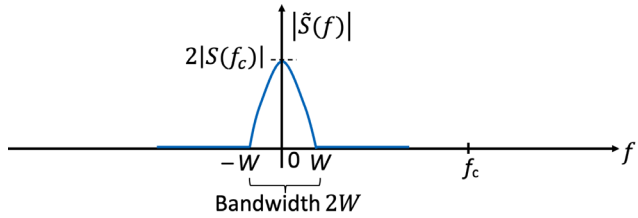
Thus, the canonical representation of bandpass signals enables their analysis in terms of complex lowpass equivalent signals and it is independent of the center frequency at which the bandpass signals occur. From a SDR perspective, the implication is that the transmitter can focus on implementing a modulation scheme for information transmission without considering the band of frequencies in which the modulated signal should be transmitted, while the receiver can extract the information contained in the bandpass signal by baseband processing of the complex lowpass equivalent signal.



(a) Amplitude spectrum for bandpass signal $s(t)$.



(b) Amplitude spectrum for pre-envelope signal $s_+(t)$.



(c) Amplitude spectrum for complex lowpass equivalent signal $\tilde{s}(t)$.

Fig. 3. Amplitude spectra for signals used in the canonical representation of bandpass signals.

B. The In-Phase and Quadrature Signal Components

The downside of the canonical representation of bandpass signals based on the complex lowpass equivalent signal is the fact that, due to its complex-valued nature, its characteristics cannot be directly visualized using measurement equipment such as an oscilloscope or spectrum analyzer. Nevertheless, the complex lowpass equivalent signal can be used to provide an alternative representation in terms of the two real-valued signals that make up its real and imaginary parts, $s_I(t)$ and $s_Q(t)$, respectively, which are referred to as the in-phase (I) and quadrature (Q) components of the bandpass signal. Thus, for bandpass signal $s(t)$ with complex lowpass equivalent signal $\tilde{s}(t)$ we have that

$$s_I(t) = \mathfrak{Re}\{\tilde{s}(t)\} \quad \text{and} \quad s_Q(t) = \mathfrak{Im}\{\tilde{s}(t)\}, \tag{9}$$

such that equation (8) can be rewritten as

$$s(t) = \mathfrak{Re}\{[s_I(t) + js_Q(t)]e^{j2\pi f_c t}\}. \tag{10}$$

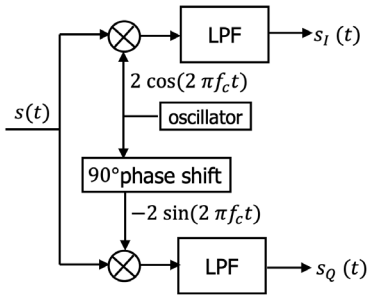


Fig. 4. Obtaining the I and Q components of a bandpass signal.

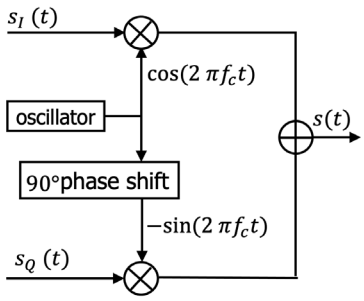


Fig. 5. Synthesizing a bandpass signal from its I and Q components.

Using Euler's formula $e^{j2\pi f_c t} = \cos(2\pi f_c t) + j\sin(2\pi f_c t)$ in (10) we obtain the equivalent expression of $s(t)$ in terms of the I and Q components as

$$s(t) = s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t), \qquad (11)$$

which, similar to (8), provides the I/Q signals as an alternative way of characterizing the information content of bandpass signal $s(t)$, in terms of real-valued signals $s_I(t)$ and $s_Q(t)$, both with lowpass spectrum and bandwidth $2W$ as implied by the spectrum of $\tilde{s}(t)$.

Given the bandpass signal $s(t)$, its I and Q components can be obtained by multiplying it with $\cos(2\pi f_c t)$ and $-\sin(2\pi f_c t)$ respectively, followed by lowpass filtering (LPF), as shown in Fig 4, where the bandwidth of the LPF used is the same for both $s_I(t)$ and $s_Q(t)$ and is equal to the bandwidth W of the complex lowpass equivalent signal $\tilde{s}(t)$.

Alternatively, when the I and Q components of the bandpass signal are available, the bandpass signal $s(t)$ can be synthesized by directly implementing (11) as shown in Fig. 5.

C. Heterodyning and Frequency Down/Up-Conversion

Heterodyning, also referred to as frequency changing or mixing [38, Section 2.4], consists of multiplying a bandpass signal $s_1(t)$ with center frequency f_{c1} with a sinusoidal signal produced by a local oscillator with frequency f_{LO} followed by an appropriate bandpass filtering operation to produce a new bandpass signal $s_2(t)$ with a different center frequency $f_{c2} = f_{c1} \pm f_{LO}$. When $f_{c2} < f_{c1}$ the operation is referred to as *frequency-down conversion*, and when $f_{c2} > f_{c1}$ the operation is referred to as *frequency-up conversion*.

A major application of heterodyning is in the superheterodyne receiver [38, Section 2.9], which has been used for decades in the reception of radio signals by converting the received RF signal to an intermediate frequency (IF) where it would be further filtered and amplified before being processed by the demodulator to extract the information. The same approach that consists of processing at an IF, which is used in the superheterodyne receiver, is also used in SDRs [42] as well as in other applications such as the emerging 5G systems [43].

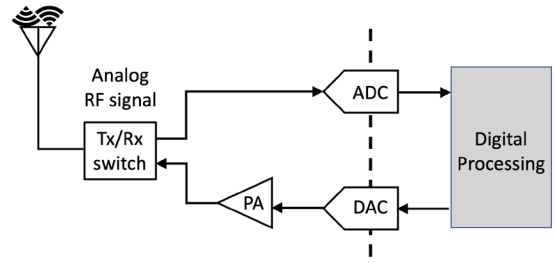


Fig. 6. Ideal architecture of a SDR. The ADC and DAC are performed on the RF signal, and a power amplifier (PA) is used on the transmit side to ensure desired RF transmit power level.

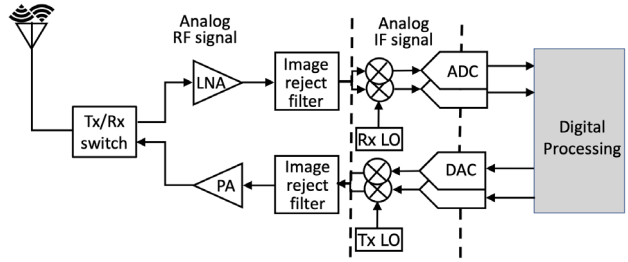


Fig. 7. The architecture of existing SDRs. The RF signal is shifted to IF where ADC and DAC are performed. A low noise amplifier (LNA) is used on the receive side prior to IF, with a PA on the transmit side.

IV. THE SDR ARCHITECTURE

Ideally, a SDR platform should perform all processing digitally, with the conversions from analog-to-digital and digital-to-analog occurring directly on the RF signal at the antenna, as shown in Fig. 6. This architecture, which corresponds to the one envisioned by Mitola [2], is applicable currently to lower frequencies, mostly in the high-frequency (HF) and very high frequency (VHF) bands, due to limitations of existing ADC and DAC converter technology. We note that, to support operation over a wide RF range, from HF (tens of MHz) to super high frequencies (SHF) up to 6 GHz, the ADC and DAC used must have extremely high resolutions with a wide dynamic range. This is a critical requirement that was acknowledged in the microelectronics and IC community from the early days of SDRs [44].

A. The IF Stage

During the late 1990s and early 2000 years, it was realized that capabilities of integrated ADCs and DACs [45] were increasing at a slower pace than those of other types of ICs, which were following Moore's law [46]. Thus, alternative architectures for SDRs had to be pursued. The solution was found in the form of the SDR architecture shown in Fig 7, which is present in current SDRs and uses an IF processing stage where the ADC and DAC take place [42].

The IF stage bridges the RF front end of the SDR with its digital processing core where information is extracted from received signals or embedded in signals synthesized for transmission:

- On the receiver side, the IF stage translates the analog RF signal to the IF and enables a subsampling (or sampling translation) approach for ADC [47], which takes advantage of the fact that a conventional ADC can digitize an analog signal with a compact spectrum, as is the case with bandpass signals, using an undersampling approach relative to the RF or IF frequencies, but oversampling with respect to the information bandwidth of the bandpass signal [48], [49]. The RF front end of the SDR may include on the receiver side a LNA to strengthen weak signals without significantly impacting the signal-to-noise ratio (SNR) [50].
- On the transmitter side, the I and Q components of the modulated signal are digitally synthesized and translated to IF, followed by DAC and analog frequency translation to RF. The RF front end of the SDR includes on the transmitter side also a PA, which is a critical component as it impacts the power consumption and overall cost of the SDR [51].

We note that in current implementations of SDR platforms, the two analog stages of the SDR (the RF front end and the IF) are usually integrated on the same chip. For example, this is the case with Ettus Research bus series SDR platforms, which use AD 936x agile transceiver chips [52][3], as well as with the RTL-SDR receive only SDR, which uses the R820T tuner [53].

B. Digital Processing

On the digital processing side of a SDR platform one can also distinguish two distinct stages, the digital front end and the baseband processing stage [54].

The digital front end performs two functions [55]:

- Sample rate conversion, which adapts the sampling rate corresponding to the IF stage and the sampling rate at which the digital baseband processing is accomplished in the subsequent signal processing stage.
- Channelization, which includes channel filtering to extract specific frequency bands and conversion between the digital IF and baseband.

The baseband processing stage is where the actual operations related to the communication signal synthesis/analysis takes place and covers functions that include physical layer processing such as implementing modulation/demodulation and error correction encoding/decoding as well as MAC layer functions that connect the physical layer with the upper protocol layers.

In existing implementations of SDR platforms, the digital front end is implemented on an FPGA that can be co-located on the same board as the RF front end and IF chip or on a different board, while baseband processing is accomplished on a general purpose processor (GPP), which, in most cases, is a host computer programmed to run specific applications handling the digital stream of I/Q data. In this case, the FPGA includes also the communication interface between the digital front end and the host computer, which can be over Universal Serial Bus (USB), Ethernet, or PCIe [56].

Baseband processing can also be implemented on the same FPGA as the digital front end if the FPGA fabric has sufficient resources available, which is the case with high-end FPGAs and Systems-on-a-Chip (SoCs) that integrate powerful FPGAs with ARM processing cores on the same IC enabling standalone SDR platforms that can be deployed in the field [54].

V. SDR CHOICES AND PROGRAMMING

A wide range of SDR platforms are currently available on the commercial market, and providing a comprehensive listing of all SDR choices is beyond the scope of the paper. Rather, we would like to highlight several SDR platforms that have attracted the attention of a wider audience and have been used for wireless systems research, development, and teaching in industry and academic settings.

A. The Universal Software Radio Peripheral – USRP

The USRP is among the most widely used SDR platforms for wireless research and teaching [57], being available in many flavors [58]:

- At the low end, the USRP family has the bus series, which provides a fully integrated, single board SDR platform with continuous frequency coverage from 70 MHz to 6 GHz and up to 56 MHz of real-time bandwidth.
- At the high end of the USRP spectrum, the X series offers a high-performance scalable architecture that includes large user-programmable FPGAs and the RF front end covering the range from DC to 6 GHz with up to 120 MHz of baseband bandwidth.
- The top member of the USRP family, the X410, features a Zynq Ultrascale RFSoC that includes a quad-core ARM Cortex-A53 processor for standalone applications and is designed for frequencies from 1 MHz to 7.2 GHz with 4 independent transmit/receive channels and a two-stage superheterodyne architecture, being capable of supporting up 400 MHz of instantaneous bandwidth on each channel.

Over the past decade the USRP SDRs have become a leading choice for teaching fundamental concepts in communication systems and for hands-on experimentation with wireless communications, and many references are available in the published literature [59], [60], [61], [62].

B. The RTL-SDR Receiver

Another popular SDR choice is the RTL-SDR, which is a receive only platform, with the name acronym coming from the use of the RealTek RTL2832U chip for its digital front-end. Different versions of the RTL-SDR are available, that are distinguished by the different tuner chips used to receive the RF signal, which include [53]:

- The Rafael Micro R820T covers the frequency range from 24 MHz to 1.766 GHz and uses an IF processing stage to provide a down-converted IF signal with a bandwidth of about 6 MHz to the RTL2832U, which extracts the digital I/Q data.

[3]The AD 9364 implements a single transceiver and is used in the USRP B200. The AD9361 provides two independent transceivers and is used in the USRP B210.

- The Elonics E4000 operates from 52 MHz to 2.2 GHz, with a gap between 1.1 GHz and 1.25 GHz, and has no IF stage, converting the analog RF signal to a baseband one with a roughly 10 MHz bandwidth and feeding the analog I/Q signals to the RTL2832U, which samples them to extract the I/Q data.

The I/Q data at the output of the RTL2832, which has a bandwidth of about 2.8 MHz and is encoded on 8 bits, is provided over USB to the host computer for baseband processing.

The RTL-SDR receiver is very affordable, with kits that include the RTL-SDR USB dongle, antennas and cables, available online for very low prices. Despite its lower capabilities in terms of frequency range or bandwidth when compared to the USRP, the RTL-SDR is a main choice for radio enthusiasts, with numerous projects using it featured on the internet [63]. In addition, reference book [53], which can be used for hands-on teaching SDR concepts, has also contributed to the popularity of the RTL-SDR receiver.

C. Other SDR Choices

The ADALM-PLUTO is a SDR platform that aims academic teaching and is marketed as "*an active learning module*" that "*helps introduce electrical engineering students to the fundamentals of SDR, RF, and wireless communications*" [64]. Its RF front end features an AD9363 highly integrated RF agile transceiver with the digital front end using a Zynq FPGA, operating over the frequency range from 325 MHz to 3.8 GHz with up to 20 MHz of real-time bandwidth and communicating with the host computer over USB. The ADALM-PLUTO is supported in MATLAB and Simulink, and is a good candidate for integrating it in the electrical engineering curriculum to support teaching a wide range of concepts related to RF and wireless systems, digital communications and signal processing, or embedded systems [65], [66].

Two other SDRs have also been mentioned alongside the USRP in a recent study of SDR platforms that meet minimum specifications for existing wireless technologies [56], the HackRF One [67] and the Lime SDR [68]. They have also been used in academic projects [69], [70] and are also popular with radio enthusiasts, with various projects using them also featured online [63]. Their main characteristics are:

- The HackRF operates over the frequency range from 1 MHz to 6 GHz with an instantaneous bandwidth of 20 MHz, communicating with the host computer over USB 2.0. It features a MAX2837 chip for the RF front end, which has no IF and converts the RF signals to baseband, followed by the MAX5864 ADC/DAC and the LPC4300 series ARM Cortex-M4 microcontroller for its digital front end.
- The Lime SDR operates over the frequency range from 100 kHz to 3.8 GHz with an instantaneous bandwidth of 160 MHz, communicating with the host computer over USB 3.0. The RF front end uses the LMS7002M field programmable RFIC dual transceiver, which supports 2×2 MIMO configurations and has on chip integrated 12-bit ADC and DAC to provide the digital I/Q signal data to an Alterra Cyclone IV FPGA.

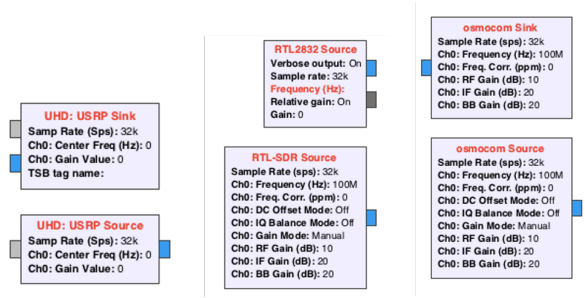


Fig. 8. GNU Radio blocks required for programming SDR platforms: (a) Source and sink blocks for USRP SDR; (b) Different versions of source blocks for RTL-SDR; (c) Osmocom source and sink blocks that can be used with various SDRs.

Finally, the Myriad RF SDR board is also worth mentioning [71]. The original Myriad–RF 1 is a multi-band, multi-standard RF module from Lime Microsystems that is based on their LMS6002D transceiver IC, featuring one RF broadband output and one RF broadband input with digital baseband interface. The Myriad-RF 1 contained everything needed for it to be connected to baseband chipsets, FPGAs, or to run in standalone mode. Currently, the Myriad-RF 1 design has been adapted for use with the Novena open hardware computing platform, which can be configured for embedded applications [72].

D. SDR Programming

Software development toolkits available to program SDR platforms include GNU Radio [73], MATLAB and Simulink [74], and LabVIEW [75]. They provide graphical interfaces in which blocks performing specific signal processing functions are interconnected to implement various physical and MAC layer functions on SDRs and to run standalone applications. We note that GNU Radio is open source and free to use, while MATLAB and Simulink as well as LabVIEW require a valid license to be able to use them.

Behind every block available in GNU Radio there is a Python script supporting it. Due to the open source nature of GNU Radio, code can be modified by the user as needed by adding out-of-tree (OOT) modules containing new functionalities and blocks [76], thus effectively leveraging the power of open-source SDR community [73]. We note that GNU Radio was originally designed for use with the open-source Linux operating system and the Ubuntu distribution of Linux continues to be preferred for developing applications with GNU Radio by SDR developers. Nevertheless, installation options for running GNU Radio under Windows and Mac operating systems are also available, albeit taking advantage of the open-source features of GNU Radio such as adding OOT modules may not be as friendly under these operating systems as under Ubuntu. We also note that, while GNU Radio can be used without any hardware as a simulation and development environment, its power lies in the ability to simulate complete transmit/receive chains that include RF, analog, and other relevant impairments encountered in practical systems and implementations. Thus, using GNU Radio with specific SDR platforms requires that the manufacturers provide support for

GNU Radio to ensure that blocks corresponding to their specific SDR platforms such as the ones illustrated in Fig. 8 are available for use:

- Sink blocks represent transmitters and correspond to the RF front end of the SDR platform that synthesizes the RF signal. The input to a SDR sink is in general a complex variable whose real and imaginary parts, respectively, are sampled versions of the I and Q components of the RF signal that is transmitted by the SDR, with the sampling rate specified as one of the parameters of the sink block.
- Source blocks represent receivers and correspond to the RF front end of the SDR platform that acquires the RF signal. The output of a SDR source is in general a complex variable whose real and imaginary parts, respectively, are sampled versions of the I and Q components of the RF signal that is acquired by the SDR, with the sampling rate specified as one of the parameters of the source block.

Programming SDR platforms using MATLAB and Simulink requires also the Communications Toolbox, which needs to be included with the MATLAB and Simulink license to be available for use. In addition, hardware support packages specific to the SDR platform that needs to be programmed should be installed. These provide Simulink blocks similar to the source and sink blocks in GNU Radio, which communicate with external SDR devices to process live radio signals captured over the air. Currently, the USRP, the RTL-SDR, and the ADALM-PLUTO are supported with MATLAB, but the HackRF One and the Lime SDR are not [74]. Since MATLAB is not open source, the blocks available for SDR programming cannot be modified by the users. However, free open-source MATLAB and Simulink code published by users is available on the MATLAB Central File Exchange [77].

Programming using LabVIEW is currently limited to the NI and Ettus USRP SDRs, and other SDR platforms are not officially supported. However, some examples of using LabVIEW with RTL-SDR can be found by searching the NI Community website [78].

VI. CASE STUDIES

We illustrate the use of SDR platforms with two case studies that have been completed in recent years in academic projects outside of a formal course on wireless communications:

- In the first project a Lime SDR platform is used as receiver to collect RF measurements for an empirical characterization of man-made noise in the 900 MHz frequency band [70]. This study demonstrates one of the many applications of SDR platforms, to replace conventional equipment used for performing RF measurements, which leads to lighter and more portable system and is beneficial for reducing overall system costs.
- In the second project an experimental study of the RF transmit power for a USRP B200 is presented [79]. The study is important since it highlights the need to test SDR platforms prior to using them in practical implementations, to confirm the RF power level at which they are programmed to transmit, and to understand dependence of RF power on frequency and other parameters, such as transmit gain(s) for example, which may be available when programming the SDRs.

Fig. 9. Full size van converted into a mobile platform to perform noise measurements in the early 1990s [70].

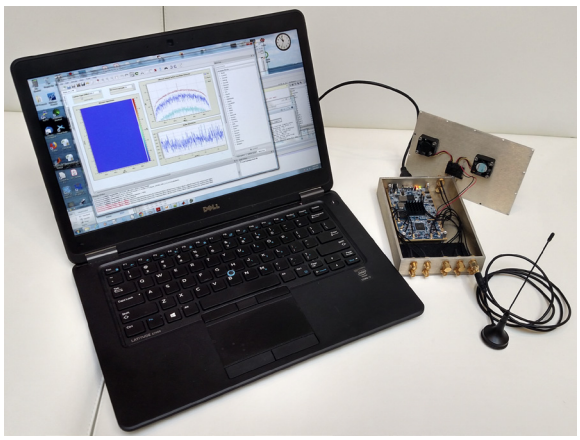


Fig. 10. The mobile platform used to perform noise measurements in [70].

A. Using SDR Platforms for Empirical Noise Characterization

The items needed to perform an empirical characterization of impulsive noise include a radio receiver, a spectrum analyzer, a logarithmic detector, a digital oscilloscope, and a computer. We note that performing an empirical noise characterization study during the early 1990s required access to a well-equipped lab dedicated to communication systems where all the items, which had significant price tags at the time, had to be available. Furthermore, in order to incorporate this equipment into a mobile platform a full size van had to be converted into a measurement vehicle as shown in Fig. 9.

By contrast, the setup used for taking measurements of impulsive noise these days can be accomplished using inexpensive components that can be acquired with the limited budget of an undergraduate research project, consisting of a SDR platform such as the Lime SDR, along with an average laptop computer as seen in Fig. 10. The SDR platform is configured as receiver providing access to the I/Q noise data, while all of the other items (the logarithmic detector, the spectrum analyzer, and the digital oscilloscope) are integrated in the signal processing software running on the laptop.

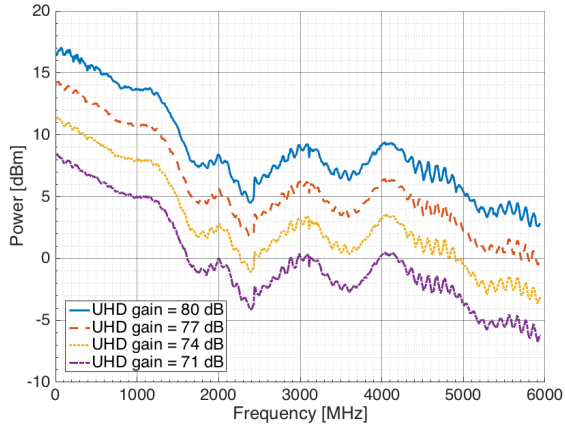
The specific application presented in [70] focuses on a narrowband system with RF bandwidth of 250 Hz operating in the 900 MHz band, but the setup can be easily adapted to other applications due to the versatility of the Lime SDR platform, which operates over a wide range of frequencies, covering the HF, VHF and UHF bands. Thus, the measurement setup in [70] requires minimal software changes to be applied to impulsive noise measurements and characterization in other scenarios, such as complementing the numerical simulation results in [80] with actual RF measurements of the noise radiated by a microwave oven over a narrowband channel with 300 kHz bandwidth at 2 GHz, similar to the one in [81]. Alternatively, the same system in [70] can be adapted to characterize wideband UHF digital TV channels with a bandwidth of 10 MHz in the 700 MHz band [82].

The Lime SDR, or other similar SDR platform, can also be used in transmit mode to synthesize impulsive noise signals using computer generated I and Q components as outlined in [83]. Accomplishing this requires in essence developing the software for generating the I and Q components of the artificial impulsive noise along with programming the SDR platform to transmit it over the frequency band of interest, and can be useful for testing purposes in a lab setup.

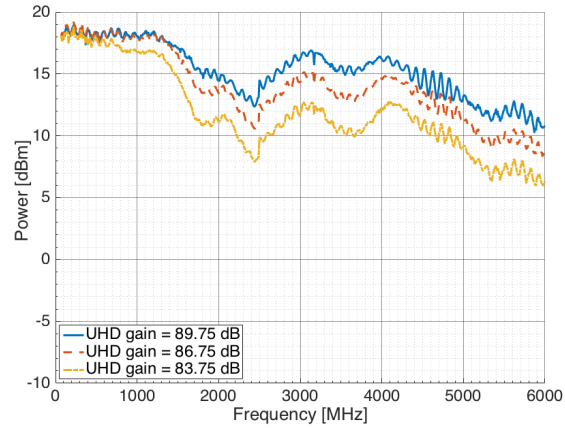
The noise measurement approach presented in [70] allows RF site-surveyors to accumulate noise data and to identify best and worst-case noise scenarios by using a lightweight and highly portable battery-powered measurement setup that includes a SDR platform and a laptop computer. The setup can be easily adapted to perform measurements of broadband or narrowband electromagnetic interference emissions, which are usually performed by experts at nationally recognized testing laboratories. Such measurements can be both expensive and time consuming, since they require specialized personnel and test fixtures, as well as hours of calibration and measurements to produce highly accurate certified results. However, when stringent accuracy and certification are not absolutely necessary, as may be the case with consumer-type applications, taking advantage of the versatility of SDR platforms, to establish dedicated systems for electromagnetic compatibility testing can significantly reduce cost, while still providing useful information for system design.

B. Transmit Power Variation for USRP B200 SDR

The study of transmit RF power in [79] is motivated by the fact that the USRP SDRs are not calibrated devices that can be used for measurement and/or testing, and USRP data sheets are vague when it comes to the specification of the RF transmit power level. A USRP B200 is considered in



(a) Lower UHD gain values.



(b) Higher UHD gain values.

Fig. 11. Measured output RF power for USRP B200 SDR [79].

[79], for which the RF specifications mention only that its RF transmit power is above 10 dBm [52]. Furthermore, the transmit power of the USRP B200 is programmable through the transmit gain parameter in the USRP Hardware Driver (UHD), but exact specification of RF power as a function of the UHD gain is elusive. Nevertheless, precise knowledge of the RF transmit power is desirable for both experimentation and practical implementations, in particular for radio links with low margins such as, for example, those occurring in satellite communications [84], where every dB matters.

To perform the measurements of transmit RF power in [79] the USRP B200 was programmed using MATLAB and Simulink to transmit tones with frequency starting from 70 MHz to 6 GHz in 10 MHz increments. The USRP was configured to transmit using various UHD gain settings with values between 70 dB and the maximum allowed UHD value of 89.75 dB, and the RF power of the corresponding transmitted signal was recorded using a spectrum analyzer. A separate Matlab script is run to collect the power measurements automatically using the Instrument Control Toolbox, and the results obtained are separated into two categories as follows:

- The first category, shown in Fig. 11(a), includes lower UHD gains with values starting at 71 dB and increasing in 3 dB increments to 80 dB.
- The second category, shown in Fig. 11(b), includes high UHD gains with values starting at 83.75 dB and increasing in 3 dB increments to a UHD gain value of 89.75 dB, which is the maximum allowed UHD gain setting for the USRP B200.

The experimental results show that, with the UHD gain set at the maximum allowed value of 89.75 dB, the transmit power of the USRP B200 is indeed above 10 dBm over all its operating range as specified by the manufacturer [52]. Results in Fig. 11(b) also show that for UHD gains above 83 dB, the transmitter appears to be outside its linear operating range as changes in the UHD gain value do not reflect linearly in the transmit power variations. For UHD gains below 80 dB results in Fig. 11(a) show that the transmitter operation is linear as one can observe the clear 3 dB separation between neighboring curves, which is expected from the UHD gain settings under linear operation. It is thus conceivable to extrapolate the measurements done for these UHD gain values (71 dB, 74 dB, 77 dB, and 80 dB) to estimate transmit power levels for UHD gains below 70 dB.

Another interesting observation that can be based on the plots in Fig. 11 is that the transmit RF power of the USRP B200 varies with frequency. For example, looking at the curves for lower UHD gain values in Fig. 11(a), one can notice a variation of about 12 dB between the output power values at lower frequencies (70 MHz – 100 MHz) and those at higher frequencies (5.5 GHz – 6 GHz). The decrease in the transmit RF power of the USRP B200 observed when the operating frequency increases is likely due to the fact that the transmitter impedance is better matched at lower frequencies than at higher ones, and should be considered when the USRP is programmed to transmit at a specific power level in a given frequency band.

The specific application setup presented in [79], which focuses on a USRP B200 SDR, can be easily adapted to other SDR platforms. If MATLAB hardware support package for the SDR platform is not available, which is currently the case with the HackRF One and the Lime SDR platforms, then the SDR may be programmed to sweep its operating range using GNU Radio, while collecting the power measurements can still be accomplished using the MATLAB script that calls functions in the Instrument Control Toolbox.

VII. CONCLUSIONS

In this paper we provided a comprehensive introduction to SDRs, which are the building blocks of modern communication systems and are used in research, development, implementation, and teaching of wireless communications. The paper starts with a brief review of drivers and enabling technologies that contributed to the advancement of SDR platforms and to their ubiquitous presence in current and emerging wireless systems and networks, mentioning also future trends that will continue to keep SDRs at the forefront of communication technologies. Next, a brief theoretical background is given, reviewing bandpass signal representations in terms of I/Q components along with heterodyning for frequency up- and down-conversions, which are concepts that are essential to the understanding of SDR operation. The current architecture of SDR platforms is then presented, with details on the various processing stages in SDRs that include the RF front end, the IF processing, the digital front end, and the baseband processing. Finally, the paper presents several SDR platforms that have emerged as preferred choices for research, development, teaching, and radio enthusiast projects, reviews SDR programming alternatives, and presents two case studies demonstrating SDR applications and uses.

We are confident that, with the continued interest for SDRs in existing and future generations of wireless communication systems, the paper will serve as a useful reference for wireless researchers and developers working in the industry as well as for instructors teaching courses on wireless communications.

ACKNOWLEDGEMENT

The work of Dimitrie Popescu was supported by a Fulbright US Scholar grant awarded for the Spring 2022 semester.

REFERENCES

[1] J. Mitola, "Software Radios: Survey, Critical Evaluation and Future Directions," *IEEE Aerospace and Electronic Systems Magazine*, vol. 8, no. 4, pp. 25–36, 1993. DOI: 10.1109/62.210638

[2] ——, "The Software Radio Architecture," *IEEE Communications Magazine*, vol. 33, no. 5, pp. 26–38, 1995. DOI: 10.1109/35.393001

[3] Wireless Innovation Forum, "What is a Software Defined Radio," available online at https://www.wirelessinnovation.org/assets/documents/SoftwareDefinedRadio.pdf, accessed: April 24, 2022.

[4] National Instruments, "Software Defined Radio: Past, Present, and Future," available online at https://www.ni.com/hu-hu/innovations/white-papers/17/software-defined-radio--past--present--and-future.html, accessed: April 24, 2022.

[5] J. Melby, "JTRS and the Evolution Toward Software-Defined Radio," in *Proceedings 2002 IEEE Military Communications Conference*, vol. 2, pp. 1286–1290, Anaheim, CA, 2002. DOI: 10.1109/MILCOM.2002.1179664

[6] F. Vergari, "Software-Defined Radio: Finding Its Use in Public Safety," *IEEE Vehicular Technology Magazine*, vol. 8, no. 2, pp. 71–82, 2013. DOI: 10.1109/MVT.2013.2252292

[7] M. Öner and F. Jondral, "Air Interface Recognition for a Software Radio System Exploiting Cyclostationarity," in *Proceedings 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2004)*, vol. 3, pp. 1947–1951, Barcelona, Spain, September 2004. DOI: 10.1109/PIMRC.2004.1368338

[8] O. A. Dobre, "Signal Identification for Emerging Intelligent Radios: Classical Problems and New Challenges," *IEEE Instrumentation Measurement Magazine*, vol. 18, no. 2, pp. 11–18, 2015. DOI: 10.1109/MIM.2015.7066677

[9] G. Retz, H. Shanan, K. Mulvaney, S. O'Mahony, M. Chanca, P. Crowley, C. Billon, M. Kalimuddin Khan, J. J. Lopez Orive, and P. Quinlan, "Radio Transceivers for Wireless Personal Area Networks Using IEEE 802.15.4," I*EEE Communications Magazine*, vol. 47, no. 9, pp. 150–158, 2009. DOI: 10.1109/MCOM.2009.5277469

[10] M. Cummings and S. Haruyama, "FPGA in the Software Radio," *IEEE Communications Magazine*, vol. 37, no. 2, pp. 108–112, February 1999. DOI: 10.1109/35.747258

[11] S. Balasubramanian, S. Boumaiza, H. Sarbishaei, T. Quach, P. Orlando, J. Volakis, G. Creech, J. Wilson, and W. Khalil, "Ultimate Transmission," *IEEE Microwave Magazine*, vol. 13, no. 1, pp. 64–82, 2012. DOI: 10.1109/MMM.2011.2173983

[12] L. Pucker, "Is There Really Such a Thing as a "DSP" Anymore?" *IEEE Communications Magazine*, vol. 44, no. 9, pp. 34–36, 2006. DOI: 10.1109/MCOM.2006.1705976

[13] C. Dick and F. Harris, "FPGA Signal Processing Using Sigma-Delta Modulation," *IEEE Signal Processing Magazine*, vol. 17, no. 1, pp. 20–35, 2000. DOI: 10.1109/79.814644

[14] P. Burns, Software Defined Radio for 3G. Norwood, MA: Artech House, 2002.

[15] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-Advanced: Next-Generation Wireless Broadband Technology," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, 2010. DOI: 10.1109/MWC.2010.5490974

[16] R. Schneiderman, "LTE Base Stations, Mobile Devices Flood Telecom, Consumer Markets," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 9–14, 2012. DOI: 10.1109/MSP.2012.2186185

[17] Y. Kyung, T. M. Nguyen, K. Hong, J. Park, and J. Park, "Software Defined Service Migration Through Legacy Service Integration Into 4G Networks and Future Evolutions," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 108–114, 2015. DOI: 10.1109/MCOM.2015.7263353

[18] T. Ulversøy, "Software Defined Radio: Challenges and Opportunities," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 531–550, 2010. DOI: 10.1109/SURV.2010.032910.00019

[19] B. van Liempd, J. Borremans, E. Martens, S. Cha, H. Suys, B. Verbruggen, and J. Craninckx, "A 0.9 V 0.4–6 GHz Harmonic Recombination SDR Receiver in 28 nm CMOS With HR3/HR5 and IIP2 Calibration," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 8, pp. 1815–1826, 2014. DOI: 10.1109/JSSC.2014.2321148

[20] G. Wang, B. Yin, K. Amiri, Y. Sun, M. Wu, and J. R. Cavallaro, "FPGA Prototyping of a High Data Rate LTE Uplink Baseband Receiver," in *Proceedings of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 248–252, Pacific Grove, CA, November 2009. DOI: 10.1109/ACSSC.2009.5470112.

[21] R. G. Machado and A. M. Wyglinski, "Software-Defined Radio: Bridg- ing the Analog–Digital Divide," *Proceedings of the IEEE*, vol. 103, no. 3, pp. 409–423, 2015. DOI: 10.1109/JPROC.2015.2399173

[22] C. Erdmann, D. Lowney, A. Lynam, A. Keady, J. McGrath, E. Cullen, D. Breathnach, D. Keane, P. Lynch, M. De La Torre, R. De La Torre, P. Lim, A. Collins, B. Farley, and L. Madden, "A Heterogeneous 3D-IC Consisting of Two 28 nm FPGA Die and 32 Reconfigurable High-Performance Data Converters," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 258–269, 2015. DOI: 10.1109/JSSC.2014.2357432

[23] F. Gringoli, P. Patras, C. Donato, P. Serrano, and Y. Grunenberger, "Performance Assessment of Open Software Platforms for 5G Prototyping," *IEEE Wireless Communications*, vol. 25, no. 5, pp. 10–15, October 2018. DOI: 10.1109/MWC.2018.1800049

[24] K. Lin, W. Wang, X. Wang, W. Ji, and J. Wan, "QoE-Driven Spectrum Assignment for 5G Wireless Networks Using SDR," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 48–55, December 2015. DOI: 10.1109/MWC.2015.7368824

[25] X. Xiong, W. Xiang, K. Zheng, H. Shen, and X. Wei, "An Open Source SDR-Based NOMA System for 5G Networks," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 24–32, December 2015. DOI: 10.1109/MWC.2015.7368821

[26] S. Gökceli, P. P. Campo, T. Levanen, J. Yli-Kaakinen, M. Turunen, M. Aléln, T. Riihonen, A. Palin, M. Renfors, and M. Valkama, "SDR Prototype for Clipped and Fast-Convolution Filtered OFDM for 5G New Radio Uplink," *IEEE Access*, vol. 8, pp. 89 946–89 963, May 2020. DOI: 10.1109/ACCESS.2020.2993871

[27] Y. Liu, C. Li, X. Xia, X. Quan, D. Liu, Q. Xu, W. Pan, Y. Tang, and K. Kang, "Multiband User Equipment Prototype Hardware Design for 5G Communications in Sub-6-GHz Band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 2916–2927, July 2019. DOI: 10.1109/TMTT.2019.2904234

[28] G. P. Fettweis, "5G and the Future of IoT," in *Proceedings 42nd European Solid-State Circuits Conference – ESSCIRC 2016*, Lausanne, Switzerland, September 2016., pp. 21–24. DOI: 10.1109/ESSCIRC.2016.7598234

[29] J. Liang, H. Chen, and S. C. Liew, "Design and Implementation of Time-Sensitive Wireless IoT Networks on Software-Defined Radio," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 2361–2374, February 2022. DOI: 10.1109/JIOT.2021.3094667

[30] P. Solic, Z. Blazevic, M. Skiljo, L. Patrono, R. Colella, and J. J. P. C. Rodrigues, "Gen2 RFID as IoT Enabler: Characterization and Performance Improvement," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 33–39, June 2017. DOI: 10.1109/MWC.2017.1600431

[31] G. Saxl, L. Görtschacher, T. Ussmueller, and J. Grosinger, "Software-Defined RFID Readers: Wireless Reader Testbeds Exploiting Software-Defined Radios for Enhancements in UHF RFID Systems," *IEEE Microwave Magazine*, vol. 22, no. 3, pp. 46–56, March 2021. DOI: 10.1109/MMM.2020.3042408

[32] SEMTECH, "What is LoRa®?" available online at https://www.semtech.com/lora/what-is-lora, accessed: May 23, 2022.

[33] R. Ghanaatian, O. Afisiadis, M. Cotting, and A. Burg, "Lora Digital Receiver Analysis and Implementation," in *Proceedings 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1498–1502., Brighton, UK, May 2019. DOI: 10.1109/ICASSP.2019.8683504

[34] C. Bernier, F. Dehmas, and N. Deparis, "Low Complexity LoRa Frame Synchronization for Ultra-Low Power Software-Defined Radios," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 3140–3152, May 2020. DOI: 10.1109/TCOMM.2020.2974464

[35] P. I. Theoharis, R. Raad, F. Tubbal, M. U. Ali Khan, and S. Liu, "Software-Defined Radios for CubeSat Applications: A Brief Review and Methodology," *IEEE Journal on Miniaturization for Air and Space Systems*, vol. 2, no. 1, pp. 10–16, March 2021. DOI: 10.1109/JMASS.2020.3032071

[36] M. De Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite Communications Supporting Internet of Remote Things," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 113–123, February 2016. DOI: 10.1109/JIOT.2015.2487046

[37] C. Gavrilă, V. Popescu, M. Alexandru, M. Murroni, and C. Sacchi, "An SDR-Based Satellite Gateway for Internet of Remote Things (IoRT) Applications," *IEEE Access*, vol. 8, pp. 115423–115436, July 2020. DOI: 10.1109/ACCESS.2020.3004480

[38] S. Haykin, *Communication Systems*, 4th ed. NewYork, NY: John Wiley & Sons, Inc., 2001.

[39] G. Kolumban, T. I. Krebesz, and F. C. Lau, "Theory and Application of Software Defined Electronics: Design Concepts for the Next Generation of Telecommunications and Measurement Systems," *IEEE Circuits and Systems Magazine*, vol. 12, no. 2, pp. 8–34, 2012. DOI: 10.1109/MCAS.2012.2193435

[40] G. Kolumban, "Software Defined Electronics: A Revolutionary Change in Design and Teaching Paradigm of RF Radio Communications Systems," *ICT Express*, vol. 1, no. 1, pp. 44–54, 2015. DOI: 10.1016/S2405-9595(15)30021-7

[41] Wikipedia, "FM Broadcasting," available online at https://en.wikipedia.org/wiki/FM_broadcasting, accessed: April 30, 2022.

[42] P. Cruz, N. B. Carvalho, and K. A. Remley, "Designing and Testing Software-Defined Radios," *IEEE Microwave Magazine*, vol. 11, no. 4, pp. 83–94, June 2010. DOI: 10.1109/MMM.2010.936493

[43] A. Udalcovs, M. Levantesi, P. Urban, D. A. A. Mello, R. Gaudino, O. Ozolins, and P. Monti, "Total Cost of Ownership of Digital vs. Analog Radio-Over-Fiber Architectures for 5G Fronthauling," *IEEE Access*, vol. 8, pp. 223 562–223 573, December 2020. DOI: 10.1109/ACCESS.2020.3044396

[44] B. Brannon, "Wideband Radios Need Wide Dynamic Range Converters," Analog Dialogue, vol. 29, no. 2, pp. 11–12, April 1995, available online at https://www.analog.com/en/analog-dialogue/articles/wideband-radios-need-wide-dynamic-range-converters.html, accessed May 3, 2022.

[45] R. Walden, "Analog-to-Digital Converter Survey and Analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, 1999. DOI: 10.1109/49.761034

[46] R. Schaller, "Moore's Law: Past, Present and Future," *IEEE Spectrum*, vol. 34, no. 6, pp. 52–59, 1997. DOI: 10.1109/6.591665

[47] A. A. Abidi, "The Path to the Software-Defined Radio Receiver," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 954–966, 2007. DOI: 10.1109/JSSC.2007.894307

[48] A. I. Hussein and W. B. Kuhn, "Bandpass ΣΔ Modulator Employing Undersampling of RF Signals for Wireless Communication," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 7, pp. 614–620, 2000. DOI: 10.1109/82.850420

[49] N. Beilleau, H. Aboushady, F. Montaudon, and A. Cathelin, "A 1.3V 26mW 3.2GS/s Undersampled LC Bandpass ΣΔADC for a SDR ISM-band Receiver in 130nm CMOS," in *Proceedings 2009 IEEE Radio Frequency Integrated Circuits Symposium*, Boston, MA, June 2009., pp. 383–386., DOI: 10.1109/RFIC.2009.5135563

[50] A. Aneja and X. J. Li, "Multiband LNAs for Software-Defined Radios: Recent Advances in the Design of Multiband Reconfigurable LNAs for SDRs in CMOS, Microwave Integrated Circuits Technology," *IEEE Microwave Magazine*, vol. 21, no. 7, pp. 37–53, July 2020. DOI: 10.1109/MMM.2020.2985189

[51] F. M. Ghannouchi, "Power Amplifier and Transmitter Architectures for Software Defined Radio Systems," *IEEE Circuits and Systems Magazine*, vol. 10, no. 4, pp. 56–63, fourth quarter 2010. DOI: 10.1109/MCAS.2010.938639

[52] Ettus Research, "The USRP B200/B210/B200mini/B205mini," available online at https://kb.ettus.com/B200/B210/B200mini/B205mini, accessed: May 2, 2022.

[53] R. W. Stewart, K. W. Barlee, D. S. W. Atkinson, and L. H. Crockett, *Software Defined Radio Using MATLAB & Simulink and the RTL-SDR*, 1st ed. Glasgow, Scotland, UK: Strathclyde Academic Media, 2015, available online at https://www.desktopsdr.com.

[54] R. Akeela and B. Dezfouli, "Software-Defined Radios: Architecture, State-of-the-Art, and Challenges," *Computer Communications*, vol. 128, pp. 106–125, September 2018. **DOI**: 10.1016/j.comcom.2018.07.012

[55] T. Hentschel, M. Henker, and G. Fettweis, "The Digital Front-End of Software Radio Terminals," *IEEE Personal Communications*, vol. 6, no. 4, pp. 40–46, August 1999. **DOI**: 10.1109/98.788214

[56] D. M. Molla, H. Badis, L. George, and M. Berbineau, "Software Defined Radio Platforms for Wireless Technologies," *IEEE Access*, vol. 10, pp. 26 203–26 229, March 2022. **DOI**: 10.1109/ACCESS.2022.3154364

[57] A. M. Wyglinski, D. P. Orofino, M. N. Ettus, and T. W. Rondeau, "Revolutionizing Software Defined Radio: Case Studies in Hardware, Software, and Education," *IEEE Communications Magazine*, vol. 54, no. 1, pp. 68–75, 2016. **DOI**: 10.1109/MCOM.2016.7378428

[58] Ettus Research, "Products," available online at https://www.ettus.com/products/, accessed: May 2, 2022.

[59] T. B. Welch and S. Shearman, "Teaching Software Defined Radio Using the USRP and Labview," in *Proceedings 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing – ICASSP 2012*, pp. 2789–2792, Kyoto, Japan, March 2012. **DOI**: 10.1109/ICASSP.2012.6288496

[60] D. Pu and A. Wyglinski, Digital Communication Systems with Software-Defined Radio. Boston, MA: Artech House, 2013.

[61] M. El-Hajjar, Q. A. Nguyen, R. G. Maunder, and S. X. Ng, "Demonstrating the Practical Challenges of Wireless Communications Using USRP," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 194–201, May 2014. **DOI**: 10.1109/MCOM.2014.6815912

[62] R. Heath, *Introduction to Wireless Digital Communication: A Signal Processing Perspective*. Hoboken, NJ: Prentice Hall, 2017.

[63] "RTL-SDR and Software Defined Radio News and Projects," available online at https://www.rtl-sdr.com, accessed: May 26, 2022.

[64] Analog Devices, "ADALM-PLUTO: Software-Defined Radio Active Learning Module," available online at https://www.analog.com/en/design-center/evaluation-hardware-and-software/evaluation-boards-kits/ADALM-PLUTO.html, accessed: May 26, 2022.

[65] S. G. Bilén, A. M. Wyglinski, C. R. Anderson, T. Cooklev, C. Dietrich, B. Farhang-Boroujeny, J. V. Urbina, S. H. Edwards, and J. H. Reed, "Software-Defined Radio: A New Paradigm for Integrated Curriculum Delivery," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 184–193, May 2014. **DOI**: 10.1109/MCOM.2014.6815911

[66] M. Rice and M. McLernon, "Teaching Communications with SDRs: Making It Real for Students," *IEEE Communications Magazine*, vol. 57, no. 11, pp. 14–19, November 2019. **DOI**: 10.1109/MCOM.001.1900185

[67] Great Scott Gadgets, "HackRF One," available online at https://greatscottgadgets.com/hackrf/one/, accessed: May 26, 2022.

[68] Lime Microsystems, "LimeSDR," available online at https://limemicro.com/products/boards/limesdr/, accessed: May 26, 2022.

[69] M. Gummineni and T. R. Pollipali, "Implementation of Reconfigurable Transceiver using GNU Radio and HackRF One," *Wireless Personal Communications*, vol. 112, pp. 889–905, January 2020. **DOI**: 10.1109/MCOM.2014.6815911

[70] O. Popescu, J. Musson, and D. C. Popescu, "Using Open-Source Software Defined Radio Platforms for Empirical Characterization of Man-Made Impulsive Noise," *IEEE Electromagnetic Compatibility Magazine*, vol. 9, no. 4, pp. 54–61, 4th Quarter 2020. **DOI**: 10.1109/MEMC.2020.9327997

[71] Lime Microsystems, "Myriad RF," available online at https://limemicro.com/initiatives/myriad-rf/, accessed: Sep. 12, 2022.

[72] Myriad RF, "Novena Open Hardware Computing Platform," available online at https://myriadrf.org/projects/component/novena-rf/, accessed: Sep. 12, 2022.

[73] "GNU Radio," available online at https://gnuradio.org, accessed: May 26, 2022.

[74] MathWorks, "Supported Hardware – Software Defined Radio," available online at https://www.mathworks.com/help/comm/supported-hardware-software-defined-radio.html, accessed: May 26, 2022.

[75] National Instruments, "What is LabVIEW?" available online at https://www.ni.com/en-us/shop/labview.html, accessed: May 26, 2022.

[76] "GNU Radio Out of Tree Modules," available online at https://wiki.gnuradio.org/index.php/OutOfTreeModules, accessed: May 26, 2022.

[77] MathWorks, "MATLAB Central File Exchange," available online at https://www.mathworks.com/matlabcentral/fileexchange/, accessed: May 26, 2022.

[78] "NI Community," available online at https://forums.ni.com, accessed: May 26, 2022.

[79] M. W. O'Brien, J. S. Harris, O. Popescu, and D. C. Popescu, "An Experimental Study of the Transmit Power for a USRP Software-Defined Radio," in *Proceedings 12th IEEE International Communications Conference – COMM 2018*, pp. 377–380, Bucharest, Romania, June 2018. **DOI**: 10.1109/ICComm.2018.8484809.

[80] S. Miyamoto, M. Katayama, and N. Morinaga, "Performance Analysis of QAM Systems Under Class A Impulsive Noise Environment," *IEEE Transactions on Electromagnetic Compatibility*, vol. 37, no. 2, pp. 260–267, 1995. **DOI**: 10.1109/15.385891

[81] ——, "Receiver Design Using the Dependence Between Quadrature Components of Impulsive Radio Noise," in *Proceedings 1995 IEEE International Conference on Communications (ICC)*, vol. 3, pp. 1784–1789, Seattle, WA, June 1995. **DOI**: 10.1109/ICC.1995.524506

[82] M. G. Sanchez, L. de Haro, M. C. Ramon, A. Mansilla, C. M. Ortega, and D. Oliver, "Impulsive Noise Measurements and Characterization in a UHF Digital TV Channel," *IEEE Transactions on Electromagnetic Compatibility*, vol. 41, no. 2, pp. 124–136, 1999. **DOI**: 10.1109/15.765101

[83] P. Torío, M. G. Sanchez, and I. Cuinas, "An Algorithm to Simulate Impulsive Noise," in *Proceedings 19th IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Dubrovnic, Croatia, September 2011.

[84] O. Popescu, "Power Budgets for CubeSat Radios to Support Ground Communications and Inter-Satellite Links," *IEEE Access*, vol. 5, pp. 12 618–12 625, July 2017. **DOI**: 10.1109/ACCESS.2017.2721948

Dimitrie C. Popescu received the Engineering Diploma and PhD degree in electrical and computer engineering from Polytechnic Institute of Bucharest and Rutgers University, respectively. He is currently a Professor in the ECE Department, Old Dominion University, Norfolk, Virginia. His research interests are focused on wireless communication systems and include software defined and cognitive radios, spectrum sensing and dynamic spectrum access for cognitive radios, modulation classification, interference mitigation and transmitter/receiver optimization to support quality of service, vehicular networks, and signal processing for communications and radar systems. He is an active Senior Member of the IEEE currently serving as an associate editor for IEEE Open Journal of the Communications Society and participating regularly in the technical program and organizing committees for the IEEE Global Telecommunications Conference (GLOBECOM), the IEEE International Conference on Communications (ICC), the IEEE Wireless Communications and Networking Conference (WCNC), and the IEEE Vehicular Technologies Conference (VTC).

Rolland Vida graduated as valedictorian at the Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj, Romania, in 1997. He received the PhD degree in Computer Networks from Université Pierre et Marie Curie, Paris, France, in 2003. He is currently an Associate Professor with Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics, and the Head of HSN Lab, an academic research laboratory that is a strategic partner of Ericsson since 1992. In the last ten years, he has held different leading positions in IEEE Communications Society, IEEE Smart Cities, and IEEE Sensors Council. He is a member of the Steering Committee of IEEE Internet of Things journal, and Associate Editor of IEEE Sensors Letters. He served as the TPC Co-Chair for the IEEE Sensors Conference in 2019, 2020, and 2022. Rolland Vida is a Senior Member of IEEE.

Decentralized Authentication Mechanism for Mobile Ad hoc Networks

Hafida Khalfaoui, Abderrazak Farchane and Said Safi

Abstract—**Covid 19 has dramatically changed people's lives around the world. It has shut down schools, companies and workplaces, forcing individuals to stay at home and comply to quarantine orders. Thus, individuals have resorted to the Internet as a means for communicating and sharing information in different domains. Unfortunately, some communities are still unserved by commercial service providers. Mobile Adhoc Network (MANET) can be used to fill this gap. One of the core issues in MANET is the authentication of the participating nodes. This mechanism is a fundamental requirement for implementing access control to network resources by confirming a user's identity. In recent years, security experts worldwide proposed distributed authentication for MANET due to the lack of a central authority to register and authenticate nodes. In this article, decentralized authentication based on the technology of fog computing and the concept of the blockchain is proposed. The evaluation of this mechanism satisfies the diverse security requirements and strongly protects the networks from attacks.**

Index Terms—**Authentication, Blockchain, Community network, Fog computing, Mobile Ad hoc Network.**

I. INTRODUCTION

South Africa is one of the most underdeveloped societies in the world. Around 55.5% of the population is poor, with rural areas accounting for 80%. The majority of poor Africans cannot afford basic necessities or have access to resources, adequate education or essential services. Community networks give rural and underprivileged communities the chance to own, control and market their communication services, which could help bridge socioeconomic.

Zenzeleni ("Do it yourself" in isiXhosa) is the first community network in South Africa, founded in 2013 by the University of the Western Cape (UWC) PhD students and the Mankosi community's tribal government. Due to a shortage of electricity, Zenzeleni began as a local wireless network for delivering free calls by connecting analog phones via solar-powered routers. It was later updated to add an external internet connection for making calls to national numbers via a 3G modem [1]. The above has not happened without challenges: This project took six years to complete, and communities have faced several legal, technological, financial and social restrictions, which have necessitated collaboration to overcome them.

Mobile Ad hoc network (MANET) is a good solution to reduce communications costs in these community networks. They are composed of various computer systems or mobile phones called nodes, which can connect autonomously by radio waves without needing a fixed infrastructure such as routers or access points. This solution can open the door to many individuals to communicate, share information and have novels of what happens in their society with less cost and effort.

The authors are with the LIMATI Laboratory, Department of Mathematics and Computer Science, Polydisciplinary Faculty, Sultan Moulay Slimane University, PO Box 592, Beni Mellal, 23000, Morocco. (E-mail: hafidakhalfaoui1996@gmail.com, a.farchane@gmail.com, safi.said@gmail.com)

DOI: 10.36244/ICJ.2022.3.4

A. Problem statement

MANET has advanced dramatically due to the proliferation of mobile devices and enhancements in wireless communication. In some situations, people need to set up an instant and temporary multimedia network where devices can communicate immediately, such as in conferences, classrooms, home networks or other civilian locations like a sports stadium, ship or small aircraft [2].

The specifics of MANET show how these networks are, by nature, a great challenge for IT security. The operating approach for nodes is to receive packets and transmit them to the next-hop through intermediate nodes until they arrive at their destination [3]. They are resource-constrained devices that cannot secure and defend themselves, making them vulnerable to hacking and compromise. Indeed, the security of nodes and exchanged messages is essential. This work will focus precisely on authentication as a starting point of security, as it is required for both old and new nodes to connect to the network for communication. This security objective will protect the network by permitting only legitimate nodes to gain access to their resources and will defeat all attacks using the identity or role impersonation. Consequently, the authentication scheme must be robust, scalable and resistant to known threats. For MANET, many theoretical studies exist, but ultimately few practical applications can satisfy all the constraints inherent in Ad hoc networks.

One of the most popular authentication mechanisms in MANETs is clustering, such as proposed in [4], [5]. This algorithm separates the network into clusters, with the cluster head (CH) is chosen from the node with the highest weight. The CH then utilizes the AUCRES (authentication response) technique to authenticate nodes using unique identification and long keys. The problem with this algorithm appears when the cluster head is attacked. To improve this technique, security experts worldwide have been concentrating on the blockchain. This technology is well suited for MANET authentication and access control services due to its decentralized system, its cryptographic features and other characteristics such as enhanced unforgeability, reliability and fault tolerance.

B. Paper's contributions

This work proposes and evaluates the security of an authentication mechanism based on blockchain technology for

MANET's nodes. The architecture of this system includes mobile nodes and Ethereum Blockchain managed by smart contract rules. Fog nodes are also used to synchronize data associated with registered nodes. This paper's principal contributions can be outlined as follows:

- A decentralized authentication approach is presented without needing an intermediary or trusted third party. It uses the blockchain to register and authenticate devices, with access tokens calculated by the smart contract.
- The complete system is shown in detail, including the architecture, sequence diagrams and the explication of different interactions between nodes and the smart contract.
- Security analysis and discussion are explained about how this proposed authentication method satisfies security goals (identification, confidentiality, integrity and non-repudiation) and resists certain types of attacks.

C. Paper structure

The rest of this paper is shown as follows: Section II shows a brief overview of MANET and defines blockchain and fog computing technologies. Section III discusses a view of decentralized authentication methods in the literature. Section IV spots the light of the proposed mechanism. Section V presents its implementation and talks about the evaluation of results. Finally, Section VI gives the conclusion and the future work.

II. BACKGROUND

This section gives a summary of MANET, blockchain and fog computing technologies.

A. Mobile Ad hoc Network

A Mobile Ad hoc Network is an instant network of the mobile nodes without a fixed infrastructure. There are two modes of MANET: The first mode is when the nodes can directly interact with other nodes in their radio range. The second one is called multi-hop communication, when the intermediate nodes are employed to communicate with the nodes beyond their radio ranges [6].
MANET is vulnerable to various security attacks because of its characteristics, including dynamic topology, lack of central management and unsecured medium. The absence of a centralized administration pushes the nodes to communicate on a level of mutual trust. This property renders MANET more vulnerable to exploitation by internal attackers. Moreover, wireless links make it also more accessible for malicious nodes to get network resources.
There are two classifications of attacks according to [7], [8]. On the one hand, the attacks in MANET can be classified according to the attack's behavior as passive or active attacks:

- **Passive attacks**: Intercept data when it transits a network. It is difficult to detect the intrusion because this type of attack does not cause any noticeable perturbation or malicious activity disrupting the network's normal function. Traffic analysis, Traffic monitoring and Eavesdropping are examples of these attacks.
- **Active attacks** : Are more disruptive because the malicious nodes affect the network traffic and the transmissions by generating congestion and false routing information. Still, the dynamic nature of these attacks is quite easy to detect and prevent. Modification, Impersonation and Fabrication are examples of active attacks.

On the other hand, attacks can also be classified as either internal or external:

- **External attacks**: Are launched by unlicensed nodes that aren't part of the network, and they may flood the network with fake packets and sometimes imitate legitimate nodes. External malicious nodes mainly aim to create congestion or disrupt normal network operations.
- **Internal attacks**: Are initiated by the network's legitimate nodes; these malicious nodes take resources from other network nodes arbitrarily and selfishly like battery power, processing power and bandwidth.

B. Blockchain technology

Blockchain is a distributed ledger for storing cryptographically signed data. When a user creates a transaction over a blockchain network, the new transaction is grouped with others to form a block as shown in Figure 1. Once the block is created, active nodes, known as miners, ensure that transactions inside the block follow predetermined criteria using a consensus procedure like proof of Work (PoW). After, the miner who validates the block is rewarded, and the blockchain stores the verified block. Finally, to avoid a single point of failure, each node in the network keeps a copy of the blockchain on hand. These copies are simultaneously updated and verified [9].

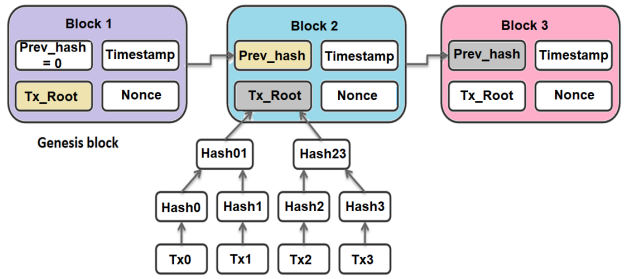


Fig. 1. Blockchain

C. Fog Computing

Fog computing is an extension of cloud computing, in which data processing is transferred to the near of devices instead of sending it to the cloud. This mechanism will lessen the burden of the Internet and give quick processing. The basic instruction of Fog computing consists of three layers: End devices, fog nodes and cloud servers [10].

III. STATE OF THE ART

Authentication is an essential step to ensure the legitimacy of nodes and the network's safety. Current state-of-the-art authentication solutions are not adequate for MANET due to the open environment, resource-constrained devices, centralization and high energy consumption. For this reason, low-power, low-storage, low-latency and low-communication-overhead authentication techniques are required. Recently, to fill gaps in existing technologies, current researchers have focused on blockchain mechanisms, each in their perspective.

Jarjis and Kadir [11] presented a blockchain implementation for Ad Hoc On-Demand Distance Vector (AODV). The AODV is considered an excellent routing protocol in MANET because it employs the least overhead by reducing information exchanged in messages. However, this weakens the authentication and checking integrity mechanism as nodes do not have enough knowledge about the identities of neighboring nodes. As a result, the authors modify the AODV protocol to BAODV, adding an extra field to the Route Request (RREQ) and Route Reply (RREP) options. This procedure begins by including a field called (Bnode) to the RREQ packet's header. This field stores the malicious node's address if it breaks into the network or tries to exploit its resources. Then, each node verifies that the previous node's address matches (Bnode). This solution provides node verification features and uses a chaining technique to detect Impersonation and Black-hole attacks as a secure system. It removes the malicious node involved path without extra processing that affects network performance with additional overhead.

Yang and Hwanseok [12] proposed a decentralized protocol for authentication by integrating blockchain and an area-based hierarchical structure. After dividing the entire network into regions, the region's most reliable node is designated as a Region Certificate Authority (RCA). The highest reliability node among the RCA will be elected and selected as Top Certificate Authority (TCA). Then, the RCA node issues a group key to the TCA node and a member key to the member nodes in the area. Only nodes that give the corresponding key can participate in data transmission. This protocol was implemented using blockchain technology through transaction creation, block packaging and verification processes. Blockchain, as a peer-to-peer network with a hierarchical structure of nodes, provides a decentralized solution for node authentication that can prevent a single point of failure of centralized mechanisms and the forgery of authentication information for nodes participating in the network.

The blockchain is also implemented in the other types of MANET like Flying Ad hoc Network (FANET) and Vehicular Ad hoc Network (VANET).

VANET is widely regarded as the main platform for vehicle-to-vehicle communication. This system improves transportation security and reduces the number of accidents [13]. However, it is more vulnerable to attacks due to its high mobility and dynamic topology. As a result, with VANET, a secure and anonymous authentication mechanism is critical. Azees et al. [14] proposed an anonymous authentication mechanism based on blockchain in which RSUs can anonymously authenticate the vehicles and perform future communications through the shared session key. Moreover, the reauthentication of automobiles between roadside units became faster thanks to the secure transfer of authentication codes between adjacent Road Side Units (RSUs). Among the advantages of this mechanism, the integrity of the transmitting message is preserved due to the support of the blockchain. In addition, the performance analysis done by the authors of this work proves its efficiency in terms of communication, computational and storage costs. So it is convenient for real-time applications.

For FANET; the network of unmanned aerial vehicles (UAVs) that can execute many activities, for example, delivery of goods, rescue missions and terrain monitoring; Kashish et al. [15] proposed a decentralized network authentication approach that adopts a blockchain-based public key infrastructure. The blockchain allows for shared public keys and the necessary information for the authentication process. This decentralized scheme reduces major security risks associated with a single point of failure. The experimental results show that this method works with constant throughput, latency and approximately constant message overhead as transaction size increase for a given network size. In addition, the authors mention that this architecture presented in the context of flying nodes can apply to other types of ad hoc network nodes.

The studies mentioned above demonstrate the utility of blockchain for different types of MANET. Also, the Internet of Things uses this technology as cited in [9]. For this reason, this work benefits the advantages of this technology and Fog computing to obtain a suitable authentication solution for MANET.

IV. PROPOSED MECHANISM

A. Architecture

This project aims to create a distributed authentication mechanism based on the blockchain that permits communication between nodes from different groups in the network. Figure 2 depicts the architecture of this mechanism divided into two layers: the device layer and the fog layer.

The device layer contains two types of mobile nodes installed in the network in clusters: admins and devices.

- **Admins**: They are in charge of controlling devices access, and they serve as the certification authority in each group. These admins should have high-performance RAM, a faster processor and a high storage capacity. In addition, It is preferable to be fixed computers connected to redundant power supplies as servers to stay permanently in service to register and authenticate nodes at any time. Many algorithms in the literature are created to select suitable admins as cited in [16]. Artificial Neural Networks (ANNs) are established to develop a clustering algorithm using weight-based parameters to choose cluster heads utilizing four inputs: mobility, packet drop, energy and the number of neighbor nodes. ANNs are computer systems inspired by the biological neural networks that form human brains [17]. This work does not focus on how selecting admins. The network creator chooses the other admins, and each one of them is assigned to the

nearest fog node that helps share the blockchain between all admins.

- **Devices**: Each device is identified by its unique identified information and can be added by one admin.

The fog layer contains, in general, a network of fog nodes that enable the implementation of fog services. It also performs localized storage and processing locally to the devices in order to reduce cloud latency and response time. Each group in the network is associated with the nearest fog node. These fog nodes communicate to ensure data synchronization for authentication. When an admin registers any device, it will add a block to the blockchain and share it with its associated fog node. This latter updates the blockchain and distributes it to other fog nodes. As a result, all admins can receive the novel copy of the blockchain and will be able to authenticate all network nodes in the case of node mobility than a group to other groups.

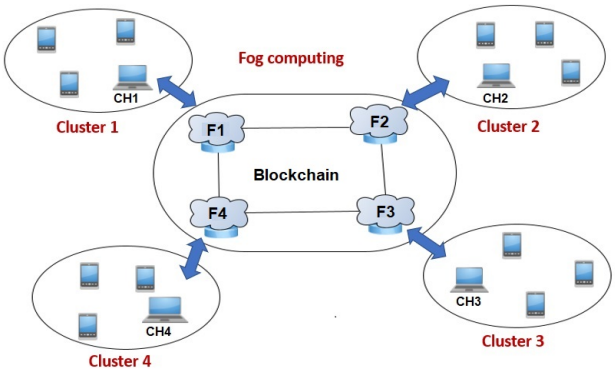


Fig. 2. The architecture of the proposed mechanism

B. Systems functioning

The network's function is to communicate by forwarding packets between nodes without any security guarantees. As a result, focusing on the authentication strategy allows reliable communication between several nodes over an untrusted network and prohibits the access of malicious users. In this paper, the transaction includes the node's registration and authentication request. The hash value of the latter is stored in the blockchain. The following sections will describe the different algorithms of the authentication mechanism.

1) Assumptions:

In this proposed mechanism, we assume that:
- Admins are trusted and all their actions are legitimate.
- Each admin chooses its cluster identifier (CID) and will share it only with trusted nodes. A new node can not sign up in the cluster if it does not have its CID.
- Each node has a private/public key pair for data encryption and integrity in the system.

2) Initialization phase:

This approach starts with an initialization phase. In the smart contract's constructor, the first admin is created using the address of the smart contract's creator, a unique identifier AID and a cluster head identifier CID. Then, the smart contract calculates the token of this first admin and saves it in the blockchain. Nodes credentials will be saved using their hash instead of keeping them in plain text.

3) Registration phase:

In the registration phase, each group with its associated nodes is registered in the blockchain. This phase is divided into two stages: Admin registration and device registration.

a- Admin registration

During this phase, each admin chooses a unique identifier (CID) to register its cluster on the blockchain. This operation is only restricted to administrator nodes. The following steps and the sequence diagram of Figure 3 describe the phase of admin registration.

1) A new admin can sign up by sending its information, including the address AIP, the cluster identifier CID and the unique identifier AID to an old admin using **addAdmin**(AIP, CID, AID) function.
2) After, the smart contract checks if the admin in question is already registered in the blockchain or not by searching the hash of its AID, AIP and CID.
3) If this admin does not exist in the blockchain, a token= **keccak256**(AIP, CID, AID) will be created, and the new admin will be saved in the blockchain. Keccak256 is a hash function.
4) Finally, the smart contract accepts the transaction triggering the **AdminAdded** event and generates a new block designated successful registration of the admin node. The event is a broadcast to all nodes in the blockchain.

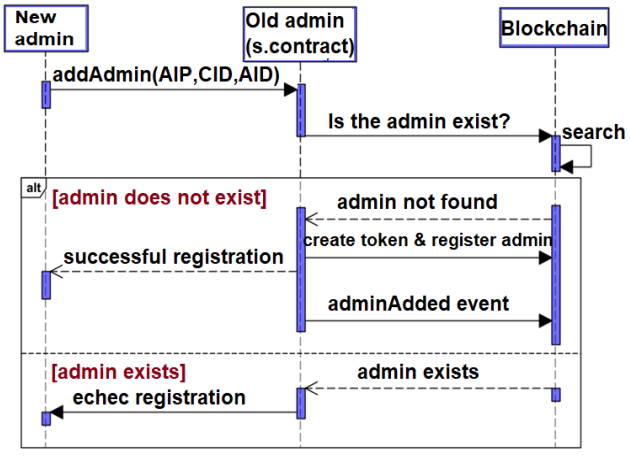


Fig. 3. Sequence diagram for clusters registration

b- Device registration phase

After the successful registration of the cluster in question, devices can register. The following steps and the sequence diagram of Figure 4 describe the device registration phase.

1) Admin can add a device to the network by sending the information, including the address of device DIP, the

cluster identifier CID and the unique device identifier DID to **addDevice**(DIP, CID, DID) function.

2) The function initially verifies if the CID presented corresponds to the adding admin. If not, the transaction will be terminated by error.
3) If yes, the function checks if this device is already existed by comparing the hash of its DIP, CID and DID to the nodes list saved in the blockchain.
4) If the device does not exist in the blockchain, a token = **keccak256**(DIP, CID, DID) is created, and the new device is registered.
5) Finally, the smart contract accepts the transaction triggering the **DeviceAdded** event and generates a new block designated successful registration of the new device.

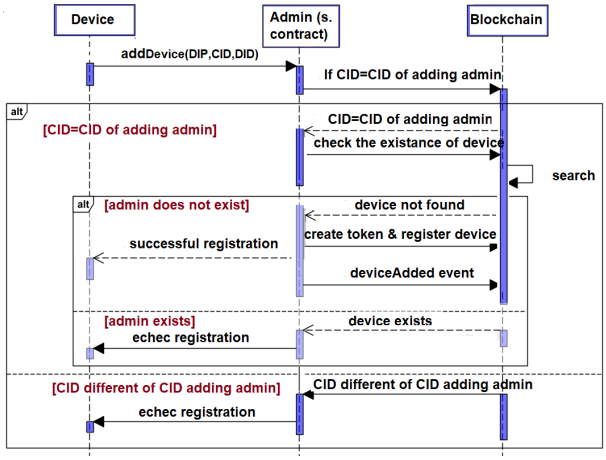


Fig. 4. Sequence diagram for devices registration

4) Authentication phase:

In the authentication phase, any admin can authenticate the devices registered in the blockchain to communicate with each other. The following steps and the sequence diagram of Figure 5 describe the device authentication phase.

1) The device sends a request authentication containing its DID, DIP and CID to any admin. This latter uses the smart contract to verify the received packet's legitimacy by creating its token and comparing it with the list of tokens stored in the blockchain.
2) If the token corresponding to the DIP exists, the admin successfully authenticates the device, and the event Authenticated is shared. Otherwise, it shares the NoAuthenticated event.

V. IMPLEMENTATION AND EVALUATION

A. Implementation

For implementing this strategy, Remix IDE is used in the first step. It is an open-source web tool that aids in the testing, debugging and deploying smart contracts and serves as a learning and teaching environment for Ethereum blockchain. The selection of Ethereum is based on the following criteria:

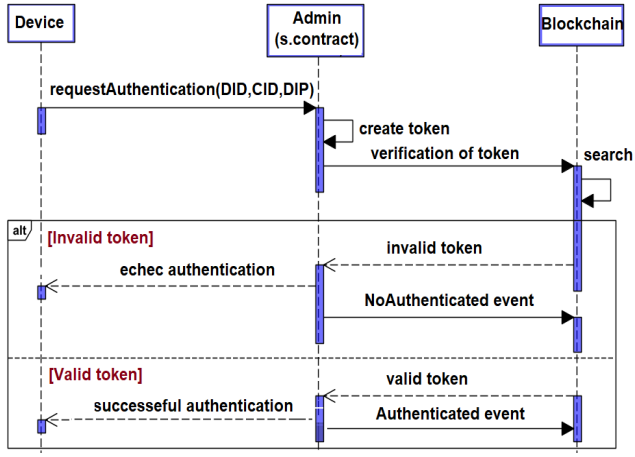


Fig. 5. Sequence diagram for the authentication phase.

1) Firstly, Ethereum is the most popular blockchain platform that simplifies the construction of decentralized applications (DApps).
2) Secondly, it allows creating, deploying and testing smart contracts.
3) Thirdly, it provides secure transactions using a lightweight and robust signature for restricted devices called Elliptic Curves Cryptography (ECC).
4) Fourthly, many Ethereum test networks do not need a faucet or mining; they can reset accounts instantly with a fixed amount of fake Ether for developing and testing the smart contract.
5) Finally, it is widespread and adopted by a large community.

In the following step, ReactJS was used to build a decentralized application (Dapp) that allows interaction between users and Ethereum. The smart contract's compilation and deployment were implemented by Hardhat. Ganache-CLI was used as an Ethereum emulator, which provided accounts with 100 ethers for testing smart contracts. It is very close to a real Ethereum implementation and shows the transactions and blocks created in the blockchain. Then, Metamask created a wallet to manage different accounts and connect the current user to the blockchain. Table I resumes the different tools adopted to validate the suggested mechanism.

TABLE I
DIFFERENT TOOLS USED IN THE IMPLEMENTATION.

Tool	Description
Remix IDE, Hardhat	Smart contracts compilation and deployment
Ganache-cli	Ethereum emulator
ReactJS	development of the front-end application
Metamask	Managing accounts

B. Evaluation

This study does not interest the proposed mechanism's execution time or power consumption because it depends on the type of Ethereum network, the communication protocol

and the node specification. This study's interest concerns the impact of the mechanism on the security of nodes and the network in general. The security mechanism was created to authenticate nodes with limited MANET resources. The blockchain used SHA-256 and ECC algorithms as solid cryptographic proof for data integrity and authentication. All transactions in the blockchain were digitally signed and validated by admin nodes in the network and then were stored and organized in blocks using timestamps and hashes. These blocks were linked together to build a chain. The ECC algorithm is suited to the MANET environment, particularly in terms of key sizes and signature times. Also, it consumes less power and has the same level of security as Rivest Shamir Adleman (RSA)[18].

This part describes how the suggested strategy satisfies the various security requirements and protects against attacks.

- **Identification**: (ID) is required for any node trying to access the network. Each node in a cluster has a unique identifier DID, a unique public address and the CID that can associate it with its unique group.
- **Confidentiality**: This requirement was satisfied by encrypting all nodes' requests to keep the privacy of identification credentials. Each admin in the network has an Ethereum address that includes asymmetric public key pairs. It shares its public key with nodes for encrypting their messages request.
- **Integrity**: Each node uses its private key to generate a signature for each request message, while the admin can use the node's public key to verify the signature. Also, transactions in the blockchain were signed using the admin's private key produced by the Ethereum-supported ECDSA. This algorithm ensures the integrity of messages and the legitimacy of nodes. Besides, the identifiers of nodes are saved using their hash to keep their privacy because hashing is an irreversible function that cannot be converted back into the original information.
- **Non-repudiation**: Private keys sign all data in the system; as a result, sending and receiving parties can never deny ever their executing a transaction.
- **Resistance against attacks**: Each node can only have one identity and only one key pair in this model. The private key associated with this identity should sign all its transactions. The cluster admin must approve all its nodes identities; as a result, an attacker cannot impersonate another node's identity because he always needs its private key. Furthermore, it will be impossible to attack the blockchain if one admin is compromised because services are distributed and duplicated over all network admins and fog nodes.

C. Open issues

The approach proposed in this article is not adapted to real-time applications because registration and authentication times rely on admins availability. This problem is due to the limited storage capacity of the nodes. The blockchain system has a scalability problem that limits its practical use for all nodes in the network; i.e., if the number of nodes increases, the number of transactions increases, and the file size of the blockchain will require a large amount of storage capacity. As a result, only administrators can maintain the blockchain and control access by other nodes. In addition, this approach also relies on the type of blockchain used. According to the consensus protocol, the transactions will be validated only after the time of consensus. This step represents a critical and challenging issue in MANET due to the limited power and computation resources. High financial costs that involve the amount of cryptocurrency spent on transaction fees are another problem in this approach. In conclusion, for blockchain to operate seamlessly in MANETs, more personalized versions are needed that require fewer resources without losing the built-in security and that handle the changing topology of the network more efficiently.

VI. Conclusion and Future work

This paper proposed a decentralized authentication mechanism for allowing secure access to network resources. The suggested technique used the distributed nature and the cryptographic characteristics of blockchain technology. Besides, fog computing was used to assure communication between admins by delivering the update of blockchain anytime nodes are added to the network. Furthermore, this study defined the security requirements to analyze and evaluate the approach's resistance against attacks. The future work will be interested in protecting the admin from flooding attacks that aim to shut down the cluster. An intrusion detection system will be integrated to detect DOS/DDOS by checking the time of sending authentication requests. Suppose a node transmits more than several authentication requests to the admin in a short time. In that case, the admin will consider it a malicious node and immediately terminate any sort of communication with it and send an alert to all the admins in the network.

References

[1] S. D. Tena and C. Rey-Moreno, *Global information society watch 2018, Community networks*, 2018.

[2] A. O. Bang and P. L. Ramteke, "Manet: History, challenges and applications," *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 2, no. 9, pp. 249–251, 2013.

[3] M. Anand and T. Sasikala, "Efficient energy optimization in mobile ad hoc network (manet) using better-quality aodv protocol," *Cluster Computing*, vol. 22, no. 5, pp. 12681–12687, 2019. DOI: 10.1007/s10586-018-1721-2.

[4] N. Sharma and A. Gangal, "Mobile node authentication in manet using enhanced cluster based aucres algorithm," *Far East J. Electron. Commun*, pp. 1–12, 2016. DOI: 10.17654/ECSV3PI16001.

[5] M. Er-Rouidi, H. Moudni, H. Faouzi, H. Mouncif, and A. Merbouha, "Improving performance of mobile ad hoc network using clustering schemes," vol. 6, pp. 69–75, 2017. DOI: 10.11591/IJICT.V6I2.PP69-75.

[6] B. U. I. Khan, R. F. Olanrewaju, F. Anwar, A. R. Najeeb, and M. Yaacob, "A survey on manets: architecture, evolution, applications, security issues and solutions," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 2, pp. 832–842, 2018. DOI: 10.11591/IJEECS.V12.I2.PP832-842.

[7] K. Gupta and P. K. Mittal, "An overview of security in manet," *International Journals of Advanced Research in Computer Science and Software Engineering ISSN*, vol. 7, pp. 2277–3128, 2017. DOI: 10.23956/IJARCSSE/V7I6/0254.

[8] M. Ichaba, "Security threats and solutions in mobile ad hoc networks; a review," *Universal J. Commun. Netw*, vol. 6, no. 2, pp. 7–17, 2018. DOI: 10.13189/ujcn.2018.060201.

[9] M. T. Hammi, B. Hammi, P. Bellot, and A. Serhrouchni, "Bubbles of trust: A decentralized blockchain-based authentication system for iot," *Computers & Security*, vol. 78, pp. 126–142, 2018. DOI: 10.1016/j.cose.2018.06.004.

[10] A. A. Laghari, A. K. Jumani, and R. A. Laghari, "Review and state of art of fog computing," *Archives of Computational Methods in Engineering*, vol. 28, no. 5, pp. 3631–3643, 2021. DOI: 10.1007/S11831-020-09517-Y.

[11] A. Jarjis and G. Kadir, "Blockchain authentication for aodv routing protocol," in *2020 Second International Conference on Blockchain Computing and Applications (BCCA)*. IEEE, 2020. DOI: 10.1109/BCCA50787.2020.9274452, pp. 78–85.

[12] H. Yang, "A study on hierarchical structure and blockchain-based node authentication mechanism in manet," *Convergence Security Journal*, vol. 19, no. 3, pp. 13–19, 2019. DOI:10.33778/kcsa.2019.19.3.013.

[13] H. Garmani, D. AitOmar, M. ElAmrani, M. Baslam, and M. Jourhmane, "Joint beacon power and beacon rate control based on game theoretic approach in vehicular ad hoc networks," *INFOCOMMUNICATIONS JOURNAL*, vol. 13, no. 1, pp. 58–67, 2021. DOI: 10.36244/ICJ.2021.1.7.

[14] A. Maria, V. Pandi, J. D. Lazarus, M. Karuppiah, and M. S. Christo, "Bbaas: Blockchain-based anonymous authentication scheme for providing secure communication in vanets," *Security and Communication Networks*, vol. 2021, pp. 1–11, 2021. DOI: 10.1155/2021/6679882.

[15] K. Khullar, Y. Malhotra, and A. Kumar, "Decentralized and secure communication architecture for fanets using blockchain," *Procedia Computer Science*, vol. 173, pp. 158–170, 2020. DOI: 10.1155/2021/6679882.

[16] B. Chatterjee and H. N. Saha, "Parameter training in manet using artificial neural network." *International Journal of Computer Network & Information Security*, vol. 11, no. 9, 2019. DOI: 10.5815/ijcnis.2019.09.01.

[17] D. Bisen, S. Mishra, and P. Saurabh, "K-means based cluster formation and head selection through artificial neural network in manet," 2021. DOI: 10.21203/RS.3.RS-667651/V1.

[18] T. K. Goyal and V. Sahula, "Lightweight security algorithm for low power iot devices," in *2016 international conference on advances in computing, communications and informatics (ICACCI)*. IEEE, 2016. DOI: 10.1109/ICACCI.2016.7732296, pp. 1725–1729.

Hafida Khalfaoui obtained her B.Sc. in Electronic and Telecommunication Engineering and her M.Sc. in Telecommunication Systems and Computer Networks from Sultan Moulay Slimane University, Beni Mellal, Morocco, in 2017 and 2019, respectively. She is following her Ph.D. in Mathematics and Computer Science at Sultan Moulay Slimane University. Her research interests include computer science and network security.

Abderrazak Farchane received his B.Sc. in Computer Science and Engineering in June 2001 and M.Sc. in Computer Science and Telecommunication from the university of Mohammed V Agdal, Rabat, Morocco, in 2003. He obtained his Ph.D. in Computer Science and Engineering at ENSIAS, Rabat, Morocco. He is currently an Associate Professor of Computer Science in the Polydisciplinary Faculty, Sultan Moulay Slimane University, Morocco. His areas of interest are Information Coding Theory, Cryptography, and Security.

Said Safi received his B.Sc. degree in Electronics from Cadi Ayyad University, Marrakech, Morocco, in 1995. He obtained his M.Sc. and Ph.D. from Chouaib Doukkali University and Cadi Ayyad University in 1997 and 2002, respectively. He is currently a Professor of Science at the Multidisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco. His general interests span the areas of communications and signal processing, estimation, time-series analysis and system identification. Safi has more than 160 publications. His research currently focuses on transmitter and receiver diversity techniques for single and multi-user fading communication channels and on broadband wireless communication systems.

A Comprehensive Survey on the Most Important IPv4aaS IPv6 Transition Technologies, their Implementations and Performance Analysis

Omar D'yab

***Abstract*—As the central public IPv4 address pool has already been exhausted, the deployment of IPv6 has become inevitable. However, the users still require IPv4 Internet access due to some IPv4-only applications. The IPv4aaS (IPv4-as-a-Service) IPv6 transition technologies facilitate that ISPs provide IPv4 service to their customers while using only IPv6 in their access and core networks. This paper discusses the widely used IPv4aaS IPv6 transition technologies in ISP/enterprise networks; we explain their operations, advantages, properties and consider their performances. There are currently many IPv6 transition technologies, nevertheless, in this paper, the five most prominent IPv4aaS IPv6 transition technologies are discussed, namely 464XLAT, Dual-Stack Lite, Lightweight 4over6, MAP-E, and MAP-T. Moreover, the deployment and implementations of these technologies are being analysed and inspected. This paper also overviews the benchmarking methodology for IPv6 transition technologies and surveys several papers that investigated metrics and tools utilized in analysing the performance of different IPv6 transition technologies.**

***Index Terms*—464XLAT, DS-Lite, Lw4o6, MAP-E, MAP-T**

I. INTRODUCTION

WE have already given an overview of the five IPv4aaS technologies, their operation, advantages and disadvantages, as well as their most important implementations [1]. As expected, years ago, the world is now running out of IPv4 addresses. In February 2011, IANA, the global body responsible for managing Internet addresses, distributed the last five "/8" sets of IPv4 Internet addresses to the five regional Internet registries [2]. IPv4 uses a 32-bit addressing scheme, which was thought to be enough to support billions of devices, yet the more devices connected, the more we need IPv6 to solve the problem many predicted.

IPv6 activation is the main solution to the problem of lack of IPv4 addresses. IPv6 is the next generation of Internet Protocol and is designed to replace the existing IPv4 protocol, however, it is still not easy to deploy, and as the network environment needs to be converted from IPv4 to IPv6, especially that IPv4 is still widely used, it may take a long time because of some factors, such as the inability of the IPv4 network devices to be completely replaced.

O. D'yab is with 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: omardyab@hit.bme.hu).

The deployment of the IPv6 protocol around the world is relatively slow, the following may address the possibilities of this issue:

- Service providers do not want to activate IPv6, because there is no demand from subscribers, and subscribers do not request IPv6 because of the lack of content that works on it, hence content providers do not want to activate IPv6 until it becomes a demand from users.
- IPv6 hosting provides a greater number of available internet addresses and many other features, however, ISPs do not yet offer IPv6 services or support many of the features of this version of the IP.
- If one wants to deploy something new into the network, there is an impact on the stability of the network, routers must be upgraded, sometimes firmware must be changed, so an upgrade is needed more often, debugging this software is necessary and it costs extra efforts. IPv6 and IPv4 are incompatible protocols meaning that if one has at least one application that does not support it, then both protocols must be run.

This paper [3] surveys some tools and methods for measuring the deployment of IPv6, grouping them into different categories and comparing them from different aspects, distinguishing sources of data, whether public, private, or restricted, and the extent of the measurement duration, aiming to give an estimation of the IPv6 portion.

There are plenty of IPv6 technologies that have been developed to facilitate the co-operation of the two incompatible versions of IP (IPv4 and IPv6) for different scenarios [4]. One important scenario is, when IPv4 addresses ran out and only IPv6 addresses are being distributed to the clients, but there are still many old servers, which have only IPv4 addresses. A suitable solution for this scenario is the combination of NAT64 [5] and DNS64 [6]. This technology works well with the majority of the generally used client-server network applications [7]; however, there are some applications such as Skype which unable to use IPv6. For this reason, many providers, who would like to forget about IPv4 in the access and

DOI: 10.36244/ICJ.2022.3.5

core network, still must provide IPv4 to the customers, while they use solely IPv6 in their access and core network. It is called "IPv4 as a Service" (IPv4aaS) and there are several solutions were developed for this purpose. The advantages and disadvantages of the five most important IPv4aaS technologies are discussed in the following Internet Draft [8].

The remainder of this paper is organized as follows: Section II introduces the five most important IPv4aaS technologies and their proposed applied systems. Section III deals with their implementations. Section IV gives an introduction about benchmarking methodologies for IPv6 Transition Technologies. Section V is a conclusion of this paper.

II. THE FIVE MOST IMPORTANT IPV4AAS TECHNOLOGIES

A. 464XLAT

IPv6 hosts cannot communicate directly with IPv4 hosts and for this reason, several transitions methods have been developed: 464XLAT (RFC 6877) technology [9] is essentially an extension to NAT64 that provides the IPv4 access by combining stateful (RFC 6146) and stateless translation (RFC 6145).

464XLAT as a combination of stateless NAT64 (RFC 6145) and stateful NAT64 (RFC 6146) provides a lot of benefits [10] such as:

- It is easy to deploy and troubleshoot, using open-source standard technologies and based on RFC.
- It is efficient in terms of using IPv4 at minimum resource requirements and maximum efficiency.
- 464XLAT allows for full functionality and solves IPv4 numbering issues.
- IPv6-only networks are less expensive and simpler to operate, already proven by multi-vendor: Cisco, Juniper and F5.

464XLAT main components as shown in Fig. 1 are:

- CLAT (customer side translator) is a small piece of code that enables the client to have an IPv4 address. It translates 1:1 private IPv4 addresses to global IPv6 addresses, and vice versa [9].
- PLAT provider-side translator translates statefully IPv6 to IPv4 using stateful NAT64, it translates N:1 global IPv6 addresses to public IPv4 addresses and vice versa [9].

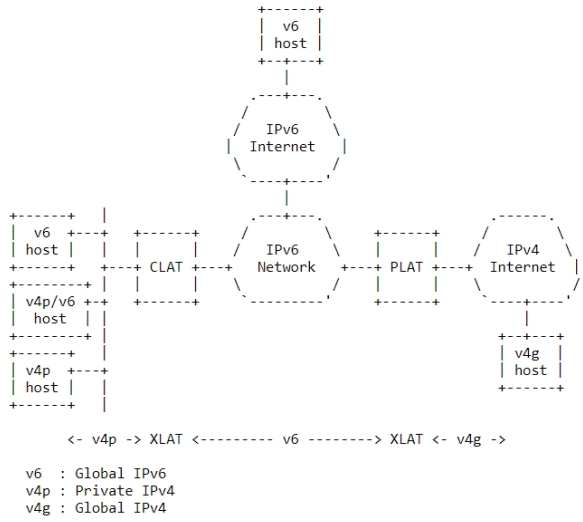


Fig. 1 464XLAT Wireline Network Topology [9]

We have elaborated an example (taken from RFC 6877 [9]) of IPv4/IPv6 address translation on the 464XLAT architecture:

- The IPv4 client with 192.168.1.2/24 private IP address is aiming to access the IPv4 server with 198.51.100.1 public IP address across an IPv6-only network.
- At the CLAT, IP routing is performed and different IPv6 prefixes are used for translation, the CLAT and the PLAT at this stage both know their IPv6 prefixes (the CLAT IPv6 prefix is 2001:8db:aaaa::/96, the PLAT IPv6 prefix is 2001:8db:1234::/96 in our example).
- The CLAT must do the translation process for the IPv4 packet to reach the IPv4 server, which means and as per our example, the destination address will be translated to 2001:db8:1234::198.51.100.1 and the source address will be translated to 2001:db8:aaaa::192.168.1.2. However, for reaching IPv6 hosts, the CLAT function is clearly dispensable.
- At the PLAT, before reaching the IPv4 server, the destination address is being extracted reversely back to its original 198.51.100.1, for the source IP address, 192.0.2.1 was chosen.
- At the server, the packets have successfully reached their destination.

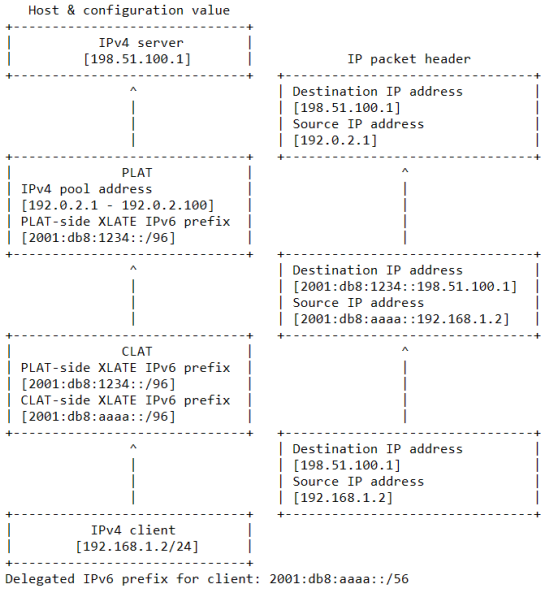


Fig. 2 464XLAT scenario [9]

As shown in Fig. 2, the example mentioned above requires double translation, however, if the client is an IPv6, then a single translation is necessary only at the PLAT, in which single stateful translation is enabled and in conjunction with a configured DNS64, as in RFC6146 in [9], CLAT is no longer needed. The DNS64 server in this case is responsible for constructing and returning a special IPv6 address called IPv4-Embedded IPv6 Address [9].

This solution is similar to NAT64, but the main difference is that the CLAT service needs to be installed on the mobile equipment. For example: Skype is an IPv4 only application, so it does not work with IPv6, the CLAT is to translate Skype clients IPv4 packets into IPv6 packets, the packets are then sent over an IPv6 only network to a NAT64 translator which translates them back into IPv4 and sends the packets to an IPv4 only server (Skype server). 464XLAT have helped a lot of mobile providers with the IPv6 implementations, because customers with 464XLAT can have an IPv6 only connection and still access all IPv4 only applications and content.

This recent paper [11] has been analysing the security aspects of this transition technology using STRIDE and Data Flow Diagram (DFD) methods, observing threats that the PLAT might face.

B. Dual Stack Lite or DS-Lite

DS-Lite stands for dual stack light (dual stack environment is one that has version 4 and version 6 addresses, too), DS-Light combines IP in IP (IPv4-in-IPv6) and Network Address Translation (NAT) technologies and allows IPv4 traffic to be encapsulated into IPv6 [12].

There are mainly two elements -as shown in Fig. 3- of DS-Lite as following:

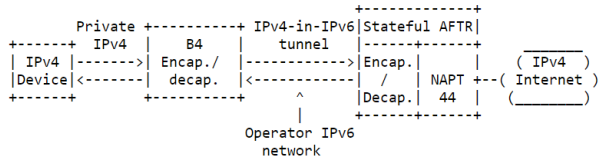


Fig. 3 Overview of the DS-Lite architecture [8]

1. B4, Basic Bridging Broadband element, which encapsulates IPv4 within IPv6, those IPv4 packets will go through an IPv6 network, B4 creates a multipoint-to-point IPv4-in-IPv6 tunnel to an AFTR [8].
2. AFTR, Address Family Transition Router receives the packets handled by the B4 and de-capsulates them. AFTR can reconstruct IPv6 when IPv4 packets come back from the Internet by doing a reverse lookup in the NAT binding table. AFTR is combination of IPv4-in-IPv6 tunnel endpoint and an IPv4-IPv4 NAT implemented on the same node [8].

IANA has defined a well-known range, 192.0.0.0/29 for numbering the interfaces of both B4 and AFTR [12].

As explained in RFC 6333 [12] and shown in Fig. 4, the goal is to carry IPv4 traffic over the IPv6 access and core network. In the case of outbound traffic, the message is first sent to the DS-Lite home router (B4) as “IPv4 datagram 1”, in which it's encapsulated and forwarded to the AFTR as “IPv6 datagram 2”. The AFTR decapsulates the IPv4 datagram from the IPv6 datagram and then the carrier-grade NAT44 is performed, “IPv4 datagram 3” is sent out [12]. In the case of inbound traffic, “IPv4 datagram 3” is received by the AFTR, NAT checks the information of its translation table and changes TCP destination port and sets the IP destination address, then it’s encapsulated into and IPv6 packet and forwarded to the home router B4. B4 decapsulates and extracts the IPv4 datagram and forwards it to the host. However, the packets are being dropped at the AFTR, when addresses are out of range, or the information does not match the NAT table.

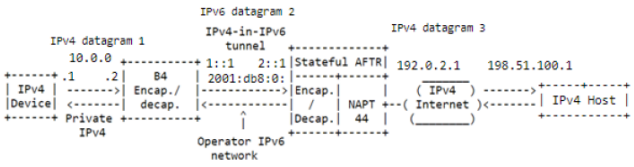


Fig. 4 Inbound and Outbound Datagram [based on 12]

C. Lightweight 4over6

Lightweight 4over6 is a transition mechanism as an extension of DS-lite; it has some of its concepts in providing IPv4 connectivity over IPv6, as well as the following main components:

- Lightweight B4 is the Lightweight Basic Bridging Broadband "lwB4" element that performs NAPT44 and creates a tunnel to a lwAFTR.

- Lightweight AFTR is the Lightweight Address Family Transition Router that is an IPv4-in-IPv6 tunnel.

The Lightweight 4over6 mechanism of sharing the addresses among clients is different from that of DS-Lite, as lw4o6 gives a portion of the port space to each client, the CPE is going to do a NAT and encapsulates the packets into IPv6, and the packets get through the border router. The lwAFTR task is to do a lookup in the binding table, which is a static table, once there is a match, lwAFTR de-capsulates the packets and forwards it to the Internet.

Lightweight 4over6 is a technology that flips the complexity of the dynamic address translation (between the LAN interface and the given public address) back to the client, where every CPE does the address translation and port-based NAT-ting, the difference in lightweight AFTR is that customers share the public IP address and each client gets the same IP address over limited port range to use, which makes this function stateless, where they all share the same binding table.

Lightweight 4over6 is a scalable solution where all routers are configured equally to load balance the traffic, for single flow packets can be distributed, and once the routing updates do not get through an instance fails then it's quickly picked up as another negligible hop and the traffic gets distributed to the other one.

As explained in RFC 7596 [13] and shown in Fig. 5, the following are the working scenarios of lwB4 and lwAFTR:

- LwB4 performs a NAPT44 function once receives an IPv4 packet, encapsulates it with an IPv6 header and forwards it to the lwAFTR as configured, while for the packet coming back from lwAFTR, lwB4 obtains IPv4 packet and performs NAPT44 translation based on the information in its NAPT44 table including the destination and port number, however, and when there is no match within what is configured, whether it's IPv6 packet at the lwAFTR, or its IPv4 packet at the lwB4, in both cases the packet is being discarded [13].

- LwAFTR performs a decapsulation and verification once receives an IPv4-in-IPv6 packet coming from lwB4. Based on the information in the binding table, lwAFTR verifies its source addresses and port, once there is a match the packet is forwarded to the IPv4 destination, otherwise it is discarded [13].

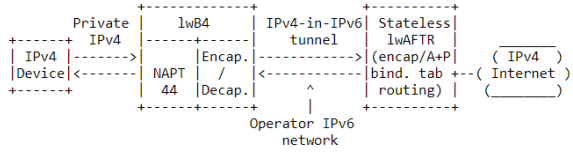


Fig. 5 Overview of the Lightweight 4over6 mechanism [8]

Paper [14] has been dealing with the design of an RFC 8219 compliant software tester for the performance analysis of the lw4o6 transition technology, disclosing the first lw4o6 tester, design considerations and important details of its operational requirements.

D. MAP

MAP stands for mapping of address and port, it's another transition mechanism; it basically maps the addresses and ports of IPv4 into the IPv6 addresses to serve IPv4 connectivity over IPv6 network, where IPv4aaS on top of IPv6 is being delivered using this stateless technology.

MAP (as in Fig. 6) is targeted access customer of broadband service providers, where it allows them to deploy an IPv6 infrastructure, MAP main components are:

- MAP Customer Edge: A home gateway acts as a CE router and provides IPv4/IPv6 stateless translation.

- MAP Border Relay (BR): a router from provider side supports stateless translation.

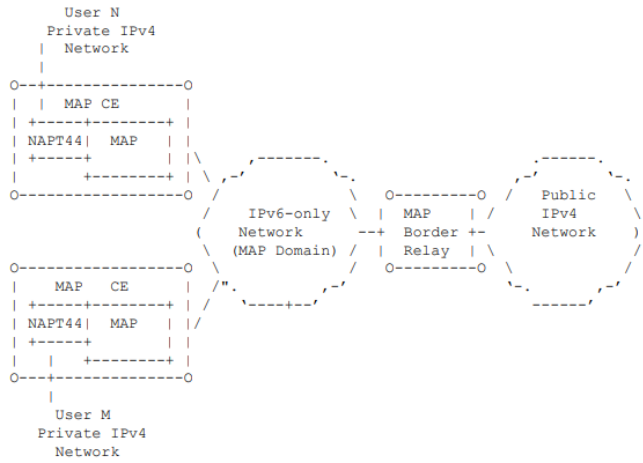


Fig. 6 MAP Network Topology [15]

MAP has two subtypes:

- MAP-T: Mapping of Address and Port using Translation.

- MAP-E. Mapping of Address and Port with Encapsulation.

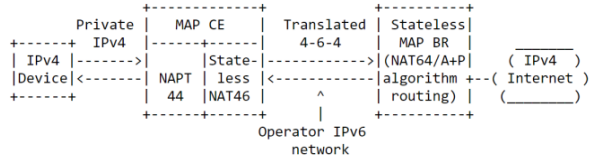


Fig. 7 MAP-T architecture [8]

TABLE I
DIFFERENT IMPLEMENTATIONS OF IPV6 TRANSITION TECHNOLOGIES

Name	*System*	*License*	*Function(s)*	*Technology*
CLATD [18]	Linux	open source	CLAT / SIIT-DC Edge Relay implementation	464XLAT
Android CLAT [19]	Android OS 4.3 jellybean or above.	open source	CLAT services through Wi-Fi connection	464XLAT
Cisco CGv6 [20]	Cisco	Licensed hardware and software	Supports stateless MAP technology to deliver both IPv4 and IPv6 services.	Stateless MAP Technology
Map [21]	Linux and OpenWrt	open-source repository	Supports both MAP-T and MAP-E and can be configured with or without NAPT44 function	MAP-T and MAP-E
SNABB [22]	Linux	open-source software	Has a large binding table with high performance.	Lw4o6
MAEMO [23]	MAEMO (OS2008 version)	licensed and open source	Tunnelling IPv6 through a tunnel broker.	DS-lite
PF [24]	BSD systems	Free software	Filter and manipulate IP packets.	464XLAT
Thunder CGN [25]	A10	Licensed hardware and software	Managing transition technologies, enabling providers to smoothly extend IPv4 connectivity and transition to IPv6	DS-Lite, lw4o6, MAP-T and MAP-E
Jool SIIT/NAT64 [27]	Linux	open-source software	BR as PLAT is stateful NAT64 and CLAT is an SIIT. High availability across Jool instances.	Stateful 464XLAT expected to support MAP-T
TAYGA[28]	Linux-based	open-source software	TAYGA is fast, flexible, and secure implementation.	Stateless NAT64
BIG-IP (CGNAT) [29]	F5	Licensed hardware and software	Stateful translation	464XLAT
Cisco ASR 9000[30]	Cisco ISM	Licensed hardware and software	ISM provides scalability in delivering services which supports CGN.	DS-Lite, Stateful NAT64, (MAP-T)
ASAMAP Vyatta [31]	Linux	Open source	Stateless address auto-configuration.	MAP-E, MAP-T, DS-Lite and 464XLAT
FD.io VPP [32]	Linux	Open source	Vector processing graph	Stateful NAT64, MAP-E, MAP-T and lw4o6.

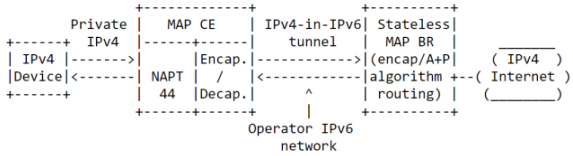


Fig. 8 MAP-E architecture [8]

MAP-T works as follows. When a CE (Customer Edge) device receives a packet that is destined to the public Internet, performs two transformations:

1. It does NAPT but with limited set of ports, the source port is being replaced, the source IP address is also being replaced with public IP address, while the destination address and port number remain the same [16].
2. Then it translates the IPv4 header into an IPv6 header using a stateless NAT46 translation [16].

Then the CE forwards the IPv6 packet to the MAP BR (Border Relay) device. BR performs just the inverse of the second translation (that is, a stateless NAT64) and forwards the resulting IPv4 packet to the public Internet. (Fig. 7).

MAP-E operates similarly to MAP-T, but it uses encapsulation and de-encapsulation instead of stateless NAT46 and stateless NAT64, respectively. (Fig. 8).

Their operation determined by various mapping rules, (Basic Mapping Rule, Forwarding Mapping Rule, Default Mapping Rule). All the details can be found in their RFCs.

The main benefit of this technology is that it is stateless at the center of the network, where no additional hardware is required even with the growth of the traffic. MAP-T [17] is one of two transform modes of the parent technology MAP, which aims to transport IPv4 over an IPv6 domain, while MAP-E [15] uses encapsulation, similarly to DS-Lite, where an IPv4 Packet is prepended with an IPv6 header and transported across the network, MAP-T uses IPv4 and IPv6 stateless translation, so the header translation as opposed to encapsulation. With MAP-T, the IPv4 addresses are embedded within the corresponding IPv6 address.

III. IMPLEMENTATIONS

There are many implementations for the most IPv4aaS (IPv4-as-a-Service) technologies, most of the implementations are free open source and they are usually preferred. **Table I** provides a summary for IPv6 transition technologies implementations.

A. CLATD

CLATD [18] - a CLAT / SIIT-DC Edge Relay implementation for Linux is free software available to implement the CLAT component of the 464XLAT network.

B. Android CLAT

Android CLAT [19] is an open -source application already installed for any Android OS 4.3 jellybean or above. This solution relies on the routing table in order to separate traffic, this implementation does not support IPv6 content only; it was mainly designed to offer CLAT services through Wi-Fi connection.

C. Cisco CGv6

Cisco CGv6 [20] supports stateless MAP technology to deliver both IPv4 and IPv6 services more efficiently on a high scale at a lower cost and less latency. Machine to Machine services is an advantage of this technology.

D. Map

Map [21] is an open-source repository supports both MAP-T and MAP-E and can be configured with or without NAPT44 function. This software is also compatible with AFTR of DS-Lite and NAT64 (stateful and stateless), this CPE implementation runs on Linux and OpenWrt.

E. SNABB

SNABB [22] is fully compatible open-source software with Lightweight 4over6 that has a large binding table with high performance, it consists mainly three elements: APP (Filter, lwAFTR), Programs and links to connect applications together. The 3rd version of SNABB supports YANG IETF. The 4-th version of SNABB supports RSS (Receive Side Scaling) multiprocessor and YANG Alarm Module.

F. MAEMO

MAEMO, this implementation requires an N810 Nokia tablet as hardware and one of the supported software's listed in [23], the idea behind this is tunnelling IPv6 through a tunnel broker sending and receiving IPv4 Packets.

G. OpenBSD Packet Filter

PF [24] is an abbreviation of Packet Filter subsystem which is a free software released with OpenBSD 3.0 in 2001, and contained a rather complete implementation of packet filtering, including network address translation (NAT64) [25]. Packet filter controls the flow of the packets on interfaces, it differentiates whether its TCP or UDP, it recognizes the source and destination IP addresses or layer 3 addresses.

H. Thunder CGN:

The Thunder CGN [26] is a scalable secure implementation thorough hardware and software solutions provided by A10, managing transition technologies, and enabling providers to smoothly extend IPv4 connectivity and transition to IPv6, that is including DS-Lite, lw4o6, MAP-T, and MAP-E transition technologies. However, it's worth mentioning that it's not free and prices vary from device to another that usually comes with extra yearly service and maintenance cost.

I. Jool SIIT/NAT64

Jool [27] is an open-source software and reflects an implementation of 464XLAT transition technology in which PLAT is a stateful NAT64, whereas CLAT is an SIIT. One of the most important features is the high availability across Jool instances. It's worth mentioning that Jool is presently in late development for MAP-T transition technology.

J. TAYGA

TAYGA [28] is a Linux-based stateless NAT64 implementation, packets are exchanged with the help of TUN driver. TAYGA is also: fast, flexible, compatible, secure, and most importantly it is free, however, it could not offer stateful solution. It is usually combined with iptables (stateful NAT44 for Linux) to implement a stateful NAT64 solution.

K. F5 BIG-IP Carrier-Grade NAT (CGNAT)

Widely deployed, provides scalable and high-performance network, F5 [29] implemented 464XLAT transition mechanism to deliver IPv4 and IPv6 connectivity.

L. Cisco ASR 9000

Cisco ASR 9000 [30] is an Integrated Service Module (ISM) that provides scalability in delivering services which supports Carrier Grade Network Address Translation (NAT) or CGN, Dual-Stack Lite, Stateful NAT64, and Mapping of Address and Port Translation (MAP-T), in which multiple can coexist on multiple ISMs with a lot of major features and benefits, yet this option is costly.

M. ASAMAP Vyatta

Vyatta [31] is a system that supports stateless configuration with the help of SLAAC protocol which has a host and a router as main components; however, DHCPv6 is not supported by Vyatta. ASAMAP Vyatta supports MAP-E, MAP-T, DS-Lite and 464XLAT.

N. FD.io VPP

Vector Packet Processing [32] is the heart of FD.io; it is the open-source version of Cisco's Vector Packet Processing (VPP) technology, the VPP is faster than current technologies; as it processes through vector processing graph at extreme performance, VPP supports Stateful NAT64, MAP-E, MAP-T and lw4o6.

IV. BENCHMARKING METHODOLOGY FOR IPV6 TRANSITION TECHNOLOGIES

In this section, a short introduction is given to the benchmarking methodology for IPv6 transition technologies, furthermore, a summary of several papers that investigated the performance of IPv6 transition technologies are added.

The goal of RFC 8219 [33] is to provide meaningful and unbiased results by measuring performance characteristics of various IPv6 transition technologies. There are two well-known RFCs about the benchmarking methodology for network interconnect devices: RFC 2544, which is theoretically IP version independent, but relies on IPv4 and the specificities of IPv6 are addressed in RFC 5280, in which IPv6 transition

technologies excluded. On the other hand, RFC 8219 is a new one handling IPv6 transition technologies.

RFC 8219 [33] helped in classifying a massive number of transition technologies into a much smaller number of categories. Model of production network was used to achieve this purpose, in which we have two different domains IPvX and IPvY and called domain A and domain B respectively. The scenarios are as following:

- Single translation: Domain A needs to be translated to be able to communicate to Domain B. Stateful NAT64, SIIT, and IVI are transition technologies that can solve this problem.
- Double translation: There are three domains, in which domain A and B are version 4 specific and the core domain is version 6 specific. 464XLAT and MAP-T both are examples of this production model.
- Encapsulation: Any version can be encapsulated/decapsulated into another version, DS-Lite, and MAP-E both are examples of this production model.

DNS64 is an additional protocol and does not transfer data packets, which is just required to support NAT64, thus it does not fit into any of these categories mentioned above, and it is to be dealt with separately. The problem with these scenarios, the packet format and size can be changed during the process of translation, that is why those methods must be calibrated, two test setups are defined to solve this issue:

Test Setup 1, for Single-Translation where the DUT (device under test) is translating the IPvX packets into IPvY packets as shown in Fig. 9. As for Test Setup 2 (Double-Translation), there are two DUTs. One DUT implements the reverse operation of the other one: if one DUT does encapsulation the other is decapsulation, if one is translating 4 to 6, the other is translating 6 to 4 as in Fig. 10. In case of testing as peers together we may use RFC 2544 [33] tester, however, if there is asymmetric behaviour, then we will not be able to observe it and in this case, we should use Test Setup 1.

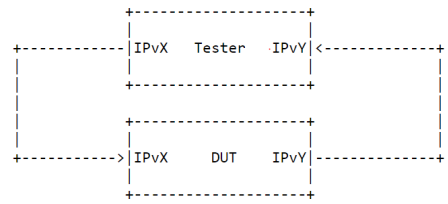


Fig. 9 Single DUT Test Setup for benchmarking[33]

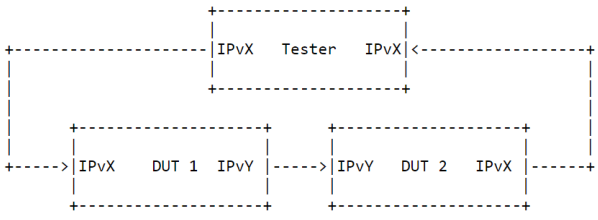


Fig. 10 Dual DUT Test Setup for benchmarking [33]

RFC 8219 recommended important benchmarking measurement tests, each with different requirements, such as: Throughput, Latency, Frame Loss Rate, Packet Delay Variation.

For double translation (either in stateless or stateful) same tests can be used, as well as different test setups for example dual and single DUT, the latter is recommended to observe asymmetric behaviour. Similar procedures for encapsulation, however packets that are encapsulated must be provided to prepare a tester. For stateless tests, UDP is used, for stateful tests, (all RFC 3511) TCP is used.

As for DNS64 benchmarking, based on RFC8219 in [33], the tester implements two different logical functions: version 6 only-client and an authoritative DNS server, it can be implemented by two different devices or similar devices. The test traffic of the DNS64 benchmarking is as following (shown in Fig. 11):

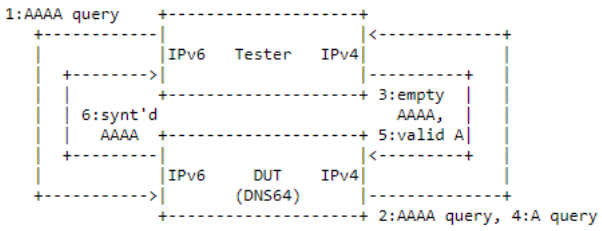


Fig. 11 DNS64 DUT Test Setup for benchmarking [33]

1. The IPv6-only client sends “AAAA” record query (IPv6 address) for a domain name.
2. The DNS64 server receives the request, sends “AAAA” record query for the given domain name to the authoritative DNS server.
3. If there is no such “AAAA” record, then an empty “AAAA” record is being returned.
4. The DNS64 server sends another query asking for “A” record of the same domain name.
5. The authoritative DNS System replies with a valid “A” record (IPv4 address).
6. The DNS64 server synthesizes an *IPv4-embedded in IPv6 address*, which is returned to the IPv6-only client.

When the DNS64 server implements caching and there is a cache hit, then step 1 is followed by step 6, and for message 1 the answer is message 6. The goal here again is to determine performance (requests processed per second), in other words, the *rate* between messages sent and received. A test should last at least 60 seconds and *timeout* should be not more than 1 second. However, the measurement may be influenced by the tasks executed by the device in the background, so the *median* of the results of the repetitive measurements is calculated to get a better understanding of the performance.

As mentioned before, while the IPv6 demands in solving the IP address shortage is expanding, there are several papers experimenting transition technologies of IPv6, utilizing

TABLE II
SUMMARY OF TRANSITION TECHNOLOGIES COVERED BY EACH REFERENCE

IPv6 Transition	*[39]*	*[41]*	*[42]*	*[43]*	*[44]*	*[45]*	*[46]*
464XLAT	✓						
MAP-E	✓						
MAP-T	✓						
Dual Stack		✓	✓		✓	✓	✓
DS-lite	✓			✓			
4over6				✓			✓
6to4		✓					
4rd				✓			

different implementations and network environments, each depending on different factor such as: type of test, configuration, technology, topology and more as discussed in [34].

Moving toward benchmarking methodologies and tools, siitperf is "an RFC 8219 compliant SIIT (stateless NAT64) tester written in C++ using DPDK" [35]. The accuracy of siitperf is examined in [36], by structuring an error model and discussing what could influence the measurements and cause inaccuracy, observing the effect of Ethernet flow-control, concluding that calibrating siitperf is a necessity.

There is a novel Internet Draft about an RFC 8219 compliant methodology for benchmarking stateful NATxy (x, y are in {4, 6}) gateways [37]. Its proposed benchmarking procedures are implemented as an extension of siitperf for stateful tests [38].

IPv6NET is a network evaluation testbed built as a combination of closed and open environments [39]. For the closed environment, ASAMAP Vyatta implementation was used, multiple IPv6 transition technologies were considered including MAP (both MAP-E and MAP-T), 464XLAT, and DS-Lite. The traffic was generated by a distributed Internet traffic generator (D-ITG), two functions were performed in each computer, one to send (ITGSend) and one to receive (ITGRecv). During the process and based on the recommendation of RFC5180, frame size and frame rates were considered. They monitored the following network performance metrics: Round-trip-delay, jitter, packet loss, and throughput. Overall MAP-E achieved the best performance in a closed environment.

As for the open environment, three associated operational feasibility metrics were introduced: configuration, troubleshooting, and application capabilities. Inspired by [40] three configuration task groups were organized associated with a task code: initial setup, reconfiguration, and confirmation. They concluded that applications capability was running smoothly for all four technologies, in regard to configuration capability, an addition of a guided self-configuration would be beneficial, for troubleshooting capabilities improvements are needed.

Based on the empirical results, it was found that MAP-E was more feasible compared to other transition technologies. MAP-T and 464XLAT had a better performance in terms of latency as translation-based technology, on the other hand, MAP-E and Ds-Lite had a better performance in terms of throughput as encapsulation-based technology, IPv6NET has shown that it has a high level of repeatability, one flaw in IPv6NET is the lack of control data.

This research [41] examined three IPv4/IPv6 transition mechanisms of dual stack, the manual tunnel, and the 6to4 automatic tunnel through three metrics: delay, delay variation, and packet loss by using the Optimized Network Engineering Tool (OPNET) Modeler simulator, on a real-time application (video conferencing). The performance results show that dual-stack had better performance than the others with the lowest average delay, dual-stack has shown efficiency in terms of packet delay variation and with a lower loss rate. Hence, the Dual-Stack was the best. Consequently, both tunnelling mechanisms results were deficient, and this is due to the encapsulation and decapsulation processes.

Network analysis was performed in [42] for three different transition technologies, namely: Dual-Stack, 6in4, and NAT-PT, they were compared using Cisco packet tracer, and for this purpose, three main performance metrics were taken into consideration: Round Trip Delay Time (RTT), Bandwidth and Throughput. The results have shown that NAT-PT due to its high latency and low throughput, was neglected, dual-stack had better performance and was preferred.

Chuangchunsong et al. [43] compared delay time, and reliability for four different transition scenarios: 4over6, DS-lite, 4rd: NAT Centralization, and 4rd: NAT Distribution by using OPNET. Results have shown that both 4rd have high performance and high reliability, but both are inflexible in IP address allocation. 4over6 has also shown a similar result to 4rd, but with lower performance compared to other transition mechanisms. On the other hand, DS-Lite only on inter-communication has shown relatively high performance and reliability, but also high flexibility. Conversely, for intra-communication, the DS-Lite has low performance and low reliability, but rather has less complexity and higher compatibility compared to other mechanisms.

This empirical measurement in [44] conducted a performance study of IPv6 and IPv4 through dual-stack sites from all over the world, using performance metrics: connectivity, throughput, packet loss, hop count, and round-trip time (RTT), considering different regions and times. Compared with IPv6, IPv4 had higher latency and lower throughput with intangible improvements since 2004, IPv6, however, had lower packet loss rate and better connectivity. The average hop count of the IPv6 network is very similar to that of IPv4.

Proving that dual stack is the best technique, achieving better performance in solving the limitations of IPv4, [45] proposed a methodology of four phases: Build & Design network, Statistics, Simulator, and the results of the analysis. The analysis and based on three different scenarios (IPv4, IPv6, and

Dual-Stack), were compared using Riverbed simulator, evaluating five performance metrics*:* Delay, Traffic dropped, Jitter, Packet delay, and CPU Utilization. The results have shown that Dual-stack surpassed IPv4 and gave a better performance.

This paper [46] proposed comparing 4over6 and dual stack by tracking seven nodes and measuring the average time, the results have shown that the average time for Dual stack had a higher performance by over 17%.

Table II concludes a summary of those different references that has been analysed for different IPv6 transition technologies.

V. CONCLUSION

The future of communications and networks, IPv6 transition technologies are the key to solving the shortage and limitations of IPv4, the question remains how both would be compatible and coexist together effectively, paving the way to develop transition technologies based on different metrics, and factors such as throughput, jitter, packet loss, delay and so forth. Consequently, this paper has fully disclosed the necessity of deploying IPv6 technologies, explained their most promising IPv4aaS transition technologies, namely, 464XLAT, DS-Lite, lw4over6, MAP-E, MAP-T, their operation mechanism, advantages, and disadvantages, analysed and examined existing solutions, collected their most important implementation cases, and gave an introduction about benchmarking methodologies, surveyed some papers, considering and analysing their outcomes. To summarize, 464XLAT is easy to deploy and efficient in using minimum resources, dual-stack lite can work with Interoperability; allowing IPv4 and IPv6 content to reach hosts simultaneously, MAP is a stateless scalable transition technology. Contrarily, 464XLAT needs additional service on the client or on the network, Lw4o6, MAP-E, and MAP-T require more planning and re-provisioning, 464XLAT and Ds-lite may have scalability issues being a stateful and keeping per-flow mapping information between IPv4 and IPv6 addresses.

VI. ACKNOWLEDGEMENTS

The author thanks his supervisor; Dr. Gabor Lencse, Budapest University of Technology & Economics, for helping him by reviewing and commenting the manuscript.

REFERENCES

[1] O. D'yab, "An overview of the most important implementations of IPv4aaS technologies", *AIS 2019*, University of Óbuda, Székesfehérvár, Hungary (2019) pp. 143–146.

[2] IANA IPv4 Address Space Registry. [Online]. Available: https://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xhtml

[3] V. Iró, G. Lencse, "Survey on Measurement Methods for IPv6 Deployment", *Acta Technica Jaurinensis*, vol 13, no 2, pp. 112–130, May 27, 2020, DOI: 10.14513/actatechjaur.v13.n2.544

[4] G. Lencse and Y. Kadobayashi, "Comprehensive survey of IPv6 transition technologies: A subjective classification for security analysis", *IEICE Transactions on Communications*, vol. E102-B, no.10, pp. 2021–2035. DOI: 10.1587/transcom.2018EBR0002

[5] M. Bagnulo, P. Matthews, I. V. Beijnum, Stateful NAT64: "Network address and protocol translation from IPv6 clients to IPv4 servers", IETF RFC 6146 (2011). DOI: 10.17487/RFC6146

[6] M. Bagnulo, A. Sullivan et al., "DNS64: DNS extensions for network address translation from IPv6 clients to IPv4 servers", IETF RFC 6147 (2011), DOI: 10.17487/RFC6147

[7] S. Répás, T. Hajas, G. Lencse, "Application compatibility of the NAT64 IPv6 transition technology", in *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, 2015, pp. 1–7. DOI: 10.1109/TSP.2015.7296383

[8] G. Lencse, J. P. Martinez et al., "Pros and cons of IPv6 transition technologies for IPv4aaS", approved Internet Draft, May 23, 2022. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-ietf-v6ops-transition-comparison-03

[9] M. Mawatari, M. Kawashima, C. Byrne, "464XLAT: Combination of stateful and stateless translation", IETF RFC 6877 (2013). DOI: 10.17487/RFC6877

[10] UK IPv6 Council: 464xlat for mobile operators. [Online]. Available: https://www.ipv6.org.uk/wp-content/uploads/2018/11/Nick-Heatley_BT_EE_464xlat_UKv6Council_20180925.pdf

[11] A. Al-Azzawi and G. Lencse, "Identification of the Possible Security Issues of the 464XLAT IPv6 Transition Technology", *Infocommunications Journal*, vol. 13, no. 4, pp. 10–18, December 2021, DOI: 10.36244/ICJ.2021.4.2

[12] A. Durand, R. Droms et al., "Dual-stack lite broadband deployments following IPv4 exhaustion", IETF RFC 6333 (2011). DOI: 10.17487/RFC6333

[13] Y. Cui, Q. Sun et al., "Lightweight 4over6: An extension to the dual-stack lite architecture", IETF RFC 7596 (2015). DOI: 10.17487/RFC7596

[14] A. Al-hamadani and G. Lencse, "Design of a Software Tester for Benchmarking Lightweight 4over6 Devices", *44th International Conference on Telecommunications and Signal Processing (TSP 2021)*, Brno, Czech Republic, July 26-28, 2021, pp. 157–161, DOI: 10.1109/TSP52935.2021.9522607

[15] E. O. Troan, W. Dec et al., "Mapping of address and port with encapsulation (MAP-E)", IETF RFC 7597 (2015). DOI: 10.17487/RFC7597

[16] MAP - Solving IPv6 Deployment and IPv4 Address Exhaustion without Stateful CGN: CKN TechAdvantage Webinar. [Online]. Available: https://community.cisco.com/t5/networking-knowledge-base/ckn-techadvantage-webinar-map-solving-ipv6-deployment-and-ipv4/ta-p/3639138

[17] X. Li, C. Bao, W. Dec (ed), O. Troan, S. Matsushima, T. Murakami, "Mapping of address and port using translation (MAP-T)", IETF RFC 7599, July 2015. DOI: 10.17487/RFC7599

[18] Clatd - a CLAT / SIIT-DC Edge Relay implementation for Linux. [Online]. Available: https://github.com/toreanderson/clatd

[19] What is Android CLAT. [Online]. Available: https://dan.drown.org/android/clat

[20] Carrier Grade IPv6 over Integrated Services Module (ISM). [Online]. Available: https://www.cisco.com/c/en/us/td/docs/routers/asr9000/software/asr9k_r5-2/cg-nat/configuration/guide/b_cgnat_cg52xasr9k/b_cgnat_cg52xasr9k_chapter_010.html

[21] MAP, source code of MAP CE. [Online]. Available: https://github.com/cernet/MAP

[22] Lightweight4over6, one step further dual-stack lite networks. [Online]. Available: https://ripe76.ripe.net/presentations/105-lw4o6-ripe.pdf

[23] DS-Lite Host Profile for MAEMO (OS2008 version), HW and SW Requirements. [Online]. Available: http://ds-lite.garage.maemo.org/

[24] P. N. M. Hansteen, the Book of PF: A No-Nonsense Guide to the OpenBSD Firewall, 2nd ed., San Francisco: No Starch Press, 2010. ISBN: 978-1593272746.

[25] OpenBSD manual page server. [Online]. Available: https://man.openbsd.org/pf.4

[26] Extend IPv4 Investment and Transition from IPv4 to IPv6 Seamlessly. [Online]. Available: https://www.a10networks.com/wp-content/uploads/A10-SB-19104-EN.pdf

[27] Introduction to IPv4/IPv6 Translation. [Online]. Available: https://www.jool.mx/en/intro-xlat.html#ipv4ipv6-translation.html

[28] TAYGA, Simple, no-fuss NAT64 for Linux. [Online]. Available: http://www.litech.org/tayga/

[29] Using DS-Lite with CGNAT. [Online]. Available: https://techdocs.f5.com/kb/en-us/products/big-ip_ltm/manuals/product/cgn-implementations-11-5-0/14.html

[30] Cisco ASR 9000 Series Aggregation Services Routers. [Online]. Available: https://www.cisco.com/c/en/us/products/routers/asr-9000-series-aggregation-services-routers/index.html

[31] Guide to IPv6 Support. [Online]. Available: https://docs.huihoo.com/vyatta/6.0/Vyatta_IPv6_R6.0_v03.pdf

[32] The Vector Packet Processor (VPP). [Online]. Available: https://s3-docs.fd.io/vpp/22.02/

[33] M. Georgescu, L. Pislaru, G. Lencse, "Benchmarking Methodology for IPv6 transition technologies", IETF RFC 8219 (2017). DOI: 10.17487/RFC8219

[34] A.T.H.Al-hamadani,G.Lencse,"Asurveyontheperformanceanalysis of IPv6 transition technologies", *Acta Technica Jaurinensis*, vol. 14, no. 2, pp. 186–211, May 26, 2021. DOI: 10.14513/actatechjaur.00577

[35] G. Lencse, "Design and Implementation of a Software Tester for Benchmarking Stateless NAT64 Gateways", IEICE Transactions on Communications, vol. E104-B, no. 2, pp. 128–140. February 1, 2021. DOI: 10.1587/transcom.2019EBN0010

[36] G. Lencse, "Checking the Accuracy of Siitperf", Infocommunications Journal, vol. 13, no. 2, pp. 2–9, June 2021, DOI: 10.36244/ICJ.2021.2.1

[37] G. Lencse, K. Shima, "Benchmarking methodology for stateful NATxy gateways using RFC 4814 pseudorandom port numbers", active Internet Draft, May 17, 2021, draft-lencse-bmwg-benchmarking-stateful-00

[38] G. Lencse, "Design and implementation of a software tester for benchmarking stateful NATxy gateways: Theory and practice of extending siitperf for stateful tests", *Computer Communications*, vol. 172, no. 1, pp. 75–88, August 1, 2022, DOI: 10.1016/j.comcom.2022.05.028

[39] M. Georgescu, H. Hazeyama et al., "Empirical analysis of IPv6 transition technologies using the IPv6 Network Evaluation Testbed", *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol 2, no. 2, (2015). DOI: 10.4108/inis.2.2.e1

[40] D. Harrington, "Guidelines for Considering Operations and Management of New Protocols and Protocol Extensions." RFC 5706 (Informational), Nov. 2009.

[41] G. Altangerel, E. Tsogbaatar, D. Yamkhin, "Performance analysis on IPv6 transition technologies and transition method", in *2016 11th International Forum on Strategic Technology (IFOST)*, 2016, pp. 465–469. DOI: 10.1109/IFOST.2016.7884155

[42] J. L. Shah, J. Parvez, "An examination of next generation IP migration techniques: Constraints and evaluation", in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 2014, pp. 776–781. DOI: 10.1109/ICCICCT.2014.6993064

[43] N. Chuangchunsong, S. Kamolphiwong, T. Kamolphiwong, R. Elz and P. Pongpaibool, "Performance evaluation of IPv4/IPv6 transition mechanisms: IPv4-in-IPv6 tunneling techniques", *The International Conference on Information Networking 2014 (ICOIN2014)*, 2014, pp. 238–243. DOI: 10.1109/ICOIN.2014.6799698

[44] Li, K.-H.; Wong, K.-Y. "Empirical Analysis of IPv4 and IPv6 Networks through Dual-Stack Sites". *Information 2021*, 12, 246. DOI: 10.3390/info12060246

[45] M. R. A. Ahmed and S. S. A. Shaikhedris, "Network Migration and Performance Analysis of IPv4 and IPv6", *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, 2021, pp. 1–6. DOI: 10.1109/ICCCEEE49695.2021.9429664

[46] Lu, T.T., Wu, C.Y., Lin, W.Y., Chen, H.P., Hsueh, K.P. (2017). "Comparison of IPv4-over-IPv6 (4over6) and Dual Stack Technologies in Dynamic Configuration for IPv4/IPv6 Address". In: Pan, JS., Tsai, PW., Huang, HC. (eds) Advances in Intelligent Information Hiding and Multimedia Signal Processing. Smart Innovation, Systems and Technologies, vol 63. Springer, Cham. DOI: 10.1007/978-3-319-50209-0_32

Omar D'yab received his MSc in Computer Science Engineering from Óbuda University, Budapest, Hungary in 2019. He is now a PhD student at the Budapest University of Technology and Economics, Department of Networked Systems and Services in the MédiaNets Laboratory.

His research area covers IPv6 transition technologies for IPv4aaS and their performance analysis.

Towards Implementing a Software Tester for Benchmarking MAP-T Devices

Ahmed Al-hamadani, and Gábor Lencse

***Abstract*—Several IPv6 transition technologies have been designed and developed over the past few years to accelerate the full adoption of the IPv6 address pool. To make things more organized, the Benchmarking Working Group of IETF has standardized a comprehensive benchmarking methodology for these technologies in its RFC 8219. The Mapping of Address and Port using Translation (MAP-T) is one of the most important transition technologies that belong to the double translation category in RFC 8219. This paper aims at presenting our progress towards implementing the world's first RFC 8219 compliant Tester for the MAP-T devices, more specifically, the MAP-T Customer Edge (CE) and the MAP-T Border Relay (BR). As part of the work of this paper, we presented a typical design for the Tester, followed by a discussion about the operational requirements, the scope of measurements, and some design considerations. Then, we installed a testbed for one of the MAP-T implementations, called Jool, and showed the results of the testbed. And finally, we ended up with a brief description of the MAP-T test program and its configuration parameters in case of testing the BR device.**

***Index Terms*—Benchmarking, Border Relay, Customer Edge, IPv6 transition technologies, MAP-T**

I. INTRODUCTION

THE public IPv4 address pool of IANA was depleted in 2011 [1]. However, the full deployment of IPv6 is taking too much time because it faces several significant challenges. As a softening solution to the problem, many transition technologies have been proposed and developed over the past few years to allow IPv4 and IPv6 to coexist and work with other for some time before totally excluding IPv4 from the Internet and fully adopting IPv6 [2]. The IETF's RFC 8219 [3] categorized these transition technologies into four groups, namely, dual stack, single translation, double translation, and encapsulation technologies, and it defines an organized comprehensive methodology for their benchmarking.

In dual stack [4], both IPv4 and IPv6 stacks are included in the network nodes, and the RFC 2544 [5] and RFC 5180 [6] compliant measurement tools can sufficiently benchmark dual stack devices.

A. Al-hamadani 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: alhamadani@hit.bme.hu).

G. Lencse 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: lencse@hit.bme.hu). Submitted: June 3, 2022.

The single translation technologies translate the packets traveling from an IPv4 domain to an IPv6 domain and vice versa (i.e., reverse translation) at the edge between these two domains [3]. The single Device Under Test (DUT) setup of RFC 8219 [3] can help in benchmarking the devices of this type of technologies. Siitperf [7], an RFC 8219 compliant Tester, is an example tool that uses this type of setup to benchmark the Stateless IP/ICMP Translation (SIIT), also called stateless NAT64 single translation technology.

The double translation technologies translate the packets traveling from one IPvX domain to another IPvX domain through IPvY core domain, where X and Y are part of the set {4, 6} and $X \neq Y$. The first translation is taken place at the edge between the first IPvX domain and the IPvY core domain, while the second translation is taken place between the IPvY core domain and the second IPvX domain. However, the reply packets will be reversely double translated in the opposite direction. [3] The devices of this type of technologies (e.g. the CLAT and PLAT devices of the 464XLAT technology, or the CE and BR devices of the MAP-T technology) can be benchmarked using the dual DUT setup of RFC 8219 [3], where both interconnecting devices that are located at the two edges are benchmarked together with the same setup. However, when one of the devices forms a bottleneck, this could hide several potential asymmetries. Therefore, RFC 8219 recommends additional separate benchmarking for each one of the two devices using the single DUT setup.

The encapsulation technologies encapsulate the packets traveling from an IPvX domain by an IPvY header at the edge between the IPvX domain and the IPvY core domain and then decapsulate them at the edge between the IPvY core domain and another IPvX domain before reaching their destination. However, the reply packets will experience reverse encapsulation/decapsulation processes in the opposite direction. [3] The devices of this type of technologies (e.g., the B4 and AFTR devices of the DS-Lite technology, or the lwB4 and lwAFTR devices of the lw4o6 technology) can be benchmarked in the same way as benchmarking double translation technologies. [3]

MAP-T technology [8] is one of the most important transition technologies that belong to the double-translation category, and it is also considered an IPv4-as-a-Service (IPv4aaS) technology, which can give the network operators the option to continue providing their customers with IPv4 services while deploying only IPv6 in their core and access network [9]. This paper aims at progressing the implementation of the world's first RFC 8219 compliant MAP-T Tester.

DOI: 10.36244/ICJ.2022.3.6

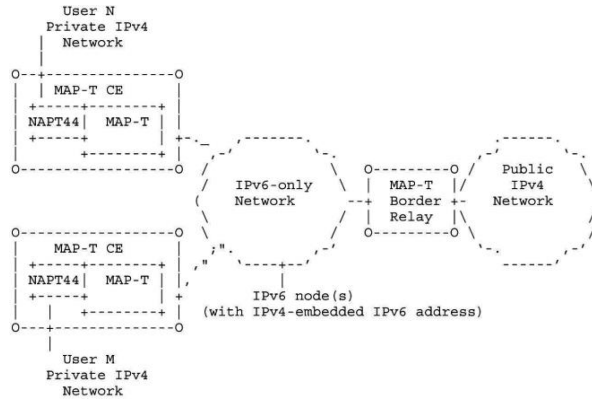


Fig. 1. MAP-T architecture [8]

The remainder of this paper is organized as follows. Section 2 introduces the MAP-T technology. Section 3 discloses the basic operational requirements for the MAP-T Tester based on the recommendations of RFC 8219, followed by a brief description of the scope of measurements. Section 4 discusses the most important design considerations for the Tester. Section 5 gives details on the installation procedure of a testbed for a MAP-T implementation, along with its results. Section 6 describes the MAP-T test programs. Section 7 summarizes further configuration parameters needed by the test programs. Section 8 presents some future plans. Section 9 concludes the paper.

II. Mapping of Address and Port Using Translation (MAP-T) Technology

MAP-T technology [8] helps in accelerating the adoption of IPv6 as it enables service providers to run IPv6-only devices for their operator network while keeping customers run applications that use socket APIs and literal IPv4 addresses. Thus, this technology is considered an *IPv4aaS* technology [10].

MAP-T can be compared to the Mapping of Address and Port using Encapsulation (MAP-E) technology [11] because it also uses mapping rules, but instead of using encapsulation, it deploys a stateless double NAT64 translation, the same as 464XLAT [12] does to perform its function. This procedure targets removing the encapsulation overhead and helps in making IPv4 traffic and IPv6 traffic be treated as similar as possible.

To accomplish its task, MAP-T deploys two types of devices: the Customer Edge (CE) device which is located at the edge of the customer's private network, and the Border Relay (BR) device which is located at the edge of the native IPv4 Internet. Fig. 1 shows the architecture of MAP-T. The CE device connects the privately addressed IPv4 users to the IPv6 network by performing two operations. First, it executes a stateful NAPT to translate the end user's private IPv4 address and port number into the public IPv4 address and port number range assigned to the subscriber, as described by RFC2663 [13]. Then, it performs a stateless NAT46 to map the public IPv4 address and port to IPv6 address. However, the IPv4 destination address will be manipulated differently by embedding it in an IPv6 address, as described by RFC6052 [14], that uses a specific IPv6 prefix to traverse the IPv6 network. When the BR device receives the IPv6 packet, it first performs a similar stateless NAT64 translation to get back the previous public IPv4 address and port and then routes the IPv4 packet into the public IPv4 network. Finally, the CE and BR devices will use the same rules to translate back and forward the reply packets to the user.

One or more CEs and BRs can be connected through an IPv6 network to form a single MAP domain which can be managed by a set of configuration parameters referred to as MAP rules. The CEs and BRs that belong to the same MAP domain will share the same MAP Rules. The service provider can utilize single or multiple MAP domains. The MAP rules are classified as Basic Mapping Rule (BMR), Forwarding Mapping Rule (FMR), and Default Mapping Rule (DMR).

The BMR specifies how the MAP address that all CE devices must share should be built and thus allows assembling MAP addresses out of IPv4 addresses, and vice versa. This rule can be identified by three fields, namely, the Rule IPv6 prefix, the IPv4 prefix assigned for CEs, and the length of Embedded Address (EA) bits. The rule IPv6 prefix is an IPv6 prefix reserved by the service provider for the MAP rule or CE usage, and this means that all CEs belonging to the same MAP domain must use the same rule IPv6 prefix. The IPv4 prefix is the public IPv4 prefix that is assigned by the service provider for the MAP rule or CE usage. The EA bits identify which CE is being used and represent two concatenated subfields: the IPv4 suffix and the Port Set ID (PSID). The PSID specifies which port range is assigned to the CE in case of sharing the IPv4 address among multiple CEs. The format of the MAP address is shown in Fig. 2. The MAP address configured by BMR will represent a customer IPv4 client behind one of the CEs and it is considered the translated form of the source address of the egress packets and the translated form of the destination address of the ingress packets.

The FMR maintains a table of a bunch of BMRs, which will be used by the BR to manage its serviced MAP domain. The BR will use the FMR to translate the source address of the packets received from one of the CEs and to translate the destination address of the packets to be forwarded to one of the CEs.

The DMR is used to form the IPv4-embedded IPv6 addresses for those destinations located outside the MAP domain by adding an IPv6 prefix, which is provisioned from its corresponding BR to the IPv4 public address of the destination. Any CE can use this rule to install an IPv4 default route and to mask the addresses of the devices on the IPv4 internet behind the BR. More precisely, the CEs and BRs will use this rule to translate the destination address of the egress packets and the source address of the ingress packets.

To make things clearer, we refer to the example given by Jool's MAP-T summary [15]. Suppose we have 256 public IP addresses within subnet 192.0.2.0/24 and the number of ports is distributed evenly among customers with 2048 ports each. Thus, we can serve 256*65536/2048= 8192 customers (i.e., we have 65536/2048= 32 port sets within each IP). The PSID will identify each port set. For instance, PSID 0 will identify the set with ports 0-2047, PSID 1 will identify the set with ports 2048-4095, and so on. Note here that the first port set will also include the well-known ports 0-1023. Then, each CE can be identified

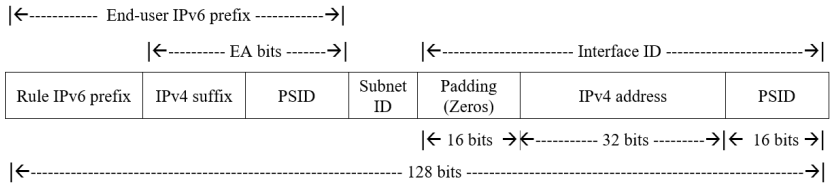


Fig. 2. MAP address format [15]

by the EA-bits, which are composed of a concatenation of the suffix of the public IPv4 address that is assigned to the CE and the PSID. For our example, the suffix is 8-bits because we have 256 IPs and the PSID is 5-bits because we have 32 port sets in each IP. So, the EA-bits are 13-bits. The End-user IPv6 prefix is the only important part of the MAP address, and it can be said that all other fields are essentially cosmetic fields. As the format of the MAP address, shown in Fig. 2, the end-user IPv6 prefix is composed of the Rule IPv6 prefix and the EA-bits. All CEs within the same MAP domain will share the same rule IPv6 prefix. Let us assume it as 2001:db8:ce::/51 for our MAP domain's CEs. Now, the BMR would be the triplet (2001:db8:ce::/51, 192.0.2.0/24, 13) according to its definition stated earlier. An example of the MAP address configured by this BMR, which represents a customer IPv4 client with IPv4 socket 192.0.2.2:2050, could be 2001:db8:ce:41::c000:202:1. Here, 41 represents the EA-bits (00000010 | 00001). The first part (i.e., 00000010) represents the suffix **2** in the last octet of the IPv4 address 192.0.2.**2**, and the second part (i.e., 00001) represents PSID 1 which refers to the second port set because port 2050 is in that port set. The c000:202 is the hexadecimal representation of the public IPv4 address. Whereas the last digit 1 represents the PSID. Now, this address will represent the source address of all IPv6 packets traveling through the related customer's CE to the connected BR via the IPv6 network. While the destination address of these packets is configured by DMR by masking the IPv4 address of the destination, assume it for example 203.0.113.56, by the default IPv6 prefix defined by DMR, assume it for example the NAT64 well-known prefix 64:ff9b::/96. Thus, the IPv4-embedded IPv6 destination address of these packets will be as 64:ff9b::203.0.113.56. However, at the CE, the source address and destination address of the reply packets coming from the BR will adhere to the same rules, but now in a reverse manner, to get their original public IPv4 form. That is, it gets the source address by using DMR and the destination address by using BMR.

On the other side, the source address of the IPv6 packets received by the BR from its connected CE will be translated to its original public IPv4 address according to FMR, which is in our example the same as BMR, while the destination address will be translated according to DMR. Again, the source address and the destination address of the reply packets going to the related CE will be translated according to the same rules, but in a reverse manner, to get their IPv6 form. That is, DMR for the source address and FMR for the destination address.

III. OPERATIONAL REQUIREMENTS AND SCOPE DECISIONS

To emulate, to some extent, the conditions of a production network environment, it is crucial to benchmark IPv6 transition technologies under various operational conditions [3]. This section gives a high-level overview of the operational requirements of the Tester and discloses some considerations about its scope decision.

A. Test and Traffic Setup

As stated earlier, MAP-T is considered a double translation technology. This means that its test setup should follow, in general, the dual DUT test setup as recommended by RFC 8219 [3]. This test setup is briefly described in section 4.2 of this RFC and is exhibited in Fig. 3. The CE should act as DUT1, and the BR should act as DUT 2. However, both have some asymmetries in their behavior. Thus, there should be a separate test based on the single DUT test setup for each one of them according to the recommendations of RFC 8219 [3]. This test setup is briefly described in section 4.1 of this RFC, and it is depicted in Fig. 4. Here, we should also note that the Tester should have translation capabilities as both DUTs to accomplish its task.

There are several test specifications that both test setups should adhere to during testing. These specifications comply with RFC 8219, and they can be summarized as follows:

- Bidirectional traffic must be generated in the tests even though the arrows of the traffic flow are shown as unidirectional in the test setups in Fig. 3 and Fig. 4. However, unidirectional traffic can also be used to get fine-grained performance test results.
- The two IP versions will be used, and they are expressed as IPvX and IPvY, where X=4 and Y=6 in the dual DUT test setup, while X and Y are part of the set {4, 6} and X ≠ Y in the single DUT test setup, that is their exact value depends on what DUT is to be tested (i.e., X is 4 and Y is 6 if the DUT is CE and vice versa if the DUT is BR).
- Although various media types can act as connection media, the tests will rely only on Ethernet.
- Based on RFC 8219, the following frame sizes should be used: 64, 128, 256, 512, 768, 1024, 1280, 1518, 1522, 2048, 4096, 8192, and 9216. In addition, RFC 8219 recommends setting all interfaces of the DUTs and the Tester to use the larger MTU between the physical NICs and virtual translation interfaces to avoid any frame loss due to the MTU mismatch between the two types of interfaces. More specifically, the recommended value to be used is the minimum IPv6 MTU size (i.e., 1280 bytes) plus the translation overhead.

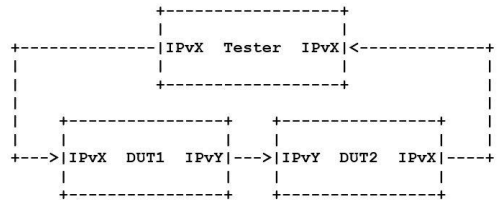


Fig. 3. Dual DUT Test Setup [3]

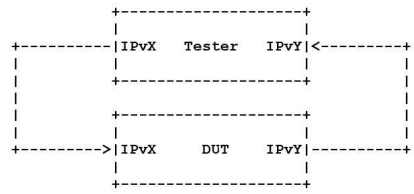


Fig. 4. Single DUT Test Setup [3]

- The IPv4 addresses should be selected based on the recommendations of section 12 of RFC 2544 [5], while the IPv6 addresses should be selected based on the recommendations of section 5 of RFC 5180 [6].
- UDP should be employed as the transport layer protocol for tests.

It is also required that non-translated or native IPv6 traffic, also called "background traffic", must be used besides the translated traffic, and different proportions of the two types must be generated. To accomplish a well-organized testing procedure and to get more precise testing results, we decided to run three types of testing:

1) CE Testing

This test adheres to the single DUT test setup. Here, the CE device will act as the DUT. Both the Tester and the DUT must have one interface configured as IPv4 and another one configured as IPv6. The test starts by sending IPv4 packets from the Tester via its IPv4 interface. The DUT should receive these packets from its IPv4 interface, then performs stateful NAPT translation on their private source address, then translates them into IPv6 packets after forming the IPv6 source address based on BMR (i.e., this address is also called the MAP address) and the IPv6 destination address based on DMR, and then forwards the translated packets via its IPv6 interface to the Tester. When the Tester receives the IPv6 packets from its IPv6 interface, it should be able to get its original IPv4 traffic after translating back the source address based on the same BMR as the CE's one and the destination address based on the same DMR as the CE's one.

To continue testing in the reverse direction, the Tester should first pre-generate IPv6 packets that adhere to the DMR rule (for source address) and BMR rule (for destination address) ahead of starting the test to enhance the speed of the Tester and then send them via its IPv6 interface. When the DUT receives these packets via its IPv6 interface, it should be able to translate them back into IPv4 packets using the same DMR and BMR rules (i.e., to get the IPv4 source address and the public destination address respectively), performs stateful NAPT translation according to the information it has in its local NAPT table, and then forwards the resulted IPv4 packets via its IPv4 interface. Finally, the Tester should, in turn, receive these packets from its IPv4 interface and they should be the same as the original ones before sending them the first time.

2) BR Testing

This test adheres to the single DUT test setup, too. But, here, the BR will act as the DUT. Similarly, both the DUT and the Tester must have one interface configured as IPv4 and another one configured as IPv6. Here, the Tester should first pre-generate templates of IPv6 packets which adhere to BMR (for source address) and DMR (for destination address) ahead of starting the test to enhance the speed of the Tester. The test starts by sending the pre-generated IPv6 packets via its IPv6 interface to the DUT. Once the DUT receives the IPv6 packets from its IPv6 interface, it should be able to translate them back into IPv4 packets after applying the appropriate FMR (i.e., it should select the same BMR as the Tester's one) to get the IPv4 source address and the DMR to get the IPv4 destination address, and then it forwards the IPv4 packets from its IPv4 interface. The Tester should, in turn, receive these packets from its IPv4 interface and they should have similar IPv4 traffic.

To continue testing in the reverse direction, the Tester should first send the IPv4 packets via its IPv4 interface. When the DUT receives these packets, it should translate them into IPv6 packets after applying the DMR to get the IPv6 source address and the appropriate FMR to get the IPv6 destination address, and then it forwards the IPv6 packets via its IPv6 interface. Finally, when the Tester receives these packets via its IPv6 interface, it should translate them back into IPv4 packets after applying the DMR and BMR rules (i.e., to get the IPv4 source and destination addresses respectively) and get its original IPv4 traffic before sending it the first time.

The challenge in the BR Testing is that the Tester should simulate a high number of CEs connected to the BR. This is usually what happens in the production network. Therefore, the Tester should have some approach and configuration settings to make this possible.

3) Overall Testing

Testing both devices under the dual DUT test setup as required by RFC 8219, where the CE device will act as the DUT1 and the BR device will act as the DUT2, can be done with the help of an existing testing tool, which is the stateful branch of `siitperf` [16]. This testing tool follows the benchmarking methodology described in [17].

B. Scope of Measurements

RFC 8219 requires performing different types of performance measurements. Practically, some of these benchmarking measurements are implemented by some existing RFC 2544 testers, while others are either omitted or seldom used such as back-to-back frames, system recovery, and reset. The first two measurement tests require that the Tester must be able to send at the maximum possible rate of the media, which could not be necessarily met by the deployed devices that run the Tester. The latter measurement test requires causing or sensing a DUT reset, and this means we need supplementary hardware. Thus, only those measurement tests that we see as important will be supported by our Tester. In this section, we introduce them and their requirements.

1) Throughput

RFC 8219 reuses the RFC 2544's definition of throughput as it is "the fastest rate at which the count of test frames transmitted by the DUT is equal to the number of test frames sent to it by the test equipment" [5]. That is, no frame loss should occur. Here, the Tester must be able to send frames at any given rate for some period and then count the number of the sent and received frames in that period. It is possible to use a binary search algorithm to find the fastest possible rate and consequently apply this measurement effectively. We should also note that RFC 8219 specified several frame sizes to perform this test.

2) Latency

This measurement is practically based on throughput. Here, a stream of frames, whose duration is at least 120 seconds, should be sent at a particular frame size from the Tester through the targeted DUT at the calculated throughput rate. Some of the frames should be tagged. At least 500 tagged frames should be recognized after 60 seconds from the start of the transmission. For each tagged frame, two timestamps should be recorded, one at the time of fully sending the frame and another at the time of receiving it. The latency will represent the difference value of the two timestamps. Then, the test should calculate two important quantities, the Typical Latency (TL), which represents the median value of the latencies of at least 500 tagged frames, and the Worst-Case Latency (WCL), which represents the 99.9^{th} percentile of them. To get more accurate results, the test must be repeated at least 20 times, and it should eventually record the median value of all TLs and the median value of all WCLs.

3) Packet Delay Variation (PDV)

This measurement includes two variations of tests, Packet Delay Variation (PDV) and Inter Packet Delay Variation (IPDV), both are significantly important, especially for real-time applications. However, our Tester will primarily focus only on calculating PDV as RFC 8219 marks it as recommended, while it marks IPDV as optional for fine-grained analysis of delay variation. In this test, also, a stream of frames should be sent at a particular frame size from the Tester through the targeted DUT at the calculated throughput rate. But the duration of the stream should rather be at least 60 seconds and the one-way latency value of *all* frames should be calculated. Thus, the PDV will represent the difference value between the 99.9^{th} percentile and the minimum delay value in the stream. Similarly, the test must be run at least 20 times and the final recorded value will be calculated from the median of all calculated PDVs.

4) Frame Loss Rate (FLR)

This measurement is similar to the throughput and is also done by sending a stream of frames at some rate through the targeted DUT, but here, we will count the number of received frames by the Tester and then calculate the FLR as in (1):

$$FLR = ((sent - received)/sent) * 100\%. \quad (1)$$

To run this test, a different frame rate will be used at each new trial, starting from the maximum frame rate of the media, and then decreased by some percentage (typically 10%) at each new trial. The test will finish once we find two consecutive trials in which no frames are lost.

IV. DESIGN CONSIDERATIONS

What follows are some important design factors that should be considered when implementing the MAP-T Tester. They are similar to those of the work done in our earlier paper [2] and in the work done in [7], and they follow the same approach of them.

A. Integration or Separation

Building a fully integrated Tester, which can automatically run all measurement tests, can be a desirable solution in case we plan to perform a commodity Tester for routine tests. However, our Tester is intended to be used primarily for research purposes. Thus, we aim to design a more flexible measurement tool that gives the ability to extract some interesting intermediate results by performing only certain important subtasks. For this purpose, the primary functions of our Tester will be implemented using high-performance programs executed by modifiable bash scripts that accept input parameters instead of built-in constants in the programs (e.g., 60 seconds duration or 500 timestamps) as allowed by RFC 8219.

B. Software Architecture and Hardware Requirements

Generally, RFC 8219 requires generating bidirectional traffic in the tests. To build a simple yet efficient program structure, two thread pairs will be used, one for the forward direction (i.e., for processing the packets from the client to the server in Section II) and another for the reverse direction (i.e., for processing the reply packets). Each thread pair consists of a thread for sending and another for receiving. In case each thread will be executed by a single CPU core, this means that we need four CPU cores for communication processes plus an extra CPU core to run the main program. It should also be said that, at any given time, either one of the two directions might be inactive. In addition, each one of the two DUTs and the Tester must have two NICs for testing and an optional extra one for network communication.

C. Input and Output

Building the program structure in such a way that supports separation could help the shell scripts to run the programs multiple times with two forms of parameters, static and dynamic. Those parameters whose values do not change during the execution (e.g., IP addresses, MAC addresses, and so on) will be statically provided in a configuration file. In contrast, those parameters whose values may change during the execution (e.g., frame size, frame rate, and so on) will be provided as command-line arguments.

The results that the shell scripts need to make some decisions during the execution should be printed out on the standard output to be used for further processing. On the other hand, those results which are big or have no longer been used by the shell script should be stored in an output file.

V. MAP-T TESTBED

Before going further with the implementation of the MAP-T Tester, we built a testbed for one of the implementations of MAP-T called Jool [15] to make sure it is working as expected and to check the operation of the implemented Tester on a valid MAP-T implementation. This testbed is shown in Fig. 5, which uses the same IP addresses used in the example of [15]. This example is also discussed in Section II. The testbed is installed using a workstation with the following specifications: Intel(R)

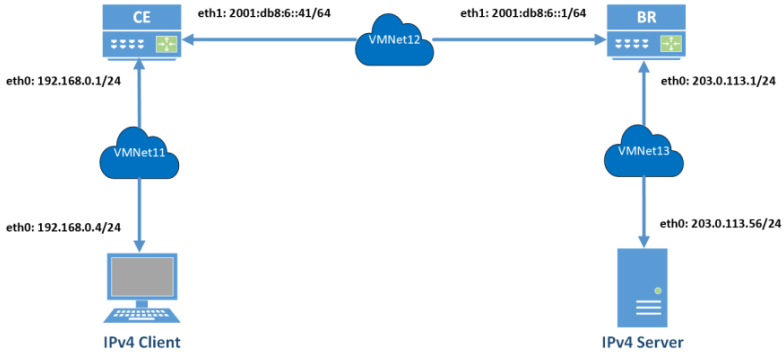


Fig. 5. MAP-T Testbed

Core(TM) i7-3612QM CPU @ 2.10GHz CPU, 8.00 GB RAM:, 1TB HDD, and Windows 10 64-bit Operating System.

Each network node in the testbed is created as a virtual machine using VMware with Debian 10 installed on a single-core processor and supplied with 1GB of RAM and 40GB of Hard Disk. The network nodes are connected via three virtual networks as follows:

- VMNet11 connects the IPv4 Client's eth0 interface to the CE's eth0 interface and it is IPv4 only.
- VMNet12 connects the CE's eth1 interface to the BR's eth1 interface and it is IPv6 only.
- VMNet13 connects the BR's eth0 interface to the IPv4 Server's eth0 interface and it is IPv4 only.

TABLE I shows the Debian and VMware network settings used for each one of the virtual machines.

Furthermore, two shell scripts are written, one is executed at the CE and the other is executed at the BR. Most of the code of the shell scripts follows the configuration steps mentioned in [18]. The **`CE-script.sh`** and **`BR-script.sh`** are available on GitHub [19].

Then, the HTTP service is activated at the IPv4 server by running this Linux command:

```
root@ipv4server# service apache2 start
```

To test the functionality of the network including the CE and the BR, an HTTP request is sent from the IPv4 client to the IPv4 server via this Linux command:

```
root@ipv4client# wget http://203.0.113.56/index.html
```

Consequently, the IPv4 client received the `index.html` page successfully, and this is the result when the "`head`" Linux command was run:

```
root@ipv4client# head -2 index.html

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0
Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">
```

Moreover, when the traffic was captured at eth1 of the CE, as shown in Fig. 6, and eth0 of the BR, as shown in Fig. 7, using tcpdump, it showed proper translations for the HTTP packets.

This gives the result that the Jool's MAP-T implementation of the CE and the BR is working properly and could be tested by our MAP-T Tester.

VI. Description of Test Programs

Three test programs are intended to implement the different four measurements: throughput, latency, PDV, and FLR. These programs are the **maptperf-tp**, which measures the throughput as well as the FLR, the **maptperf-lat**, which measures the latency, and the **maptperf-pdv**, which measures the PDV. The design and implementation of these programs are inspired by the approach followed in [7]. The three programs use the following common parameters, which can be inserted as command-line arguments:

- **IPv6 frame size**: The size of the frames must be according to what is mentioned in Section III.A. The IPv4 frames will automatically be 20 bytes shorter.

TABLE I
Debian and VMware Network Settings

VM Name	Linux Settings			VMware Settings		
	eth0	eth1	eth2	eth0	eth1	eth2
IPv4 Client	Static 192.168.0.4/24	DHCP	N/A	VMNet11	NAT	N/A
CE	Static 192.168.0.1/24	Static 2001:db8:6::41/64	DHCP	VMNet11	VMNet12	NAT
BR	Static 203.0.113.1/24	Static 2001:db8:6::1/64	DHCP	VMNet13	VMNet12	NAT
IPv4 Server	Static 203.0.113.56/24	DHCP	N/A	VMNet13	NAT	N/A

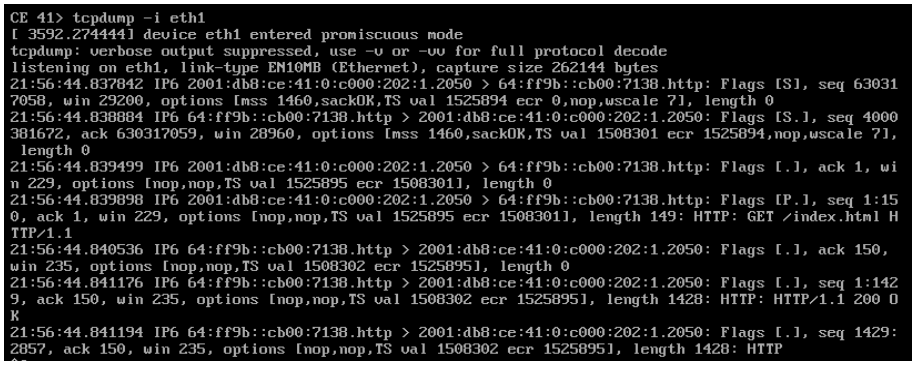


Fig. 6. CE's tcpdump traffic capture at eth1

- **frame rate**: the rate at which the frames will be sent, and it is calculated as frames per second.
- **test duration**: It must be 1-3600 seconds.
- **stream timeout**: the Tester must stop receiving frames when this timeout elapses after sending them completely. This parameter could be compared to the 2000 milliseconds "after sending timeout" recommended by RFC 2544 in its section 23 and complied with RFC 8219 recommendations.
- **n**: a relatively prime number to m, which both help in specifying the proportions of foreground and background frames.
- **m**: a relatively prime number to **n**, which represents the number of foreground frames. Thus, the background traffic will be (**n-m**) frames.

In addition to the abovementioned common parameters, **maptperf-lat** has two further ones:

- **first tagged delay**: The time required to be spent before the first tagged frame can be sent. It must be 0-3600 seconds.
- **tagged**: The number of tagged frames. It must be 1-50,000.

While **maptperf-pdv** has this further one:

- **frame timeout**: In case the value of this parameter is greater than 0 milliseconds, then the delay of every single frame will be checked against it. Then, the frame will be considered "lost" if its delay exceeded the value of this parameter. However, if the value of this parameter is 0, then, the **maptperf-pdv** will calculate PDV as described by RFC 8219.

What follows is the description of each one of these programs:

A. maptperf-tp

This test program sends a stream of frames to the DUT for **test duration** seconds and continues receiving them simultaneously from the DUT until the **stream timeout** elapses. Then, it reports the number of sent frames and the number of received frames for each one of the active directions (i.e. forward and reverse (one of them could be disabled)). The throughput test could be passed if and only if the number of sent frames and the number of received frames are equal for the active directions. However, the FLR could, then, be easily calculated as in (1) (see Section III .B.4).

B. maptperf-lat

This test program also sends a stream of frames to the DUT for **test duration** seconds, but some of these frames are tagged. The first tagged frame will not be sent until the **first tagged delay** elapses. The selected frames to be tagged should be exactly equal to the **tagged** parameter value and should be identified according to the uniform time distribution and sent during the remaining test time (i.e., **test duration** – **first tagged delay**). The test program must continue receiving frames from the DUT until the **stream timeout** expires. For each tagged frame, the test program must report the time at which it completes sending the entire frame and the time at which it finishes receiving it completely. Any tagged frame could be

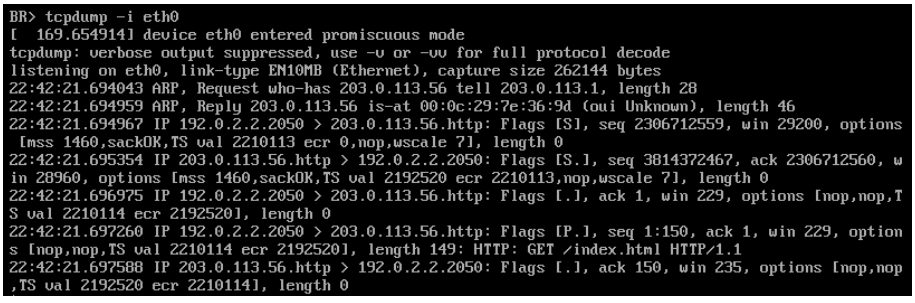


Fig. 7. BR's tcpdump traffic capture at eth0

considered "lost" if not received within a time equal to **test duration - first tagged delay + stream timeout**. Then, the TL and WCL values could be calculated with the help of all reported sending and receiving time values of the tagged frames, as described in Section III.B.2, for the active directions (i.e., forward and reverse).

C. maptperf-pdv

This test program also sends a stream of frames to the DUT for **test duration** seconds, but no tagging will be made here, instead, every frame sent must be recognized by a unique identifier. Similarly, the test program must continue receiving frames from the DUT until the **stream timeout** expires. For *all* frames, the test program must report the time at which it completes sending the entire frame and the time at which it finishes receiving it completely. Then, the test program checks the value of the **frame timeout** parameter. If it is 0, then, it will calculate the PDV as stated in Section III.B.3, for the active directions (i.e., forward and reverse), and any frame will be considered "lost" if not received within a time equal to **test duration + stream timeout**. If the value of the **frame timeout** parameter is greater than 0, then the test program will not calculate PDV, rather it will use the **frame timeout** as a frame loss specifier. That is, if the delay of any single frame exceeds its value, then it will be considered "lost". This gives a more precise approach to measuring throughput and FLR, as recommended in [20]. But the performance penalty would be greater as recording and dealing with timestamps could consume more memory and CPU cycles. Consequently, the **maptperf-pdv** can act as a dual test program. So, it will act as either a PDV tester if the value of the **frame timeout** is 0, or a precise throughput and FLR tester if the value of the **frame timeout** is greater than 0.

VII. FURTHER CONFIGURATION PARAMETERS

Besides the configuration parameters that can be provided to the test programs as command line arguments as described in Section IV, the test programs must also be supplied with some other parameters whose values could not be changed during the execution of the test. These static parameters along with their values are recorded into a configuration file, which will then be used by the test programs during their execution.

In this paper, we will focus on those parameters that must be recorded into the configuration file of the BR testing, as we see that, in practice, a high number of CEs are used together with a single BR, and the scalability of the system will mainly rely on the scalability of the BR itself. What follows is a summary describing them briefly:

A. Basic parameters

The following parameters are the basic ones:

- **Tester-L-IPv6**: The IPv6 address of the left-side interface of the Tester. It should be an address in the same subnet as that of the DUT-L-IPv6. However, it will not represent the source address of the test packets in the forward direction, instead, the source address will be formed by the BMR as the Tester will simulate many CEs. To make things simple and efficient, we deliberately skipped the NAPT function of the CEs and the IPv4 client settings. In addition, the Tester will pseudorandomize the BMR's EA-bits value depending on its length to get various (IPv4 suffix + PSID) values, each of which representing a different CE device. The BMR's IPv4 Prefix, however, is the same for all generated public IPv4 addresses of the CEs.
- **Tester-R-IPv4**: The IPv4 address of the right-side interface of the Tester. It should be an address in the same subnet as that of the DUT-R-IPv4, as Tester will simulate an IPv4 server. This address will also represent the destination IPv4 address for the test packets in the forward direction.
- **Tester-R-IPv6**: The IPv6 address that will be used by the Tester's interface for forwarding background (i.e., non-translated) traffic.
- **DUT-L-IPv6**: The IPv6 address that is currently assigned to the left-side interface of the DUT.
- **DUT-R-IPv4**: The IPv4 address that is currently assigned to the right-side interface of the DUT.
- **DUT-R-IPv6:** The IPv6 address that will be used by the DUT for forwarding background (i.e., non-translated) traffic.
- **Tester-L-MAC**: The MAC address of the left-side interface of the Tester.
- **Tester-R-MAC**: The MAC address of the right-side interface of the Tester.
- **DUT-L-MAC**: The MAC address of the left-side interface of the DUT.
- **DUT-R-MAC**: The MAC address of the right-side interface of the DUT.

The following other basic parameters specify the range of the destination port numbers in the forward direction and the range of the source port numbers in the reverse direction. They are the same as those of the extension of Siitperf [21], which implemented a random port feature originally pointed to by RFC 4814 [22]. There is no specific restriction about the values of these ranges. So, one could use the entire port space, that is 0-65535, as stated in [21]. There are also some other settings recommended by RFC 4814 [22]. We also follow the same approach of [21] to pseudorandomly generate port numbers from these ranges:

- **FW-dport-min**: The lowest number in the range of destination port numbers that can be used by the test packets in the forward direction.
- **FW-dport-max**: The highest number in the range of destination port numbers that can be used by the test packets in the forward direction.
- **RV-sport-min**: The lowest number in the range of source port numbers that can be used by the test packets in the reverse direction.
- **RV-sport-max**: The highest number in the range of source port numbers that can be used by the test packets in the reverse direction.

It may be noticed that the parameters of the source port range in the forward direction and the parameters of the destination port range in the reverse direction are not included with the BR testing parameters because the ports must be

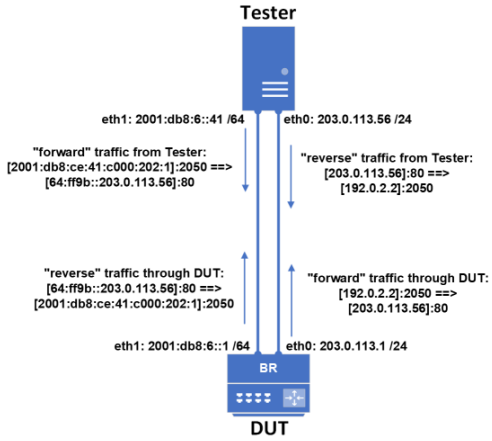


Fig. 8. Translated traffic flow during benchmarking MAP-T BR

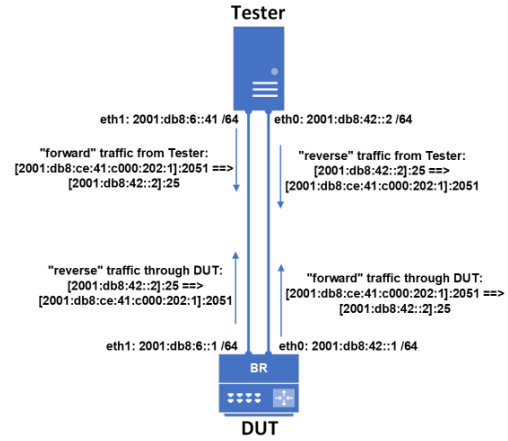


Fig. 9. Background traffic flow during benchmarking MAP-T BR

selected from a specific port set that is assigned to the simulated CE. Therefore, they will be computed by the Tester program and not prerecorded in the configuration file.

The pseudorandomized PSID will determine which port set will be used by the Tester (i.e. the simulated CE). The PSID value can be extracted from the first right "X" bits of the pseudorandomized EA-bits, where X = EA-Len – the IPv4 suffix length. The IPv4 suffix length can be easily determined because the BMR-IPv4-Prefix is set in the configuration file with its length, for example, /24. So, the IPv4 suffix length will equal 32 (i.e., IPv4 address length) minus the BMR-IPv4-Prefix length. Now, to know the range of port numbers in the port set identified by that PSID, we first must know how many port sets are there (we denote this as Y), that is Y = 2 ^ X, and then how many ports can be used in each port set (we denote this as Z), that is Z = (2^16 / Y), where 16 is the number of bits of any port address. Then, the FW-sport-min and the RV-dport-min will be equal to (PSID x Z), while the FW-sport-max and the RV-dport-max will be equal to (((PSID+1) x Z) – 1).

B. MAP rules parameters

The parameters that are related to MAP rules are the followings:

- **BMR-IPv6-Prefix**: The BMR's Rule IPv6 Prefix of the MAP address. As stated earlier, all CEs within the same MAP domain will share this IPv6 Prefix.
- **BMR-IPv4-Prefix**: The BMR's public IPv4 prefix that is reserved for CEs. However, the public IPv4 suffix plus the PSID (i.e., EA bits) will, then, uniquely identify each CE.
- **BMR-EA-Len**: The number of EA bits (i.e., EA length).
- **DMR-IPv6-Prefix**: The IPv6 prefix that will be added by DMR to the public IPv4 address to form the IPv4-embedded IPv6 address.

C. Device hardware parameters

The parameters that are related to the device hardware are the followings: (Some of them are needed by DPDK)

- **CPU-FW-Send**: The CPU core to be used by the thread of sending in the forward direction.
- **CPU-FW-Receive**: The CPU core to be used by the thread of receiving in the forward direction.
- **CPU-RV-Send**: The CPU core to be used by the thread of sending in the reverse direction.
- **CPU-RV-Receive**: The CPU core to be used by the thread of receiving in the reverse direction.
- **Mem-Channels**: The number of memory channels to be used. Setting this parameter is optional. The default value is 1.

D. Network traffic parameters

The parameters that are related to the network traffic are the followings:

- **FW**: The forward direction becomes active if set to 1.
- **RV**: The reverse direction becomes active if set to 1.
- **Promisc**: The promiscuous mode will be active if set to a non-zero value.

What follows are the contents of the *Tester.conf* file for the example MAP-T BR test setup depicted in Fig. 8 and Fig. 9. The figures also show the flow of the translated traffic and the flow of the background traffic, respectively. The IP addresses and the port settings are taken from the example of [15], which is also discussed in Section II.

Tester.conf:

```
#Basic parameters
    Tester-L-IPv6: 2001:db8:6::41/64
    Tester-R-IPv4: 203.0.113.56/24
    Tester-R-IPv6: 2001:db8:42::2/64
    DUT-L-IPv6: 2001:db8:6::1/64
    DUT-R-IPv4: 203.0.113.1/24
    DUT-R-IPv6: 2001:db8:42::1/64
    Tester-L-MAC: 00:0c:29:95:f6:a9
    Tester-R-MAC: 00:0c:29:95:f6:b3
    DUT-L-MAC: 00:0c:29:7f:37:48
    DUT-R-MAC: 00:0c:29:7f:37:52
#Port ranges parameters
    FW-dport-min: 1 #as RFC4814 recommends
    FW-dport-max: 49151 #as RFC4814 recommends
    RV-sport-min: 1024 #as RFC4814 recommends
    RV-sport-max: 65535 #as RFC4814 recommends
#MAP rules parameters
    BMR-IPv6-Prefix: 2001:db8:ce::/51
    BMR-IPv4-Prefix: 192.0.2.0/24
    BMR-EA-Len: 13
```

```
    DMR-IPv6-Prefix: 64:ff9b::/96
#Device hardware parameters
    CPU-FW-Send: 2
    CPU-FW-Receive: 4
    CPU-RV-Send: 6
    CPU-RV-Receive: 8
    Mem-Channels: 2
#Network traffic parameters
    FW: 1
    RV: 1
    Promisc: 0
```

VIII. FUTURE WORK

The next step is to implement the MAP-T Tester using C++ language and the DPDK framework [23], a high-performance user-space networking solution that offers fast packet processing and efficient memory, queue, and buffer management. Next, the Tester should be validated by comprehensive benchmarking tests.

The Tester will be developed as free software under the GPL license, and it could be reused as a model for developing testers for other IPv6 transition technologies, especially those IPv4aaS technologies that have not been benchmarked yet.

IX. CONCLUSION

In this paper, we described a brief design structure for a MAP-T technology Tester that complies with the RFC 8219 guidelines and recommendations, followed by a high-level view of the operational requirements, the scope of measurements, and the design factors that should be considered when implementing the Tester program. Then, we presented the installation steps of a testbed for a MAP-T implementation and showed its results. And finally, we disclosed our planned MAP-T BR test program and its related parameters, accompanied by a discussion about how to set the values of these configuration parameters.

REFERENCES

[1] A. Al-Azzawi, "Towards the security analysis of the five most prominent IPv4aaS technologies," *Acta Technica Jaurinensis*, vol. 13, no. 2, pp. 85–98, Mar. 2020, DOI: 10.14513/actatechjaur.v13.n2.530.

[2] A. Al-hamadani, and G. Lencse, "Design of a Software Tester for Benchmarking Lightweight 4over6 Devices," in *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, 2021, pp. 157–161, DOI: 10.1109/TSP52935.2021.9522607.

[3] M. Georgescu, L. Pislaru, and G. Lencse, "Benchmarking methodology for IPv6 transition technologies," IETF RFC 8219, Aug. 2017, DOI: 10.17487/RFC8219.

[4] E. Nordmark, and R. Gilligan, "Basic transition mechanisms for IPv6 hosts and routers," IETF RFC 4213, Oct. 2005, DOI: 10.17487/RFC4213.

[5] S. Bradner, and J. McQuaid, "Benchmarking methodology for network interconnect devices," IETF RFC 2544, Mar. 1999, DOI: 10.17487/RFC2544.

[6] C. Popoviciu, A. Hamza, G. V. d. Velde, and D. Dugatkin, "IPv6 benchmarking methodology for network interconnect devices," IETF RFC 5180, May. 2008, DOI: 10.17487/RFC5180.

[7] G. Lencse, "Design and implementation of a software tester for benchmarking stateless NAT64 gateways," *IEICE Transactions on Communications*, vol. E104-B, no. 2, pp. 128-140, Feb. 2021, DOI: 10.1587/transcom.2019EBN0010.

[8] X. Li, C. Bao, E. W. Dec, O. Troan, S. Matsushima et al., "Mapping of Address and Port using Translation (MAP-T)," IETF RFC 7599, Jul. 2015, DOI: 10.17487/RFC7599.

[9] G. Lencse, J. P. Martinez, L. Howard, R. Patterson, and I. Farrer, "Pros and cons of IPv6 transition technologies for IPv4aaS," active Internet Draft, Jan. 2021, https://tools.ietf.org/html/draft-ietf-v6ops-transition-comparison-00

[10] A. T. H. Al-hamadani, and G. Lencse, "A survey on the performance analysis of IPv6 transition technologies," *Acta Technica Jaurinensis,* vol. 14, no. 2, pp. 186–211, May. 2021, DOI: 10.14513/actatechjaur.00577.

[11] E. O. Troan, W. Dec, X. Li, C. Bao, S. Matsushima et al., "Mapping of Address and Port with Encapsulation (MAP-E)," IETF RFC 7597, Jul. 2015, DOI: 10.17487/RFC7597.

[12] M. Mawatari, M. Kawashima, and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation," IETF RFC 6877, Apr. 2013, DOI: 10.17487/RFC6877.

[13] P. Srisuresh, and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations," IETF RFC 2663, Aug. 1999, DOI: 10.17487/RFC2663.

[14] C. Bao, C. Huitema, M. Bagnulo, M. Boucadair, and X. Li, "IPv6 addressing of IPv4/IPv6 translators," IETF RFC 6052, Oct. 2010, DOI: 10.17487/RFC6052.

[15] Jool. "Jool MAP-T Summary," [Online]. Available: https://www.jool.mx/en/map-t.html.

[16] G. Lencse, "Design and implementation of a software tester for benchmarking stateful NATxy gateways: Theory and practice of extending siitperf for stateful tests," *Computer Communications*, Jun. 2022, DOI: 10.1016/j.comcom.2022.05.028.

[17] G. Lencse, and K. Shima, "Benchmarking Methodology for Stateful NATxy Gateways using RFC 4814 Pseudorandom Port Numbers," active Internet Draft, Mar. 2022, https://datatracker.ietf.org/doc/html/draft-lencse-bmwg-benchmarking-stateful

[18] Jool. "Jool MAP-T Run," [Online]. Available: https://www.jool.mx/en/run-mapt.html.

[19] A. Al-hamadani., "MAP-T Testbed," [Online]. Available: https://github.com/alhamadani-ahmed/MAP-T-Testbed.

[20] G. Lencse, and K. Shima, "Performance analysis of SIIT implementations: Testing and improving the methodology," *Computer Communications*, vol. 156, pp. 54–67, Apr. 2020, DOI: 10.1016/j.comcom.2020.03.034.

[21] G. Lencse, "Adding RFC 4814 random port feature to siitperf: Design, implementation and performance estimation," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 9, no. 3.

[22] D. Newman, and T. Player, "Hash and Stuffing: Overlooked Factors in Network Device Benchmarking," IETF RFC 4814, Mar. 2007, DOI: 10.17487/RFC4814.

[23] D. Scholz, "A look at Intel's dataplane development kit," in *Proc. Seminars Future Internet (FI) and Innovative Internet Technologies and Mobile Communications (IITM)*, Munich, 2014, pp. 115–122, DOI: 10.2313/NET-2014-08-1_15.

Ahmed Al-hamadani graduated from Florida Institute of Technology (FIT), the USA in 2013 with an MSc degree in Computer Science. Since then, he has worked as a lecturer at the Department of Computer Engineering, University of Mosul, Iraq. Ahmed has been awarded the Stipendium Hungaricum scholarship to do his Ph.D. in informatics at Budapest University of Technology and Economics (BME), Hungary. He started his study in the Department of Networked Systems and Services in September 2020. His research field is the benchmarking and performance analysis of IPv6 Transition Technologies.

Gábor Lencse received his MSc and Ph.D. in computer science from the Budapest University of Technology and Economics, Budapest, Hungary in 1994 and 2001, respectively. He has been working full-time for the Department of Telecommunications, Széchenyi István University, Győr, Hungary since 1997. Now, he is a Professor. He has been working part-time for the Department of Networked Systems and Services, Budapest University of Technology and Economics as a Senior Research Fellow since 2005.
His research interests include the performance and security analysis of IPv6 transition technologies. He is a co-author of RFC 8219.

Speaker Adaptation Experiments with Limited Data for End-to-End Text-To-Speech Synthesis using Tacotron2

Ali Raheem Mandeel, Mohammed Salah Al-Radhi, and Tamás Gábor Csapó

***Abstract*—Speech synthesis has the aim of generating human-like speech from text. Nowadays, with end-to-end systems, highly natural synthesized speech can be achieved if a large enough dataset is available from the target speaker. However, often it would be necessary to adapt to a target speaker for whom only a few training samples are available. Limited data speaker adaptation might be a difficult problem due to the overly few training samples. Issues might appear with a limited speaker dataset, such as the irregular allocation of linguistic tokens (i.e., some speech sounds are left out from the synthesized speech). To build lightweight systems, measuring the number of minimum data samples and training epochs is crucial to acquire a reasonable quality. We conducted detailed experiments with four target speakers for adaptive speaker text-to-speech (TTS) synthesis to show the performance of the end-to-end Tacotron2 model and the WaveGlow neural vocoder with an English dataset at several training data samples and training lengths. According to our investigation of objective and subjective evaluations, the Tacotron2 model exhibits good performance in terms of speech quality and similarity for unseen target speakers at 100 sentences of data (pair of text and audio) with a relatively low training time.**

***Index Terms*—speech synthesis, TTS, Tacotron2, WaveGlow, Few-shot.**

I. INTRODUCTION

Speech technology is modern, rapidly developing interdisciplinary field dealing with the artificial intelligence (AI) implementation of any element of the natural human speech chain (such as the speaker, the listener, or even the transmission medium). Several disciplines such as phonetics, machine learning, signal processing, speech acoustics, and cognitive sciences have been used altogether in this field. Based on the available statistical data, it is interesting to say that today we know of just over 7000 living, spoken languages, which poses a serious challenge for speech technology professionals in terms of the uniqueness of each language and the diversity of the linguistic environment [1].

In parallel with the development of infocommunications technology, the need for average users to ensure that language differences, communication features, decrease in the size of devices and the increase in our expectations, as well as other obstacles (e.g., physical resource limitations, functional errors) do not hinder access to certain functions and developments. It is a legitimate user expectation that human speech as a "periphery" should be available for the use of information systems instead of or in addition to peripherals that can be connected to devices (keyboard, mouse, display)[1], [2].

Speech processing (including text-to-speech (TTS) synthesis and automatic speech recognition (ASR)) is beneficial in various fields such as healthcare, security, industry, education, and recreation [3], [4], [5], [6]. Voice disorders in children, such as dysphonia, could be detected early using ASR approaches [3]. The system distinguishes between healthy and pathological sounds using the noisy aware approach. Similarly, ASR might be used to enhance security by catching the target phrases in long sequential sentences or distinct speaker identities. With the advancement of Industry 4.0, machines will become more intelligent, collaborative, and multi-purpose in the future. Laborers can quickly finish the duties by listening to the synthesized speech instructions. Moreover, TTS promises considerable advantages, such that individuals would not need to lose attention by checking the instructions. Furthermore, voiced commands help to have hands-free functions [7].

A. Speech synthesis

Speech synthesis (frequently abbreviated as TTS) has the aim to create natural, human-like voice from written texts. This field has a long history in speech and natural language processing. For a long time, speech synthesis has been a challenging problem. For example, decreasing the TTS model size for real-time synthesis has required much effort and time to enhance the computational complexity. Low resource scenarios, inadequate speakers dataset, robustness problems, expressiveness, and naturalness, have occupied researchers' minds. Much research has been conducted to enhance the quality of synthesized speech in terms of prosody, intelligibility, expressiveness, emotion, robustness, style, naturalness, controllability, etc [8], [9], [10], [11].

A speech synthesizer pipeline basically includes a text analysis module, an acoustic model, and a vocoder (i.e., a speech encoder/decoder module which can decompose the speech signal to a few parameters). A text sequence is converted to linguistic characteristics or phonemes via the text analysis module. Acoustic characteristics are derived from linguistic features or phonemes via acoustic models. At last, vocoders create waveforms based on acoustic/linguistic characteristics. As opposed to the above traditional TTS pipeline, end-to-end

Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary
E-mail: {alimandeel, malradhi, csapot}@tmit.bme.hu

DOI: 10.36244/ICJ.2022.3.7

TTS systems instantly transform characters or phonemes into synthesized speech, often without a linguistic frontend and without a traditional vocoder.

One of the early text-to-speech technologies was articulatory synthesis. Articulatory synthesis generates speech by mimicking the properties and movements of the human articulators such as the glottis, tongue, lips, and vocal tract in general [12]. Another historical technique is formant synthesis, which generates speech using a reduced source-filter paradigm that is controlled by a set of manually-defined parameters [13]. Concatenative synthesis was introduced with the idea of the concatenation of well chosen speech segments from a database [14]. The database comprises audio clips from complete sentences of recorded syllables from voice actors. Unit selection synthesis has the idea that many of such elements are available, and the algorithms try to find large enough units from the natural speech recordings of several hours. Even though the sound quality generated by these methods can be excellent in intelligibility, this approach has limitations. It consumes many resources (especially memory for storing the units) and often results in a speech with reduced smoothness in prosody (pitch, stress or timing).

Statistical Parametric Speech Synthesis (SPSS) was proposed as an alternative to concatenative and unit selection synthesis [15]. It decomposes speech to acoustic parameters and then uses vocoder algorithms to recover speech from the produced acoustic parameters [15], [16], [17], [18], [19]. The essential benefit of SPSS is its adaptability in terms of voice features, speaking styles, and emotions [15]. Typically, two machine learning techniques have been used in SPSS: hidden Markov-models (HMMs) [15] and deep neural networks (DNNs) [16]. Most recently, novelties in deep learning have allowed the creation of neural network-based speech synthesis, which uses deep neural networks as the machine learning model for TTS. The benefit of DNN-TTS over prior systems is its excellent speech quality (naturalness and intelligibility) and the fact that it demands fewer engineering preprocessing and feature creation. Moreover, being efficient during synthesis is as important as obtaining high-quality synthesized speech.

B. Speaker adaptation for speech synthesis

Speaker adaptation (also called voice cloning or custom voice) is the process of customizing the synthesizer to create a voice for any target speaker. Personalizing a TTS is a popular feature in which the application creates sound employing any target speaker's voice recordings. In this case, the general TTS model is trained with an extensive multi-speaker dataset and then adapted to a target speaker. One aspect of adaptive TTS is an efficacious adaptation setting, which reduces adaptation parameters and data per target speaker. The most likely situation in which speaker adaptation is used is when the dataset for the target speaker is too small for single speaker training, but at least enough for adaptation.

More data will lead to enhanced speech quality [20], but it means at a considerable expense of gathering data. Correspondingly, extra fine-tuning parameters can improve the synthesized speech quality, but it increases memory, and implementation costs [21]. At the same time, we might also suffer from the lack of availability of the target speaker's speech data. Accordingly, creating a TTS model which can work with extremely limited data (a few sentences) would be a solution to these problems.

C. Limited data speaker adaptation

Numerous researchers have investigated the speaker adaptation options for end-to-end speech synthesis with few samples (sentences) and tried to enhance the synthesized speech quality with these models. A study on Tacotron2 proficiency of speaker adaptative speech synthesis was accomplished with a Romanian dataset [22]. Their work concluded that it is sufficient to obtain a speaker's identity (a target speaker's voice attributes) with only one sample of data (i.e., one single sentence from the target speaker). For Spanish and Basque, the performance of the Tacotron2-based system was examined with limited amounts of data [23]. Guided attention was implemented, which provided the system with the explicit duration of the phonemes to reduce lost alignment during the inference process. Otherwise, in non-end-to-end SPSS, many studies have been accomplished for speaker adaptation, such as our previous work [24], which used a Continuous vocoder with limited target speaker data for about 14 minutes.

The speech quality was examined with varying quantities of adaption data using Deep Voice 3, and Griffin-Lim [20]. They studied the combination of two methods: speaker encoding and speaker adaptation. Even with a few adaptation audios, both ways generated adequate results. Speaker adaptation achieved slightly better naturalness than speaker encoding. Another relevant paper investigated three meta-learning variations for sample efficient adaptive TTS [25]. Multi-speaker TTS architecture was updated to allow the cloning of unknown speakers using only a few shot samples (a few training data are emerging). The fine-grained and coarse-grained encoders generate two sorts of embeddings: variable-length and global embeddings. The speaker adaption approach could be improved by using a meta-learning algorithm (Model Agnostic Meta-Learning (MAML)) [25]. Overall, while several efforts have been made to improve limited data speaker adaption models, there is still a gap in investigating the amount of English target data required to achieve adequate speaker identity, similarity and high-quality synthesized speech.

D. End-to-end TTS: Tacotron2

Tacotron2 is an end-to-end neural network architecture for TTS [26]. It consists of a network to convert character sequences of input text to mel spectrogram and a neural vocoder (see next section), which can synthesize the speech (Fig. 1). The sequence-to-sequence architecture can elevate this problem by translating the input text sequences into magnitude spectrograms. As a result, this method reduces the need for sophisticated language and speech information because it uses raw data.

The architecture of the Tacotron2 network mixes long short-term memory (LSTM) and convolutional neural network

(CNN) layers to produce mel spectrogram frames from input character sequences. It employs the Short-Time Fourier Transform (STFT) to compute mel spectrograms. Further, this network has an encoder to transform character sequences to interior features representation and a decoder to change these interior features to spectrograms frames. The advantage of using mel spectrograms as an intermediate value is to enable shorter training of the network part and the vocoder. Also, it emphasizes the distinction of low-frequency speech over high-frequency sounds. These benefits ensure that using spectrograms to synthesize speech patterns is intelligible and natural, therefore we also use Tacotron2 in the current paper.

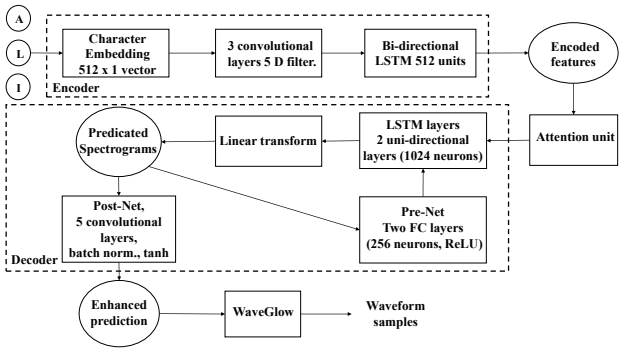


Fig. 1: Tacotron2 architecture.

E. Neural vocoder: WaveGlow

WaveGlow is a neural vocoder, i.e. a generative audio model that samples from distribution to produce waveforms [27]. The number of dimensions in this distribution must match the number of dimensions in the intended output. The samples are acquired from the distribution travel through the flow steps to rebuild the synthesized voice. Twelve coupling layers and twelve invertible 1x1 convolutions exist in the original WaveGlow architecture. Each of the eight layers of dilated convolutions in the affine coupling layer networks has 256 channels for skip links and 512 channels for residual connections [27]. It is a fast model which allows parallel synthesis at 500 kHz on an NVIDIA v100 GPU. Synthesized voices are sharp and close to the real distribution because no Mean squared error (MSE) loss is used for the model training. Moreover, it gives a tractable likelihood of the training data. According to [27], WaveGlow generates high-quality synthesized speech from mel-spectrograms and delivers a quick, efficient voice synthesis. It is faster than early WaveNet versions, therefore we use this neural vocoder in our study.

F. Goal of the current study

This work aims to use as minimal data and parameters as feasible while maintaining good synthesized speech quality during speaker adaptation for end-to-end speech synthesis. We built a TTS model based on the Tacotron2 framework and the WaveGlow neural vocoder. We tested Tacotron2 with five limited data sizes (15, 20, 35, 70, and 100 utterances) from each speaker. We defined three checkpoints (300, 700, and 900) for each target speaker and dataset scenarios. A checkpoint is a moment in the model's state where the current learning rate, weights, and other parameters are stored.

The remainder of the paper is organized in the following manner. Section II describes the experimental design, tools, and dataset. Section III delves more into the experimental results and findings. Finally, the conclusion is presented in Section IV.

II. METHODS

We utilized a multi-speaker dataset with high-quality recordings to train the average end-to-end Tacotron2 model. Then, we adapted this model with four target speakers (two females and two males). Objective and subjective evaluations have been done to test the naturalness and similarity of synthesized speech.

A. End-to-end TTS and neural vocoder

For the Tacotron2, we used the open-source solution offered by NVIDIA (https://github.com/NVIDIA/tacotron2). We employed the official pre-trained WaveGlow vocoder (on the LJ speech dataset with sampling rate 22050 Hz [28]) offered by NVIDIA (https://github.com/NVIDIA/waveglow) with the design of 12 coupling layers, eight dilated convolution layers, 512 residual, and 256 skip connections. This architecture design of WaveGlow is proposed by NVIDIA [27].

B. Speech corpus

In our study, the Tacotron2 model was trained with an English Hi-Fi multi-speaker dataset [29]. This dataset comprises texts from Project Gutenberg and audiobooks from LibriVox. It has approximately 292 hours of speech from ten native English speakers (six females and four males). Every speaker has a minimum of 17 hours of speech sampled at 44.1 kHz in WAV format. Then, we re-sampled it to 22050 Hz to be compatible with the pre-trained WaveGlow vocoder, which was trained with a dataset of 22050 Hz sample rate.

Based on signal-to-noise ratio (SNR) investigation, the Hi-Fi TTS corpus is divided into two categories:

1) The clean subset comprises high-quality audiobooks with adequate audio qualities (at least 40 dB),
2) The other set covers books with fewer SNR (a minimum of 32 dB).

C. Training topology

We used a high-performance NVidia Titan X graphics processing unit (GPU) for training the Tacotron2 model. We used four speakers (two females: Helen Taylor and Sylviamb and two males: Mike Pelton and Tony Oliva) to train the Tacotron2 average model. The total dataset of the four speakers is 88.3 hours (Helen Taylor: 24.3 hours, Sylviamb: 22.2 hours, Mike Pelton: 17.7 hours, and Tony Oliva: 24.1 hours). The dataset from the four Hi-Fi speakers was divided into the training and validation sets for the seen speakers. The validation set consisted primarily of 5% utterances from each speaker.

The Tacotron2 encoder was fed a character sequence, with

each character encoded as a 512-dim character embedding. A spectrogram was created using a 2048 point Fourier transform with Hann windowing, a shift of 16 milliseconds, and a duration of 64 milliseconds. Next, we made a mel spectrogram out of it, with frequency bins of 80 ranging from 125 Hz to 7.6 kHz. We utilized the Adam optimizer [30], a 0.000001 weight decay value, a 0.001 learning rate, hop_length equals 256, iterations_per_checkpoint=1000 (a batch of data has been passed through), and a frame size of 1024. We trained the Tacotron2 model using the four above-mentioned speakers from the Hi-Fi TTS corpus (88.3 hours) until checkpoint 870 000 (870 epochs, an epoch is the number of complete passes through the training dataset). We stopped the training at 870000 based on an investigation of the synthesis quality, which did not improve further after this point. For the whole training procedure, the batch size was set at eight. In addition, we used a pre-trained WaveGlow vocoder provided by NVIDIA (https://github.com/NVIDIA/waveglow).

D. Transfer learning / Speaker adaptation

After that, we adapted the Tacotron2 model, which trained with multi speakers, with four target speakers (two females and two males), namely Female1_other (Maria Kasper), Female2_clean (Cori Samuel), Male1_other (Phil Benson), Male2_clean (John Van Stan) from the Hi-Fi dataset. We used "clean" and "other" sets from the dataset to show the impact of the high-quality and low SNR audios on speaker adaptative synthesized speech quality. We employed two target speakers (female and male/ Female2_clean and Male2_clean) with the "clean" dataset and two target speakers (female and male/ Female1_other and Male1_other) with the "other" dataset. We used the training parameters listed in Table I. The lowest target speaker dataset was 0.48 minutes, and the greatest was 14.32 minutes. During this training, we reduced the batch size to four and iterations_per_checkpoint to 100. The durations of the used datasets are mentioned in the same table. We trained Tacotron2 with three checkpoints (300, 700, and 900). Essential facts to exemplify, we noticed during the training of Tacotron2 that the synthesized speech quality did not be enhanced much over checkpoint 900, or the quality was the same. Therefore, the model reached a stable state, and training it more leads to overfitting.

TABLE I
TARGET SPEAKERS' DATA FOR THE EXPERIMENTS.

Speakers	Dataset (sentences)	Duration (minutes)	Dataset division (training / validation)	Checkpoints
Female1_other	15	0.73	15 / 1	300, 700, 900
	20	1.72	20 / 2	
	35	0.91	35 / 2	
	70	3.84	70 / 4	
	100	5.94	100 / 5	
Female2_clean	15	0.48	15 / 1	300, 700, 900
	20	0.76	20 / 2	
	35	1.47	35 / 2	
	70	2.82	70 / 4	
	100	3.97	100 / 5	
Male1_other	15	0.7	15 / 1	300, 700, 900
	20	0.87	20 / 2	
	35	1.7	35 / 2	
	70	3.56	70 / 4	
	100	5.1	100 / 5	
Male2_clean	15	2.12	15 / 1	300, 700, 900
	20	2.9	20 / 2	
	35	5.25	35 / 2	
	70	10.29	70 / 4	
	100	14.32	100 / 5	

III. RESULTS

We compare the Tacotron2 model performance on different sizes of training data from target speakers ranging from 15 to 100 utterances and three training periods to demonstrate the training efficiency. Both objective and subjective evaluations are carried out. Using an attention-based model, we aim to achieve good alignment with minimal data and a short training period.

A. Objective evaluation

1) MCD (dB):
Mel cepstral distortion (MCD) is a metric that measures the similarity between the spectra of two sounds [31]. The lower the MCD value between synthesized and natural mel cepstral sequences, the more similar a synthetic voice is to a natural one (Eq. 1). x and y are the Mel-cepstrum of the original and synthetic voice waveforms, respectively. M is the order of Mel-cepstrum. The dynamic time warping algorithm was used before making the comparison because sequences are not aligned. We used an open-source DTW-MCD implementation (https://github.com/jasminsternkopf/mel_cepstral_distance).

$$MCD = \frac{10}{\log 10}\sqrt{\sum_{m=1}^{M}\left(x(m) - y(m)\right)^2} \qquad (1)$$

Table II details the MCD results of the synthesized speech sentences that we obtained for the four target speakers. We noticed that the minimum MCD values for the target speakers Female1_other (8.51) and Male2_clean (10.6) are obtained by the checkpoint-900 at 70 and 35 samples. At the same time, we obtained the lowest MCD value for the target speaker at checkpoint 900 with only 35 samples. Also, we noticed that the speaker Female2_clean did not offer common tendencies with increasing data and training periods. We believe the reason for this unusual behavior is because of the smaller dataset duration for this speaker compared to others (see Table I).
Fig. 2 shows the average MCD values of the four target speakers. We can conclude for each data sample of 35, 70, and 100, the MCD values decreased as the training got higher. For example, the MCD values of the 35 samples (12.46 / checkpoint-300, 11.82 / checkpoint-700, and 11.26/ checkpoint-900). Otherwise, the average MCD values did not decline as the data raised from 20 to 100 samples. Overall, the MCD metric did not reflect the expected patterns in certain circumstances as limited data and training periods increased.

2) Encoder-Decoder alignment graphs analysis:
The attention mechanism function is considered as a

TABLE II
THE RESULT OF THE MCD BASED ON THE LIMITED DATA.

Speaker	Dataset	MCD checkpoint-300	MCD checkpoint-700	MCD checkpoint-900
Female1_other	15	10.73	10.43	10.79
	20	9.52	9.65	9.66
	35	9.67	9.35	9.00
	70	9.33	8.65	**8.51**
	100	10.26	9.16	8.91
Female2_clean	15	12.59	14.58	13.22
	20	**12.42**	13.41	13.33
	35	15.52	14.71	13.01
	70	18.18	15.96	15.31
	100	16.42	15.87	14.77
Male1_other	15	16.77	15.2	16.91
	20	12.87	12.73	14.24
	35	13.45	**12.15**	12.45
	70	12.27	12.88	12.3
	100	13.03	12.71	12.87
Male2_clean	15	13.31	11.32	11.38
	20	11.64	11.1	11.4
	35	11.21	11.08	**10.6**
	70	11.58	13.73	10.64
	100	11.66	11.03	11.1

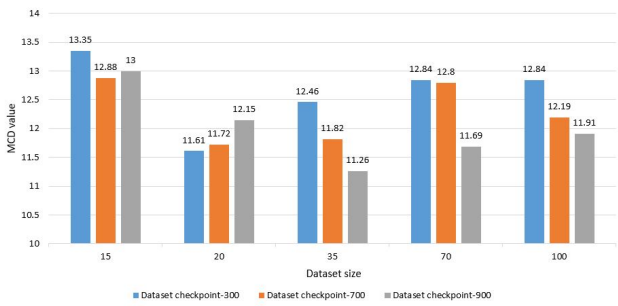


Fig. 2: The average MCD of the four target speakers.

duration model in learning the time alignment between the input text row (encoder) and the outcome acoustic sequence (decoder). The success of an end-to-end model relies on attention alignment. The attention mechanism's irregular and inexact alignment results in word repetitions, mispronunciations, and skipping [32]. The attention alignment graph is a way of showing the quality of end-to-end TTS models, as it shows how well the decoder attends to encoder input [32]. The encoder gradually receives the input and generates status vectors. It examines all status vectors and sequentially generates audio frames. The sloping line appears when audio frames are developed by concentrating on the proper input characters. In other words, the inclination of the diagonal line is an indicator of the quality of the produced speech. The clear near the diagonal line is the sign or the criteria that the alignment graph should meet and be considered as sufficient alignment. The figures (Fig. 3 and Fig. 4) depict the process of attention learning for the speaker Male1_other and the Female1_other for the synthesized sentences *"it is a curious little church,"* and *"in his own mind or that of the public,"* respectively utilizing 15, 20, 35, 70, and 100 samples of training data at three checkpoints (300, 700, and 900). These figures show how the text-to-spectrogram prediction network improved learning attention during the training process and increased data samples. With the smallest amount of training data (15 sentences), clearly the alignment path is not accurate (attention failures), indicating that this is not enough for proper training, even with a high checkpoint number. With training data of at least 20 samples, and after checkpoint 300 of training, the Tacotron2 model began to pick up on alignment. Despite that, the attention line alignment for the data samples of 35 and 70 at checkpoint 300 showed irregular behavior for the speaker Male1_other – this tells us that in general, checkpoint 300 is not enough but the network should be trained longer. In conclusion, according to the encoder-decoder alignment graphs objective evaluation, at least 20 sentences and a checkpoint of 700 shows a good alignment (close to a straight line). for these two sample cases. As we noticed, 15 sentences of data (at several training times) and checkpoint 300 (at various datasets) are insufficient to obtain a proper alignment. Therefore, at this stage of objective evaluation, we nominated 20 sentences/checkpoint 700 to be the minimum threshold that encoder-decoder alignment makes an acceptable outcome.

B. *Subjective evaluation*

In order to determine which proposed version is closer to natural speech, we conducted an online MUSHRA-like test [33]. Our aim was to compare the natural sentences with the synthesized sentences depending on the training data size (number of sentences: 15 / 35 / 70 / 100) and training time (checkpoint: 300 / 700 / 900). In the test, the listeners had to rate the naturalness of each stimulus in a randomized order relative to the reference (which was the natural sentence), from 0 (very unnatural) to 100 (very natural). As a lower anchor, we used the fewest data size and training (15 samples, checkpoint 300) because of its poor quality. We chose three sentences from the test set of the four speakers used in the adaptation experiments. The variants appeared in randomized order (different for each listener). The samples can be found at https://aliraheem.github.io/infocommunications_journal_2022/. With this test, we experimented the following seven variants for the four target speakers:

(a) natural voices,
(b) synthesized voices at 15 sentences/checkpoint-300,
(c) synthesized voices at 35 sentences/checkpoint-900,
(d) synthesized voices at 70 sentences/checkpoint-900,
(e) synthesized voices at 100 sentences/checkpoint-300,
(f) synthesized voices at 100 sentences/checkpoint-700,
(g) synthesized voices at 100 sentences/checkpoint-900.

As a result, we will have an opportunity to observe the fixed 100 sentences at different checkpoints (300, 700, and 900) and with fixed checkpoint 900 at a different number of sentences used as adaptation data (35, 70, and 100). We fitted the four target speakers and three sentences from each of them. Thus, we have 84 sentences (4 speakers x 3 sentences x 7 variants) to compare. Twenty-one subjects (18 in quiet rooms, 3 in a noisy environment; ten females, 11 males; 19- 48 years old;

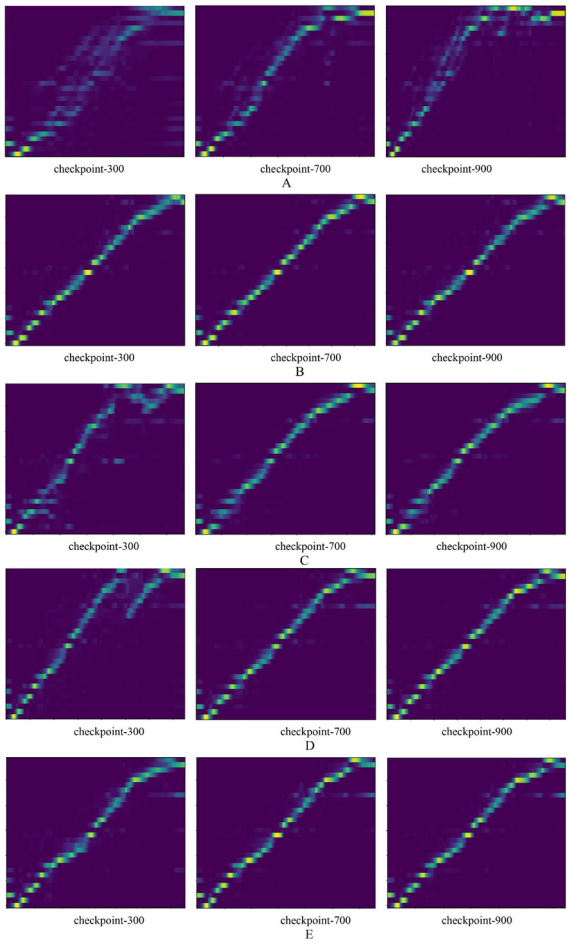


Fig. 3: The Attention alignment graph of the Male1 other speaker. The horizontal axis denotes the decoder time-steps of the creation speech sequence with a max of 175 frames. The vertical axis represents encoder time-steps a most of the 20 phonemes. A= 15 sentences, B= 20 sentences, C= 35 sentences, D= 70 sentences, and E= 100 sentences.

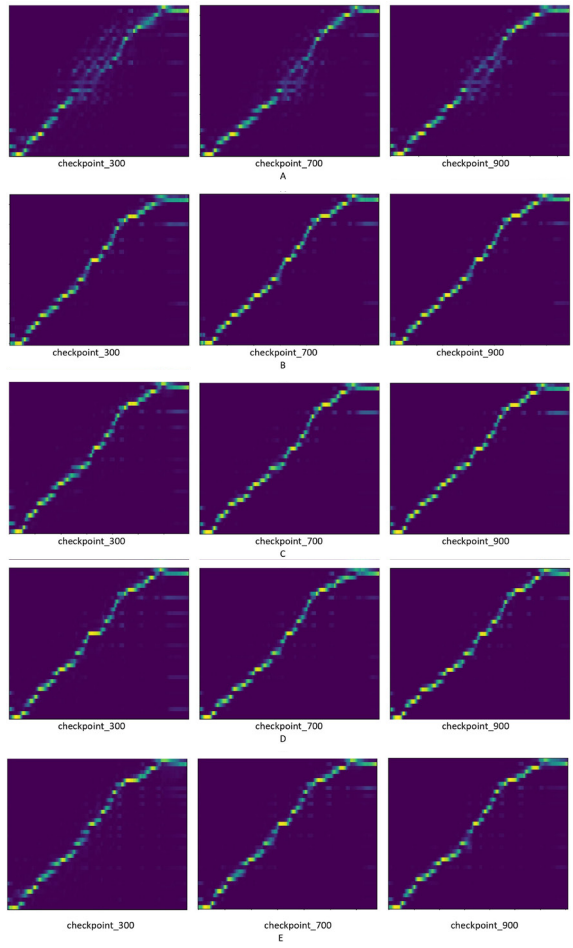


Fig. 4: The Attention alignment graph of the Female1 other speaker. The horizontal axis denotes the decoder time-steps of the creation speech sequence with a max of 175 frames. The vertical axis represents encoder time-steps a most of the 20 phonemes. A= 15 sentences, B= 20 sentences, C= 35 sentences, D= 70 sentences, and E= 100 sentences.

none of them were native English) volunteered to do the test. The test duration was 13.5 minutes, on average.

Fig. 5 presents the average naturalness scores for the different categories (data samples and checkpoints). The natural utterances obtained 88% out of 100% from the subjects. We plotted the '100 samples/checkpoint-900' two times - this way, it is easy to compare the effect of data size visually and checkpoint size. Increasing the data samples from 35 to 70 with fixing checkpoint 900 exhibited a remarkable increase in synthesized speech naturalness (35 samples= 38%, 70 samples= 45%). On the other hand, increasing the data to 100 showed a slight improvement in the synthesized speech naturalness (47%) above the naturalness of data of 70 sentences. Additionally, fixing the data samples at 100 samples and increasing the training periods (checkpoints= 300, 700, and 900) demonstrated the same behavior as the previous case by increasing the speech naturalness (checkpoint 300= 33%, checkpoint 700= 39%, and checkpoint 900= 47%). Therefore, increasing the training period creates more natural speech. For example, comparing the samples 70 and 100, we notice the naturalness of synthesized speech at 70 samples at checkpoint 900 received more scores than 100 samples at checkpoint 700. Similarly, in the case of 35 samples at checkpoint 900 has higher rates than 100 samples at checkpoint 300.

Next, Fig. 6 displays the speech naturalness speaker by speaker. We noticed that increasing the data samples from 35 to 100 with fixing checkpoint 900 showed that the 100 data samples' speech naturalness is more than 35 with all four target speakers. Nevertheless, the 70 samples' speech naturalness did not show the same behavior for all target speakers. Meanwhile, increasing the training period from checkpoint 300 to checkpoint 900 with 100 samples keeps the same scenario with the average case by gaining more speech naturalness.

Furthermore, it appears that increasing the sample size from 70 to 100 decreases the naturalness of speech for "other" data speech, whereas the opposite is true for "clean" data

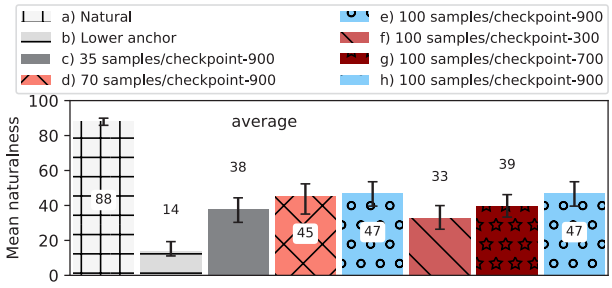


Fig. 5: Average naturalness ratings of the four speakers' speech.

speech. Similarly, using checkpoint 900 only slightly improves performance compared to using checkpoint 700 for "other" data speech. In contrast, the improvement for "clean" data speech (between options f and g) is much more noticeable.

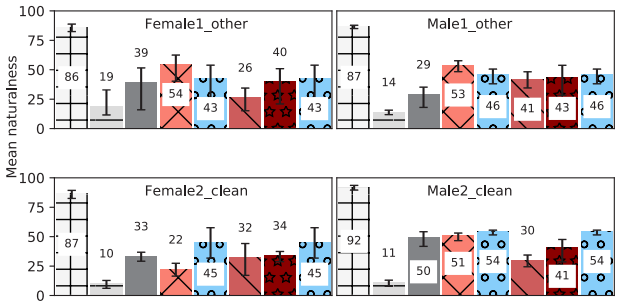


Fig. 6: Naturalness ratings of the four speakers' speech naturalness (speaker by speaker). A higher value indicates a higher level of overall quality. Errorbars represent the bootstrapped 95 percent confidence intervals.

IV. Conclusion

We investigated the minimum dataset and training period required and experimented with the Tacotron2 end-to-end TTS and WaveGlow neural vocoder to construct a TTS model with an unseen target speaker's dataset. First, we trained a general model with a multispeaker dataset of 88.3 hours, after which we applied speaker adaptation. Four target speakers (two females and two males) with two kinds of audio qualities (clean of SNR at least 40 dB and other of SNR equal to 30 dB) were used for the speaker adaptation. We conducted objective and subjective evaluation experiments. Based on our evaluations, the Tacotron2 model admirably produces synthesized speech quality and resemblance with 100 sentences of data (at least five minutes) with a relatively short training period (checkpoints 900) for both speakers of both genders. We did not find a direct relation between the SNR of the adaptation audio dataset and the quality of the synthesized speech between the four speakers' suggested data to build the system (100 sentences). These outcomes can be beneficial to building applications with personalized text-to-speech synthesis, e.g., in speech communication aids for the speaking impaired.

Acknowledgment

The research was partially sponsored by the APH-ALARM project (contract 2019-2.1.2-NEMZ-2020-00012), funded by the European Commission and the National Research, Development and Innovation Office of Hungary and supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory. The research reported in this publication, carried out by the Department of Telecommunications and Media Informatics Budapest University of Technology and Economic and IdomSoft Ltd., was supported by the Ministry of Innovation and Technology and the National Research, Development and Innovation Office within the framework of the National Laboratory of Infocommunication and Information Technology. Tamás Gábor Csapó's research was supported by the Bolyai János Research Fellowship of the Hungarian Academy of Sciences and by the ÚNKP-21-5 (identifier: ÚNKP-21-5-BME-352) New National Excellence Program of the Ministry for Innovation and Technology from the source of the National, Research, Development and Innovation Fund. The Titan X GPU used was donated by NVIDIA Corporation. We would like to thank the subjects for participating in the listening test.

References

[1] G. Németh,"Why is speech technology important and what is it good for? – Hungarian successes firsthand." in Hungarian: Miért fontos és mire jó a beszédtechnológia?– magyar sikerek első kézből," Infocommunications Journal, vol. 70, pp. 12–16, 2015.

[2] G. Németh, G. Olaszy, K. Vicsi, and T. Fegyó, "Speaking machines?! in Hungarian: "Beszélgető gépek?!," HÍRADÁSTECHNIKA: HÍRKÖZLÉS-INFORMATIKA, vol. 64, pp. 53–57, 2009.

[3] Miklós Gábriel Tulics and Klára Vicsi, "Automatic classification possibilities of the voices of children with dysphonia", Infocommunications Journal, Vol. X, No 3, September 2018, pp. 30-36. DOI: 10.36244/ICJ.2018.3.5

[4] Masakazu Kanazawa, Atsushi Ito, Kazuyuki Yamasawa, Takehiko Kasahara, Yuya Kiryu and Fubito Toyama, "Method to Predict Confidential Words in Japanese Judicial Precedents Using Neural Networks With Part-of-Speech Tags", Infocommunications Journal, Vol. XII, No 1, March 2020, pp. 17-25. DOI: 10.36244/ICJ.2020.1.3

[5] G. Du, M. Chen, C. Liu, B. Zhang and P. Zhang, "Online Robot Teaching With Natural Human–Robot Interaction," in IEEE Transactions on Industrial Electronics, vol. 65, no. 12, pp. 9571-9581, Dec. 2018, DOI: 10.1109/TIE.2018.2823667.

[6] P. -S. Chiu, J. -W. Chang, M. -C. Lee, C. -H. Chen and D. -S. Lee, "Enabling Intelligent Environment by the Design of Emotionally Aware Virtual Assistant: A Case of Smart Campus," in IEEE Access, vol. 8, pp. 62032-62041, 2020, DOI: 10.1109/ACCESS.2020.2984383.

[7] Jean D. Hallewell Haslwanter, Michael Heiml, and Josef Wolfartsberger. Lost in translation: machine translation and text-to-speech in industry 4.0. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '19). Association for Computing Machinery, USA, pp. 333–342, 2019, DOI: 10.1145/3316782.3322746.

[8] G. C. Tamás, Z. Csaba, and N. Géza, "A Study of Prosodic Variability Methods in a Corpus-Based Unit Selection Text-To-Speech System," Infocommunications Journal, vol. 65, no. 1, pp. 32–37, 2010.

[9] Y. Ren, Y. Ruan, X. Tan, T. Qin, S. Zhao, Z. Zhao, and T.-Y. Liu, "Fastspeech: Fast, robust and controllable text to speech." Advances in Neural Information Processing Systems, vol. 32, pp. 3165–3174, 2019.

[10] G. C. Tamás and G. Németh, "A novel irregular voice model for HMM-based speech synthesis," in ISCA 8th Speech Synthesis Worksop (SSW8), 2013, pp. 229–234.

[11] A. Łańcucki, "Fastpitch: Parallel Text-to-Speech with Pitch Prediction," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 6588–6592, DOI: 10.1109/ICASSP39728.2021.9413889.

[12] C. H. Shadle and R. I. Damper, "Prospects for articulatory synthesis: A position paper," in 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis, pp. 121–126, 2001.

[13] D. H. Klatt, "Software for a cascade/parallel formant synthesizer," the Journal of the Acoustical Society of America, vol. 67, no. 3, pp. 971–995, 1980, DOI: 10.1121/1.383940.

[14] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech communication, vol. 9, no. 5-6, pp. 453–467, 1990, DOI: 10.21437/eurospeech.1989-172.

[15] H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis," speech communication, vol. 51, no. 11, pp. 1039–1064, 2009, DOI: 10.1016/j.specom.2009.04.004.

[16] H. Ze, A. Senior and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 7962-7966, DOI: 10.1109/ICASSP.2013.6639215.

[17] S. King, "An introduction to statistical parametric speech synthesis," Sadhana, vol. 36, no. 5, pp. 837–852, 2011, DOI: 10.1007/s12046-011-0048-y.

[18] T. G. Csapó, G. Németh, M. Cernak and P. N. Garner, "Modeling unvoiced sounds in statistical parametric speech synthesis with a continuous vocoder," in EUSIPCO, IEEE, 2016, pp. 1338-1342, DOI: 10.1109/EUSIPCO.2016.7760466.

[19] M. S. Al-Radhi, O. Abdo, T. G. Csapó, S. Abdou, G. Németh, and M. Fashal, "A continuous vocoder for statistical parametric speech synthesis and its evaluation using an audio-visual phonetically annotated Arabic corpus," Computer Speech and Language, vol. 60, 2020, DOI: 10.1016/j.csl.2019.101025.

[20] S. Arik, J. Chen, K. Peng, W. Ping, and Y. Zhou, "Neural voice cloning with a few samples," Advances in Neural Information Processing Systems, vol. 31, pp. 10 040–10 050, 2018.

[21] X. Tan, T. Qin, F. Soong, and T.-Y. Liu, "A survey on neural speech synthesis," arXiv preprint arXiv:2106.15561, 2021.

[22] G. Săracu and A. Stan, "An analysis of the data efficiency in Tacotron2 speech synthesis system," 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2021, pp. 172–176, DOI: 10.1109/SpeD53181.2021.9587411.

[23] V. García, I. Hernáez, and E. Navas, "Evaluation of Tacotron based Synthesizers for Spanish and Basque," Applied Sciences, vol. 12, no. 3, p. 1686, 2022, DOI: 10.3390/app12031686.

[24] A. R. Mandeel, M. S. Al-Radhi, and T. G. Csapó, "Speaker Adaptation with Continuous Vocoder-Based DNN-TTS," in International Conference on Speech and Computer (SPECOM), Springer, vol. 12997, pp. 407–416, 2021, DOI: 10.1007/978-3-030-87802-3_37.

[25] S. -F. Huang, C. -J. Lin, D. -R. Liu, Y. -C. Chen and H. -y. Lee, "Meta-TTS: Meta-Learning for Few-Shot Speaker Adaptive Text-to-Speech," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1558–1571, 2022, DOI: 10.1109/TASLP.2022.3167258.

[26] J. Shen etal., "Natural TTS Synthesis by Conditioning Wavenet on MEL Spectrogram Predictions," in ICASSP, IEEE, 2018, pp. 4779–4783, DOI: 10.1109/ICASSP.2018.8461368.

[27] R. Prenger, R. Valle and B. Catanzaro, "Waveglow: A Flow-based Generative Network for Speech Synthesis," in ICASSP, IEEE, 2019, pp. 3617-3621, DOI: 10.1109/ICASSP.2019.8683143.

[28] K. Ito and L. Johnson, "The lj speech dataset," https://keithito.com/LJ-Speech-Dataset/, 2017.

[29] Bakhturina, E., Lavrukhin, V., Ginsburg, B., Zhang, Y. (2021) Hi-Fi Multi-Speaker English TTS Dataset. Proc. Interspeech 2021, pp. 2776–2780, DOI: 10.21437/Interspeech.2021-1599.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in ICLR, USA, 2015.

[31] J. Kominek, T. Schultz, and A. W. Black, "Synthesizer voice quality of new languages calibrated with mean mel cepstral distortion." in SLTU, 2008, pp. 63–68.

[32] X. Zhu, Y. Zhang, S. Yang, L. Xue and L. Xie, "Pre-Alignment Guided Attention for Improving Training Efficiency and Model Stability in End-to-End Speech Synthesis," in IEEE Access, vol. 7, pp. 65 955–65 964, 2019, DOI: 10.1109/ACCESS.2019.2914149.

[33] "ITU-R Recommendation BS.1534: Method for the subjective assessment of intermediate audio quality," 2001.

Ali Raheem Mandeel was born in Iraq. He received his M.Sc degree in computer engineering from the University of Missouri, Columbia, the US, in 2016. Currently, he is pursuing his Ph.D. in computer engineering in the Department of Telecommunications and Media Informatics at the Budapest University of Technology and Economics, Budapest, Hungary. His research interests include deep machine learning, signal processing, and speech synthesis.

Dr. Mohammed Salah Al-Radhi received a B.Sc. degree in Computer Engineering at Basra University in 2007, and a M.Sc. degree in Communication Systems Engineering at Portsmouth University, UK which was achieved with first-class honours in 2012 and awarded the MSc top student certificate in 2013. He received his Ph.D. from the Faculty of Electrical Engineering and Informatics at the Speech Technology and Smart Interactions Laboratory in Budapest University of Technology and Economics, Hungary in 2020, where he obtained it with honour (100%) and summa cum laude. Since October 2020, he has been a postdoctoral researcher in the Department of Telecommunications and Media Informatics, BME, Budapest. He served as a reviewer for several top-tier journals and conferences proceedings. His main interests are Artificial Intelligence and Machine Learning in speech signal processing and voice conversion.

Dr. Tamás Gábor Csapó obtained his Ph.D. in computer science & speech synthesis from Budapest University of Technology and Economics (BME), Hungary in 2014. He was a Fulbright scholar at Indiana University, USA in 2014, where he started to deal with ultrasound imaging of the tongue. In 2016, he joined the MTA-ELTE Lingual Articulation Research Group, focusing on investigating the Hungarian articulation during speech production. His research interests include Silent Speech Interfaces, ultrasound-based articulatory-to-acoustic mapping and articulatory-to-acoustic inversion, speech analysis and synthesis, and deep learning methods applied for speech technologies. Currently, he is a research fellow at BME.

On the Challenges of Mutual Interference between Cable Television Networks and Mobile Fixed Communication Networks in the Digital Dividend Bands

Hussein Taha[1], Péter Vári[2], and Szilvia Nagy[2]

***Abstract*—Recently, the issue of monitoring and repairing leakage from cable television networks have re-emerged, particularly after the International Telecommunication Union released a part of the ultra-high frequency spectrum to mobile broadband services. The newly allocated spectrum, known as the digital dividend bands, was traditionally used throughout Europe for digital TV broadcasting. The emerging problem is the mutual interference between the new frequency spectrum utilized by the Mobile/Fixed Communication Networks and the band used by cable TV providers to offer their services. This article is a brief overview and a starting point for extensive research in this area. We started with a simple description of the cable television system and mobile/fixed communication networks focusing on the aspects associated with ingress and egress interference issues. We also discussed the approaches for detecting and measuring mutual interference and reviewed the relevant literature. This article is concluded with some proposed measures for reducing or mitigating mutual interference.**

***Index Terms*—MFCN, LTE/5G, Cable TV, ingress/egress interference, digital dividend bands.**

I. INTRODUCTION

The monitoring and repairing leakage from cable networks operating in the Very High Frequency (VHF) band have been a priority for cable TV providers for decades. However, in the last several years, the signals sent over the cable TV system have occupied nearly the Ultra High Frequency (UHF) spectrum to offer more channels to cable TV users [1]. Many other vital services are operating in the UHF band, such as personal radio services, Public Protection and Disaster Relief (PPDR) services, government communication systems, and air navigation systems [2]. Furthermore, to address the increasing demand for frequencies suited for the implementation of mobile broadband services, the International Telecommunication Union (ITU) has reallocated part of the UHF spectrum range extending from 694 MHz to 862 MHz for Mobile/Fixed Communication Networks (MFCN), particularly for 4G/5G networks [2], [3], [4]. The new bands were allocated in two-phase: the first digital dividend (known as the 800 MHz band) and the second digital dividend (known as the 700 MHz band). Moreover, there has recently been a demand for more of the current UHF television broadcasting spectrum to be reallocated for MFCN services. Thus, the emerging problem is the mutual interference between the frequency spectrum used in MFCN, PPDR, and the spectrum used by cable TV operators to provide their services. Figure 1 outlines that the 700 MHz band (694-790 MHz spectrum) agreement in Hungary is shared by three cellular companies (mobile operators) and public safety agencies [5], [6]. Cable TV systems use the adjacent 470-694 MHz range. The two systems are theoretically independent, but in practice, the cable TV band's higher channels overlap with the communication channels where MFCN is broadcast and vice versa.

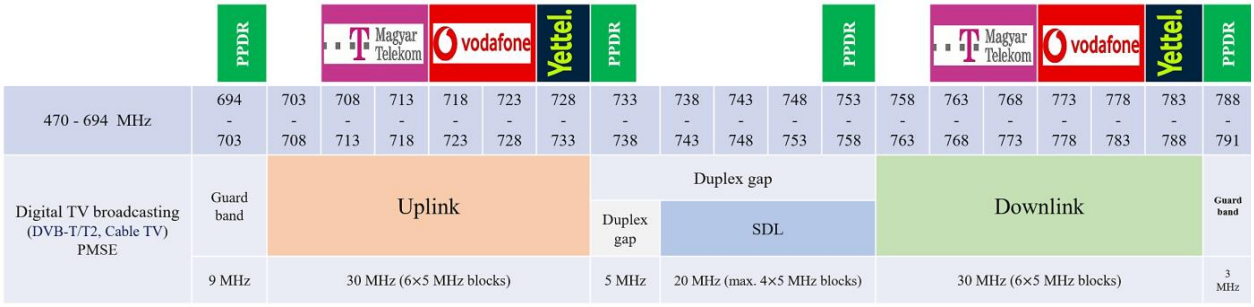


Fig. 1. 700 MHz frequency band: Rights of Use in Hungary

[1] Doctoral School of Multidisciplinary Engineering Sciences, Széchenyi István University, Győr, Hungary (e-mail: hussein.taha91@gmail.com)
[2] Department of Telecommunications, Széchenyi István University, Győr, Hungary (e-mail: varip@sze.hu, nagysz@sze.hu)

DOI: 10.36244/ICJ.2022.3.8

A cable TV system is characterized as a closed system that transmits signals on frequencies commonly utilized for multiple purposes in air broadcast environments. However, signals may leak out/into the cable network under certain conditions, causing mutual interference between over-the-air users and cable TV transmission, known as egress/ingress interference [1]. Figure 2 shows the general scenario of mutual interference between the cable TV system and MFCN.

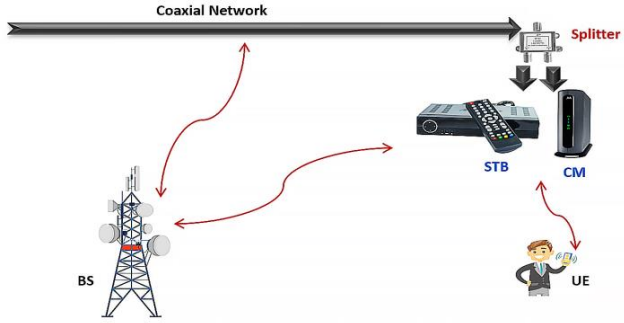


Fig. 2. General scenario of mutual interference between cable TV system and MFCN

Since frequencies in the 700 MHz and 800 MHz bands have been licensed to MFCN, spectrum regulators and public interest management must guarantee that out-of-band interference caused by cable signal leakage into authorized users of the spectrum is avoided. In this regard, the Society of Cable Telecommunications Engineers (SCTE) has issued several technical standards for cable networks. SCTE requires cable television service providers to adhere to certain signal leakage restrictions. TABLE I shows acceptable signal leakage limits as per SCTE [1], [7]:

TABLE I
ACCEPTABLE SIGNAL LEAKAGE LIMITS AS PER SCTE [1], [7]

Frequencies	Limits of signal leakage	Distance (in meters)
216 MHz < Analog signals ≤ 54 MHz	**15** $\mu V/m$	**30**
216 MHz < Digital signals ≤ 54 MHz	**13.1** $\mu V/m$	**30**
216 MHz ≥ Analog signals > 54 MHz	**20** $\mu V/m$	**3**
216 MHz ≥ Digital signals > 54 MHz	**17.4** $\mu V/m$	**3**

Over the last years, numerous field studies, papers, conferences, and symposia have been presented on ingress/egress interference issues in cable and wireless communities. Our article addresses the emerging challenges in mutual interference between cable TV systems and MFCN operating in the digital dividend bands to ensure the electromagnetic compatibility of both systems.

The rest of the article is arranged as follows: Section II presents a brief overview of cable TV system structure, and MFCNs with indicting its properties related to the mutual interference. The sources, effects, and indicators of the two main types of mutual interference are presented in section III. Section IV is devoted to understanding how ingress and egress interference is detected, located, and measured, as well as reviewing the relevant literature. Section VI provides measures that can be taken to reduce or mitigate mutual interference. This paper is concluded with section V.

II. OVERVIEW OF CABLE TV SYSTEM, AND MFCN

A. Cable TV system

A cable television system is a structure that contains a network of closed transmission lines connected with the signal generating, reception, and control hardware needed to deliver cable services. The cable TV system can now provide paying users with three services: television programs, telephone service, and high-speed internet access. Radio Frequency (RF) signals are used to deliver these services across the traditional coaxial cable networks or the Hybrid Fiber-Coaxial (HFC) network [8], [9], [10], [11].

Cable TV providers (Multiple System Operators or MSOs) deliver TV programs service in three tiers or categories, each with its own fee [8], [9]. Cable providers often provide the basic service, which is the most basic level of television programs service. The basic service category provides access to a variety of public, educational, and government television channels that are regulated by the local authority in the nation or city where the cable television operator is licensed to operate. The second category includes all cable system program channels that are not listed in the basic service category. This category may have one or more levels. Furthermore, MSOs provide separate services for each individual channel or program, which is frequently referred to as Pay-Per-View (PPV) service.

Figure 3 shows that the structure of the cable TV system consists of the operator part and the subscriber network part separated by a network termination outlet.

a) The operator part of the network

The cable TV operator constructs and maintains this part to guarantee the supply of a sufficient signal level to all customers' houses. The cable TV operator must follow the requirements defined in the standard IEC 60728-1 to determine the appropriate signal levels and characteristics for the services provided [12].

The MSOs gather all the information required to offer their services. The cable TV systems previously used the DVB-C2 (Digital Video Broadcasting - Cable 2) standard to deliver digital signals through broadband cable networks. DVB-C2 allowed for the most optimal use of cable network resources to generate variable bandwidth signals, which achieved improved spectral efficiency. Also, DVB-C2 was distinguished by its operational flexibility and adaptation to different channel conditions [10], [13].

By the end of the 1990s, MSOs were able to provide real-time communication services, namely Voice over Internet Protocol (VoIP). Later, a Content Distribution Network (CDN) was created to provide advanced multimedia services over a packet-switched network based on the Internet Protocol (IP), namely IPTV system. The IPTV content is delivered to Set-Top

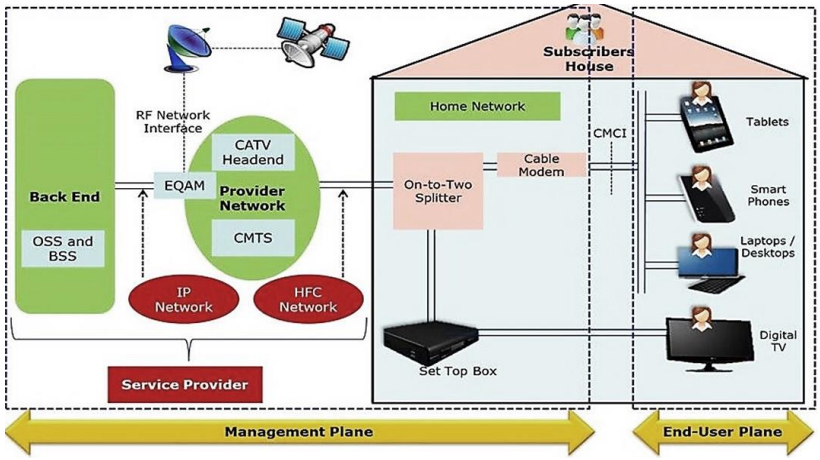


Fig. 3. The structure of cable TV system

Boxes (STBs) through a dedicated and managed network using multicast and initial unicast burst during channel change [14].

In the last few years, Over the Top entertainment (OTT) services have been launched and streamed directly to viewers using modern interactive platforms based on wireless technology. OTT services are privately owned media services that offer subscribers content streaming anywhere and at any time through an open internet and unmanaged network utilizing unicast and simulated multicast (User Datagram Protocol or UDP/Transmission Control Protocol or TCP) routing topology [14]. The evolution of OTT posed new challenges to cable operators that prompted them to develop IPTV service delivery architecture to provide interactive TV services, popular Video on Demand (VoD), and PPV services like YouTube, Netflix, Hulu, Sky Go, Amazon Prime, and others.

As a result, MSOs currently deliver IP video over their network in a controlled environment by providing content libraries and TV services anywhere through the internet as well. MSOs are also an important strategic partner for OTT providers. In order to launch OTT and IPTV, a cable operator needs a management system and a multiplatform player.

The video asset and the associated metadata are typically uploaded by the content producers to terrestrial and satellite transmitters, where they are broadcast to any cable television providers that are authorized to provide that content. The cable TV headend receives these allocated signals, which are then loaded by MSOs into the appropriate VoD distribution servers and made available to customers for a subscription.

On the other hand, MSOs operate with Internet and telephone systems to provide high-speed data transfer over the existing coaxial cable using the Data Over Cable Service Interface Specification (DOCSIS) protocol. The DOCSIS system offers the bi-directional transfer of Internet Protocol (IP) traffic or broadband service between the cable TV headend and subscriber location. The broadband service is supported by a Cable Modem Termination System (CMTS) or a Converged Cable Access Platform (CCAP) at the headend and a cable modem at the subscriber location. A Passive Optical Network (PON) system at the headend and an Optical Networking Unit (ONU) at the subscriber location are additional options for supporting this service [14].

Finally, the DOCSIS and television signals are combined, assigned to the appropriate channels, and then converted to optical signals. The headend connects to the distribution hubs over an all-coaxial or HFC cable network. In a distribution network, the optical nodes convert optical impulses from fibers to electrical RF signals while amplifiers boost weak television signals. The signal is then delivered from the distribution hub to the home network.

b) The subscriber part of the network

The subscriber's coaxial network at home connects with or without home amplifiers with access points or directly to the Set-Top Box (STB) to access TV channels or Cable Modem (CM) to access the Internet. There is a disparity in the home network's quality of components and installations, ranging from high to poor. Low-quality shielding coaxial cables, connectors, and splitters are commonly used or are improperly connected or terminated. As a result, a significant amount of cumulative attenuation of the components may deteriorate the strength and quality of the signal.

B. MFCN (Mobile/Fixed Communications Networks)

The term "MFCN" (Mobile/Fixed Communications Networks) is used in this article to refer to International Mobile Telecommunications (IMT) services and other communications networks operating in the digital dividend bands, namely Long-Term Evolution (LTE), 5G, and public safety services. MFCNs operating in digital dividend bands differ from earlier cellular technology in several aspects, including [15], [16], [17]:

a) Frequency allocations

The frequencies used in the MFCN deployments in the 700/800 MHz ranges are usually at or above the upper-frequency ranges used in most current cable networks. Figure 1 above depicts Hungary's MFCN agreement in the 700 MHz spectrum. It is worth mentioning that transmissions in the low 700 MHz range travel over longer distances but are less attenuated when passing through structures.

b) The modulation schemes

In the MFCN downlink, Orthogonal Frequency-Division Multiple-Access (OFDMA) is used with these modulation schemes: QPSK, 16 QAM, 64 QAM, 256 QAM, and 1024 QAM, whereas in the MFCN uplink, Single Carrier-FDMA (SC-FDMA) is used with QPSK, 16 QAM, and 64 QAM constellations. It is worth mentioning that OFDMA demands a high potential power spectral density, whereas the high modulation level requires a higher SNR and makes the system more vulnerable to interference.

c) Bandwidth and resource block allocation

MFCN use a flexible channel bandwidth. According to [17], MFCN operating in the 700 MHz band can support channel bandwidths of 3, 5, 10, 15, or 20 MHz, while those operating in the 800 MHz band can support channel bandwidths of 5, 10, 15, or 20 MHz. However, most mobile operators primarily use channel bandwidths of 10 and 15 MHz when deploying LTE/5G networks in the digital dividend bands.

The MFCN channel bandwidth is shared among numerous users, whereby a certain number of sub-carriers and OFDM symbols are allocated to each user in the form of a "resource block". For example, 50 resource blocks can be allocated within a 10 MHz channel bandwidth.

There is a correlation between the level of MFCN interference in cable networks and resource block allocations, which may be summarized as follows: As previously stated, utilizing OFDM in the downlink requires a potentially high-power spectral density. Figure 4 illustrates that when just a few resource blocks are allocated within the channel bandwidth, the overall signal strength is focused on a smaller portion of the available bandwidth and distributed over fewer sub-carriers [18]. As a result, this may increase the likelihood of MFCN interfering with cable networks.

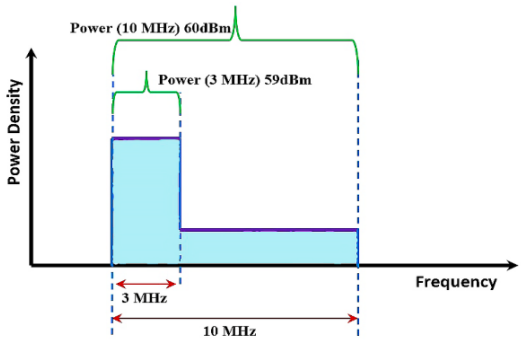


Fig. 4. Example of a correlation between power density and resource blocks allocations in channel bandwidth

III. INGRESS/EGRESS INTERFERENCE

RF interference happens when an undesired RF signal enters the frequency band used by a service, which negatively affects the quality of service. There are two types of mutual interference between cable TV systems and MFCN.

A. Ingress interference

In this type, unwanted RF signals interfere with cable transmissions [1].

MFCN uplink and downlink signals, impulse noise, and Gaussian white noise are common causes of ingress interference. The interference caused by MFCN is mainly related to how frequency allocations, modulation schemes, and bandwidth are handled. Impulse noise is defined by a short and sharp increase in decibel level, frequently caused by turning on and off electrical gadgets in the home or by loud events in the surrounding environment. Gaussian noise is identified as filtered white noise coming from Customer Premises Equipment (CPE), such as STB, CM, TV set, or other devices used in the cable network. Gaussian noise will lower the Modulation Error Ratio (MER) and Forward Error Correction (FEC) values, causing transmission disruption. Initial indications and effects of ingress interference in cable TV networks are customer complaints, for example, temporary signal loss, poor sound or picture quality, frozen images, intermittent audio, or poor or no data connectivity.

B. Egress interference

Egress interference occurs when unwanted RF signals leak from the cable TV network and interfere with MFCN [1].

Even though cable TV companies comply with SCTE-mandated signal leakage restrictions, the interference probability still exists under certain circumstances. The lack of sufficient shielding, damage to a portion of a closed cable network, or poor installation and maintenance are examples of these circumstances. The leaking signals are QAM signals that may block or degrade desired MFCN communications. MFCN operating in the 700 and 800 MHz bands are particularly vulnerable to interference because the MFCN was designed to achieve high data rates at the cost of lower signal robustness. MFCN uplink signals may have very low amplitude at the base station. Thus, if the leakage point from the cable TV network is near the base station, the QAM signals leaked in the coverage area may be equal to or higher than the MFCN uplink signals. As a result, a user near one of the leakage points might experience dropped or blocked calls and reduced data rates. On the other hand, there will be problems with the MFCN downlink if there are lots of leaks in the coverage area. This would show up as poor coverage for one or more sectors near leakage points, particularly near the edge of the base station coverage area.

Familiar sources of egress interference include poor shielding effectiveness of cable system elements, damaged or loose hardline connectors, unterminated outlets, damage from environmental phenomena, and poor-quality materials. It is worth noting that any egress points are also possible ingress points.

The initial indications and effects of the leakage on MFCN are customer complaints, such as interrupted calls, low voice quality, low data throughput, or even no data connection. Besides, the statistical information offered by the base stations themselves is a significant noise indicator, such as a low Signal-to-Noise Ratio (SNR) or high level of Received Signal Strength Index (RSSI).

On the other hand, both cellular carriers and cable TV providers perform routine tests for early detection of noise and negative impacts, including cable TV providers calculating the

Cumulative Leakage Index (CLI) in a given area and mobile operators performing drive-testing.

IV. INTERFERENCE HUNTING APPROACHES AND RELEVANT LITERATURE REVIEW

This section presents the approaches for detecting and measuring ingress and egress interference and reviews the relevant literature.

We can distinguish three main approaches for detecting and measuring leakage from cable TV networks [19]. The simplest and most cost-effective method is injecting a narrow-band carrier at a specific frequency between two adjacent QAM channels and then measuring the leakage at that frequency. The injected carriers have unique signatures for each cable TV provider, so detecting which cable TV network emitted the leaking test signal is feasible, and thus the leakage can be fixed. Using this approach, the leakage field strength can be automatically and continually monitored by installing inexpensive receivers in fleet vehicles owned by companies in the region of the cable TV provider. For instance, leakage detectors may be installed on garbage trucks that service the target area, and the cable TV operator would then get the collected data to analyze and take the necessary action. The second method depends on creating a correlation of QAM signals at the cable system's headend with signals measured in the field. However, this approach requires special equipment and independent data link at both the headend and in the field. The most flexible approach is a spectral analysis using a monitoring receiver and high gain active directional antenna [20]. Using this approach, the equipment can offer a comprehensive snapshot of the leakage signal, yet it must be manually operated by a technician.

TABLE II describes in detail the characteristics and methodology used in the leakage detection methods, including manufacturers of appropriate equipment for each technique as well as relevant reference studies.

In the paper [19], numerous authors worked to establish several detailed and updated operating procedures to mitigate digital QAM signals leakage from cable TV networks. These operational practices included defining the performance metrics to be measured, recommendations for obtaining appropriate equipment and how to calibrate and use them, explaining the requirements for continuous monitoring and maintenance of leakages, and coordinating leakage measurements and then analyzing them for troubleshooting. Several instructions were also provided on what to do when contacting the LTE operator concerning interference.

The Society of Cable Telecommunications Engineers (SCTE) in the USA performed laboratory testing in a controlled environment in two stages [1]. In the first stage, the effect of QAM signal leakage on the LTE downlink was investigated by

TABLE II
CHARACTERISTICS OF LEAKAGE DETECTION APPROACHES

Approaches	Injected carrier method	Correlation Method	Portable spectrum analyzer method
Detection system components	- A signal source "marker" is installed in a headend or hub. - A platform for detecting GPS leaks in fleets. - The handheld detection devices.	- A reference signal that is acquired by connecting to the cable network. - A leakage signal captured using an antenna at a field location. - A correlation detector.	- Portable monitoring receiver. - Active directional high gain antenna.
Methodology	Inject a narrow-band carrier "marker" at a specific frequency between two adjacent QAM channels in the headend and then measure the leakage at that frequency using handheld field detection units that are programmed to detect the corresponding marker signal.	The correlation method determines if there is a correlation between the snapshot of the QAM signal at the headend (or another site in the network) and a leakage signal received at the leaking antenna. The two signals are the same if there is a correlation and the two signals have the same components. In this case, we know that QAM egress is being monitored. The Time Difference of Arrival (TDOA) location technique compares the times of arrival of the two waveforms to locate the source and intensity of the leakage.	The approach consists of two major steps: The first step is to examine the statistical data provided by the MFCN base station, which may indicate a potential leakage issue, such as a low SNR or high RSSI. The next step is to connect a spectrum analyzer to the uplink/downlink antenna test point and check for clear signs of RF interference.
Features	The simplest and most cost-effective method.	This technique considers two crucial factors simultaneously to determine whether a leakage will negatively impact MFCN transmission. The first is the distance and beam path from the leak to the tower. The second is the leak's amplitude at the MFCN transmission frequency.	The most flexible approach that can provide a comprehensive snapshot of the leakage signal.
Manufacturers	CPAT, VIAVI, ComSonics, and Effigis.	ARCOM Digital, and VIAVI.	Rohde & Schwarz, Aaronia AG, Narda, Anritsu, and VIAVI.
Related reference studies	SCTE operational practices [17].	SCTE operational practices [17], ARCOM technical report [19].	SCTE operational practices [17], SCTE technical report [20], In Poland [21].

measuring the CQI and the user equipment data throughput during the test. Besides, numerous parameters have been adjusted for downlink tunings, such as transport block size index, code rate, and block error rate. Then, the QAM signals strength and the error correction level were varied to determine their effect on the downlink. The test results at this stage showed an inverse relationship between the QAM power leakage and the distance between the leak source and the affected receiver. As the distance doubles, the resulting field strength of a leak is decreased by 6 dB, assuming there are no reflections or obstacles in the free-space signal path. In the second stage, the effect of QAM signal leakage on the LTE uplink was studied by measuring the RSSI and the data throughput of the base station. During the test, the QAM signal strength was increased by 6 dB every 15 minutes, and then the data throughput and timestamp were measured at each level. The results indicated that digital QAM signal leakage significantly impacted LTE uplink performance to the point where a low field strength (5 μV/m) at the antenna plane produced interference.

In [21], ARCOM equipment was used by mobile operator "Verizon" to identify QAM signal leakage from cable TV networks. Firstly, the specialists first divided the coverage of each tower into quadrants. When initial indications indicate interference in the quadrant, such as low SNR or high RSSI, they drive out the affected quadrant using an analyzer and monopole antenna. Then, the source of the leakage was isolated using an analyzer and a Yagi antenna. Time Difference of Arrival (TDOA) technology has been used in some cases to locate the leakage in the GPS. According to the obtained results, the average leak per mile was 70% in high frequency only (at 717 MHz), 20% in low frequency only (at the aeronautical band), and 10% in both high and low frequency.

The authors of [22] emphasized the need to monitor leakage in both aeronautical and broadband frequencies. Since field measurements revealed no correlation between the strength of the leakage field and low or high frequencies, leaks might exist in the higher band even if they did not exist in the lower band. In addition, the leakage strength varies depending on the source and mechanism. The researchers developed a program for detecting and repairing cable signal leakage over the air at multiple frequencies, reducing harmful interference, and ensuring the quality of cable TV service.

In Poland [23], the researchers used a monitoring receiver and an active directional antenna to measure leakage from several components of a cable TV network operating at a site. The leakage field strength was first measured at a certain distance from the equipment and employed as a threshold level, and then the leakage field strength was measured near the leaking source. By comparing the level of radiation observed in the field with the maximum levels given by the Federal Communications Commission (FCC), the researchers could determine the critical leak locations at specific frequencies. They concluded that poor initial network installations were responsible for most leakage cases.

The ingress interference and direct pickup noise (impulse noise) are discussed in detail in [24], [25]. The causes, sources, and effects of ingress interference and impulse noise at UHF band were explored in [24]. This report also included many practical measures contributing to rapid detection and mitigation of ingress and impulse interference. The authors of research published in Belgium [25] reported the link between ingress interference and direct pickup noise and their effect on 16 QAM transmission. A setup was designed that emulates the return path in a cable network. The ingress interference was measured using a spectrum analyzer connected to a computer, while an oscilloscope with a high sampling rate was employed to measure the pulse noise. Then, the authors measured the effect of ingress interference and impulse noise on the performance of QAM transmission. The results were displayed in graphs through a developed program on MATLAB software. These diagrams helped cable providers choose appropriate frequencies with ingress interference and direct pickup noise.

In the Netherlands, Dutch Radiocommunications Authority and the University of Twente investigated the interference of the LTE system with the cable television system in the 800 MHz band [26], [27]. The interference probability in overall households was low. They considered two possible scenarios for interference to occur: One, if someone was making a phone call in the 800 MHz band on the same channel used to show a TV program simultaneously. In the second scenario, an interference probability of about 48% will cause the digital TV signal to be distorted if there is co-channeling. However, this percentage varies depending on many factors. They also suggested measures for individual households to reduce the interference probability more, such as using good quality cables, good plugs in the home, and not using the cellphone near cable TV system equipment in the house.

In the United Kingdom [28], the Cobham Technical company investigated potential interference from LTE user equipment operating in the 800 MHz spectrum into most types of STBs and CMs available in the local market. The researchers investigated the following factors: for STBs, they evaluated the quality of the received digital signal using the picture failure standard and the effect of varying the cable signal strength in the STBs and changing the resource block allocations used in the UE signal. For CMs, they used a data failure criterion to assess the quality of the received digital signal. Additionally, they examined the hardware design of most tested STBs and CMs. The results of the tested STBs showed that most of them were affected by interference from the maximum LTE UE power broadcast at a separation distance of 1 meter. However, a 1 dB increase in the cable signal level in the STBs significantly reduced the interference. On the other hand, the partial resource block allocation in the LTE UE signal increased the interference level in the STB. While the results of the tested CMs revealed that all of them were affected by interference from the maximum LTE UE power broadcast at a separation distance of 1 meter. Finally, the hardware design examination revealed that all the evaluated STBs had a metal design with perforations that enabled undesirable signals to pass into the sensitive circuits. The design quality of the examined CMs varied based on their components and materials. The influence of designs was evident in the test findings.

The authors of [29] performed an experimental analysis to cancel interference of a wide 20 MHz LTE signal with cable TV systems in 700 MHz band. A new design was suggested to reduce LTE interference in cable TV transmissions by adapting the phase and amplitude of multi-paths LTE signals in broadband.

In Croatia [30], in-house measurements were performed in a controlled environment to prevent both ingress and egress interference. Firstly, a suitable setup was configured to examine the effect of varying the cable TV system's modulation levels and signal strength. The researchers determined levels of immunity enhancement in the cable TV system interfered by the LTE-uplink signal at 795 MHz when the QAM modulation level decreases from 256 QAM to 64 QAM or when the input power level increases by 10 dB. They found that the increase in input power level was more beneficial. Secondly, they examined the vulnerability of the passive and active elements of the cable TV system to interference. The tested elements were: three different cables, four connector configurations, additional passive elements, and different STBs. The test results were evaluated according to the modulation error ratio (MER in dB) versus cable signal level (dBμV). According to the test findings, the researchers could classify the equipment based on its degree of sensitivity to interference. The critical components were low-shielding cables and connectors, whereas the STB was the most vulnerable to interference.

In Hungary [31], the authors measured the Shielding Effectiveness (SE) of passive cable network components, such as various cables, multi-taps, and line splitters. Additionally, they examined the impact of various factors on the measurement results, such as cases of twisted cables, damaged splitters, and loose taps. The test findings determined that the proper initial cable TV network installation, the employment of high-quality components, and appropriate shielding offer sufficient protection against interference caused by unwanted signals in the broadband. However, some faults that develop over time and produce a rise in harmful interference, such as destroyed or twisted shielding or loose connectors, need continuous maintenance.

Later, the Hungarian authorities carried out internal technical work collaborating with the mobile operator and cable TV provider to avoid harmful interference between the network elements. They determined the minimum required physical distances between the cable TV elements and MFCN-base station and MFCN-user equipment to prevent mutual interference according to modulation level, DOCSIS versions, and the shielding effectiveness of cable TV elements. According to the results of this technical work, the required minimum physical distances to prevent mutual interference between the cable TV system and MFCN increase with the decrease in the shielding effectiveness of cable TV elements and with an increase in the QAM modulation level. However, employing the most recent DOCSIS version decrease the minimum separation distance compared to the prior DOCSIS version, despite using the same QAM modulation level in the previous one.

V. PROCEDURES TO MITIGATE THE MUTUAL INTERFERENCE

The first step in mitigating interference is to create a team that is fully aware of interference issues in both cellular operators and cable TV providers. On the one hand, this team is responsible for proper initial installation and ongoing maintenance, and it is striving to enhance community awareness of interference issues on the other hand. Later in actual practice, if a problem occurs that causes interference, the party responsible for the solution is determined based on the cause, whether it is the cable provider, the cellular company, the device manufacturer, or even the user in some erroneous practices.

As the digital dividend bands are now licensed to the MFCN, leakage from the cable networks to authorized users is forbidden, and cable TV providers may face exorbitant fines from spectrum regulators. Moreover, any leakage points of the cable TV network are potential ingress points causing signal degradation for their subscribers. As a result, cable TV providers must comply with acceptable signal leakage and shielding effectiveness limits as per SCTE, besides continuous monitoring and measuring the cumulative leakage level over time.

In general, resolving radio-signal interference issues necessitates addressing one or more of the following components: an item susceptible to interference, a source of unwanted signal energy, and a propagation path.

The suggested procedures to mitigate the interference concerning the first component that is vulnerable equipment to interference include correcting installation errors of outdoor and indoor cable TV networks, maintaining the shielding effectiveness, and using improved cabling inside the home and STBs/CMs with enhanced immunity.

Regarding the source of noise, we suggest the trade-off between using a specific QAM modulation level and input power level in the cable television system to achieve an acceptable degree of immunity enhancement in the cable TV system while being exposed to MFCN interference. On the other hand, reducing the power levels of MFCN in the adjacent bands is beneficial. Additionally, cable TV providers should consider implementing forward error correction algorithms to improve the system's immunity to interference while the modulation level is upgraded.

Lastly, regarding the propagation path of unwanted signals, one of the recommendations for mitigating mutual interference is to adjust the antenna's gain, tilting, and radiation pattern at the MFCN base station [1]. These factors impact how QAM leak levels affect RSSI and, therefore, the MFCN performance. The electrical or/and mechanical tilt changes the antenna radiation pattern to optimize the performance of the MFCN under certain conditions. Figure 5 displays that the antenna should be oriented vertically, and the radiation pattern should have a narrow width of no more than 15 degrees [1]. Besides, the reduction in antenna gain caused by variations in the vertical antenna pattern near the cellular tower reduces the effect of QAM signal leakage located below the main lobe antenna pattern.

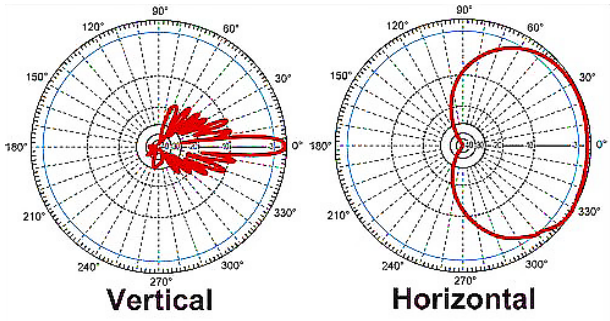


Fig. 5. The radiation pattern and tilting of the antenna

This tuning has the benefit of reducing the influence on the base station while also increasing sensitivity gain by 20 dB or more, considering the difference in effect based on the distance between the leak location and the antenna's main lobe or side lobes [1]. On the other hand, when implementing the Multiple-Input and Multiple-Output (MIMO) system in the base station, spatial multiplexing and diversity technologies will assist limit the impact of cable TV signal leakage.

Although the recommendations mentioned above regarding the propagation path of unwanted signals seem theoretically possible, in practice, they are very challenging for the following reasons: Many other variables might affect the intensity of the leak QAM signal that reaches the MFCN base station, such as physical obstructions (terrain, buildings, trees, etc.). Therefore, it is impractical to utilize fixed guidelines to avoid leakage levels that may interfere with the MFCN base station. On the other hand, those suggestions are not in line with the current practices of mobile operators. Many locations currently do not have MFCN antennas installed on towers, as in the case of microcells, have antennas installed on the sides of buildings or utility poles. In these situations, the MFCN antennas may be near the cable TV infrastructure; thus, even minor QAM leakage might interfere with the MFCN signal.

We also advocate preventative measures to avoid future leakages, such as practical and professional training on proper initial installation, monitoring service quality, periodic maintenance, and collaboration between MFCN operators and cable TV providers.

There is a proposal to provide appropriate instructions to the subscribers to avoid future problems (for example, do not use cellular devices near cable equipment to avoid poor cable TV service quality). Prohibiting the sale of unshielded devices and issuing legal rules regulating customer behavior could also help mitigate the harmful interference of cable TV services and MFCN in digital dividend bands.

VI. CONCLUSION

This article discussed the emerging challenges in mutual interference between cable TV systems and MFCN operating in the digital dividend bands. We presented the 700 MHz spectrum sharing agreement in Hungary as an example to illustrate this issue.

We started this article by explaining the structure of a cable television system and a brief description of the digital signal transmission process from the operator to the subscribers' homes. Next, we covered all the interference factors related to MFCN operating in digital dividend bands. Then, we outlined the typical sources, influences, and primary indicators of ingress and egress interference for cable TV providers and cellular carriers. Following that, we discussed the methods for detecting and evaluating mutual interferences previously employed in several European nations and the United States, whether in the laboratory or the field. Finally, based on prior research findings, we categorized all measures and procedures that may be used to avoid or mitigate mutual interference according to the three main components: an item susceptible to interference, a source of unwanted signal energy, and a propagation path.

Ultimately, despite progress in raising awareness of potential interference issues in the cable and wireless industries, resolving mutual interference will not be quick or straightforward. However, proactive maintenance and good engineering practices can help.

Moreover, the ITU is currently working on sharing and compatibility studies in the new frequency range 470-694 MHz, implying that more frequencies below 700 MHz would be allocated for mobile broadband services. As a result, MFCN and cable providers will face new challenges due to this new frequency range, emphasizing the significance of addressing ingress and egress issues.

REFERENCES

[1] Engineering committee and network operations subcommittee, "Technical Report UHF Leakage, Ingress, Direct Pickup", SCTE Technical Report, The Society of Cable Telecommunications Engineers (SCTE), 209, Exton, PA, USA, 2015, pp. 1–80. [Online]. Available: https://www.scte.org/documents/203/SCTE-209-2015-1575563912519.pdf

[2] International Telecommunication Union (ITU), Geneva, Switzerland. [Online]. Available: https://www.itu.int/

[3] P. Lamy, "Results of the work of the high-level group on the future use of the UHF band (470-790 MHZ)", Report to the European Commission 1, 2014, pp. 1–34 [Online]. Available: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=6721

[4] The European Parliament and the council of the European Union, "Decision (EU) 2017/899 of the European Parliament and of the Council of 17 May 2017 on the use of the 470-790 MHz frequency band in the Union", Strasbourg, France, 2017. pp. 1–7 [Online]. Available: http://data.europa.eu/eli/dec/2017/899/oj

[5] NMHH, "National roadmap for the utilization of the VHF III (174–230 MHz) and the UHF (470–790 MHz) frequency bands; the future of digital broadcasting and mobile broadband frequency use options", National Media and Infocommunications Authority (NMHH), Budapest, Hungary, published on 20 August 2017, pp. 1–22. [Online]. Available: https://english.nmhh.hu/document/190192/uhf_vhf_3_national_roadmap_eng.pdf

[6] T. I. Unger, "Frequency co-ordination: MFCN Agreements in Hungary; Tasks and Challenges in the Present and for the Future", National Media and Infocommunications Authority (NMHH), Budapest, Hungary, published on 2 July 2020, pp. 1–27. [Online]. Available: https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/Events/2020/Spectrum_EUR_CIS/Tamas%20Unger%20(1).pdf

[7] 76.605 Technical standards, title-47, section CFR § 76.605, Cable Rules | Federal Communications Commission, 2018. [Online]. Available: https://www.ecfr.gov/current/title-47/section-76.605

[8] L. Harte, "Introduction to Cable TV (CATV) Systems, Services, Operation, and Technology", 3rd edition, Discovery Press, 2017. ISBN: 978-1932813180, [Online]. Available: https://www.amazon.com/Introduction-Cable-CATV-Operation-Technology-ebook/dp/B06XQ7C497

[9] ITU, "The future of cable TV: Trends and implications", International Telecommunication Union (ITU), Geneva, Switzerland, 2018. [Online]. Available: http://handle.itu.int/11.1002/pub/81216af5-en

[10] ETSI Technical Specification, "Digital Video Broadcasting (DVB); Implementation Guidelines for a second-generation digital cable transmission system (DVB-C2)", ETSI TS 102 991 V1.3.1 (2016-01), 2016. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/102990_102999/102991/01.03.01_60/ts_102991v010301p.pdf

[11] L. Chiariglione and C. A. Szabó. "Multimedia Communications: Technologies, Services, Perspectives", Infocommunications Journal, 6(2), 2014, pp. 27–39. [Online]. Available: https://www.chiariglione.org/publications/papers/infocom/Pages%20from%20InfocomJ_2_komplett.pdf

[12] IEC Technical Specification, "Cable networks for television signals, sound signals and interactive services - Part 1: System performance of forward paths", IEC 60728-1: (May-2014), 2014. [Online]. Available: https://cdn.standards.iteh.ai/samples/19970/f1db2762aa2c43e29850d009bb73ade1/IEC-60728-1-2014.pdf

[13] W. Fischer, "Digital Video and Audio Broadcasting Technology. A Practical Engineering Guide", Springer, Fourth Edition, 2020. DOI: 10.1007/978-3-030-32185-7

[14] M. Toy, "Cable Networks, Services, and Management", John Wiley & Sons, Hoboken, New Jersey, 2015. ISBN: 1118837592, 9781118837597, [Online]. Available: https://www.wiley.com/en-ie/Cable+Networks%2C+Services%2C+and+Management-p-9781118837597

[15] CEPT, "To develop harmonised technical conditions for the 694 -790 MHz ('700 MHz') frequency band in the EU for the provision of wireless broadband and other uses in support of EU spectrum policy objectives", CEPT Report 053, 2014. [Online]. Available: https://docdb.cept.org/download/86

[16] ECC, "Harmonised technical conditions for mobile/fixed communications networks (MFCN) in the band 694-790 MHz including a paired frequency arrangement (Frequency Division Duplex 2x30 MHz) and an optional unpaired frequency arrangement (Supplemental Downlink)", ECC Decision (15)01, 2015. [Online]. Available: https://docdb.cept.org/download/1502

[17] ETSI Technical Specification, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (3GPP TS 36.101 version 15.9.0 Release 15)", ETSI TS 136 101 V15.9.0 (2020-02), 2020. [Online]. Available: https://https://www.etsi.org/deliver/etsi_ts/136100_136199/136101/15.09.00_60/ts_136101v150900p.pdf

[18] P. Denisowski, "Recognizing and resolving lte/catv interference issues", White Paper, *Rohde and Schwarz*, 2011. [Online]. Available: https://www.mobilewirelesstesting.com/wp-content/uploads/2017/11/LTEInterferenceIssues_WP.pdf

[19] S. Windle, R. Hranac, N. Segura, K. Couch, M. Darragh, G. Tresness, and D. Howard, "Operational Practice for Minimizing Signal Leakage in the UHF Spectrum", *JOURNAL OF NETWORK OPERATIONS*, 2016, pp. 55–78. [Online]. Available: https://www.scte.org/documents/3593/SCTE-ISBE-NOS_Journal_V1N1.pdf

[20] P. Denisowski, "Spectrum Analyzers vs. Monitoring Receivers", *Rohde & Schwarz*, 2017. Online available: https://silo.tips/download/spectrum-analyzers-vs-monitoring-receivers-paul-denisowski-application-engineer

[21] ARCOM digital, "LTE interference and CATV", 2014, pp.1-63 [Online]. Available: https://pdf4pro.com/cdn/lte-interference-and-catv-2f0440.pdf

[22] R. Hranac, G. Tresness, "Another Look at Signal Leakage; The Need to Monitor at Low and High Frequencies", A Technical Paper prepared for the Society of Cable Telecommunications Engineers, *SCTE Cable TEC*, EXPO12, Orlando, USA, 17-19 October 2012, pp. 1–32. [Online]. Available: https://www.arcomdigital.com/wp-content/media/hranac-tresness-scte-white-paper.pdf

[23] M. Sadowski, "Leakages from devices of CATV system", *IFAC-Papers Online*, 51.6, 2018, pp. 490–495. DOI: 10.1016/j.ifacol.2018.07.108

[24] J. Wider, and R. Hranac, "Operational Practice for Identifying, Locating and Mitigating UHF Ingress and Direct Pickup in Cable Networks and Devices", *Journal of network operations*, 2016, pp. 4-26. [Online]. Available: https://www.scte.org/documents/3593/SCTE-ISBE-NOS_Journal_V1N1.pdf

[25] S. Bette, V. Moeyaert, V. Tamgnoue, M. Blondel, P. Mégret, "Comparative analysis of the impact of CATV return path ingress and impulse noises in a 16QAM transmission system", Service d'Electromagnétisme et de Télécommunications, Faculté Polytechnique de Mons, Belgium, 2004, pp. 1–8. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.723.8372&rep=rep1&type=pdf

[26] R. Schiphorst et al., "Analysis of interference to cable television due to mobile usage in the Digital Dividend", Radiocommunications Agency Netherlands, Groningen, Netherlands, 2010, pp. 1–80. [Online]. Available: https://research.utwente.nl/files/5124397/ATEL%20rapport%20kabel%20stoor%20onderzoek%20EN%20DEF%2022072010%20hires.pdf

[27] J. Robijns, and R. Schiphorst. "Interference to cable television due to mobile usage in the Digital Dividend-Analysis", *10th International Symposium on Electromagnetic Compatibility,* IEEE, 2011, pp. 260–265. [Online]. Available: https://ieeexplore.ieee.org/document/6078544

[28] S. Munday, and I. Parker. "Field Tests Investigating the Potential Interference into Cable TV from LTE Deployment in the 800 MHz band", Cobham Technical Services, Report number: 2010-0792, 2010, p. 1–83. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0017/101357/2010-0792_LTE_into_CATV.pdf

[29] J. Chen, G. Wu, X. Si, D. F. Hunter, P. Wensheng, Y. Shen, and S. Shao, "An experimental RF noise cancellation analysis for cable access systems", *IEEE Communications Magazine*, 53(3), 2015, pp. 121–125. DOI: 10.1109/MCOM.2015.7060492

[30] A. Teković, "Performance challenges for LTE deployment in digital dividend in Croatia", Croatia, 2014, pp. 1–44. [Online]. Available: https://www.bib.irb.hr/717201/download/717201.IIR__LTE_in_Digital_Dividend_deployment_challenges_atekovic.pdf

[31] P. Prukner, I. Drotár, M. Liszi, and S. Nagy, "Measurement of the Shielding Effectiveness of Passive Cable Television Elements", *ICEST Conference, Serbia*, 2017, pp. 101–104. [Online]. Available: http://rcvt.tu-sofia.bg/ICEST2017_23.pdf

Hussein Taha is a PhD candidate at the Doctoral School of Multidisciplinary Engineering Sciences at Széchenyi István University in Hungary. He holds a BSc in Communication and Electronics Engineering from Tishreen University in Syria and MSc in Telecommunication Engineering from the same university in 2019. His areas of interest are mobile and wireless communications, and broadcasting.

Prof. Dr. Péter Vári is an Associate Professor in the Department of Telecommunications at Széchenyi István University, the university of Győr in Hungary. He is now Deputy Director-General for Technical Affairs at the National Media and Infocommunications Authority in Hungary. His general interests span the areas of radiocommunications, mobile services, and broadcasting.

Prof. Dr. Szilvia Nagy is a Professor in the Department of Telecommunications at Széchenyi István University, the university of Győr in Hungary. Her research interests include digital image processing, information and coding theory, DSP, EMC, quantum semiconductors.

Effect of the initial population construction on the DBMEA algorithm searching for the optimal solution of the traveling salesman problem

Ali Jawad Ibada[1]*, Boldizsár Tüű-Szabó[2], and László T. Kóczy[3]

***Abstract*—There are many factors that affect the performance of the evolutionary and memetic algorithms. One of these factors is the proper selection of the initial population, as it represents a very important criterion contributing to the convergence speed. Selecting a conveniently preprocessed initial population definitely increases the convergence speed and thus accelerates the probability of steering the search towards better regions in the search space, hence, avoiding premature convergence towards a local optimum. In this paper, we propose a new method for generating the initial individual candidate solution called Circle Group Heuristic (CGH) for Discrete Bacterial Memetic Evolutionary Algorithm (DBMEA), which is built with aid of a simple Genetic Algorithm (GA). CGH has been tested for several benchmark reference data of the Travelling Salesman Problem (TSP). The practical results show that CGH gives better tours compared with other well-known heuristic tour construction methods.**

***Index Terms*—Traveling Salesman Problem, Discrete Bacterial Memetic Evolutionary Algorithm, Genetic Algorithm, Nearest Neighbor heuristic, Second Nearest Neighbor heuristic, Alternating Nearest Neighbor heuristic, Circle Group Heuristic.**

I. INTRODUCTION

THE Traveling Salesman Problem (TSP) is one of the most prominent members of the rich set of well-known combinatorial optimization problems with real life application potential. It is a Nondeterministic Polynomial hard (NP-hard) problem [1]. Given a set of cities (graph nodes) along with the costs of travel between each pair of them (the costs or lengths assigned to the edges), the TSP goal is to find the cheapest (shortest) way of visiting all the cities exactly once, and then returning to the starting point [1][2].

It must be realized that NP-hard problems are intractable (see e.g. [3]) and thus, there is no algorithm that gives guaranteed exact solution for them within a predictable time, nevertheless, there may be partially successful and guaranteed, but approximate solution methods constructed. Over the decades, there have been numerous approaches proposed in order to find the optimum (shortest, least cost) route. They may be classified to three classes: exact solution methods, algorithms for approximate solution and heuristic approaches [4][5]. The Christofides algorithm is the most well-known approximation algorithm [6], which may be however, rather imprecise as the guaranteed solution may be maximally 50% greater than the global optimum. The most efficient heuristic solver so far is Helsgaun's implementation of the classic Lin-Kernighan heuristic [7]. Many meta-heuristic researches have been published to find optimal or near-optimal solutions for the TSP; such as the Genetic Algorithm [8], the Ant Colony Optimization Algorithm [9], the Bacterial Evolutionary Algorithm (BEA) [10], the Particle Swarm Algorithm [11], Artificial Bee Colony [12], and their respective memetic versions [13]. In the next, we will only deal with a chosen, very efficient heuristic algorithm, the discrete memetic version of the Bacterial Evolutionary Algorithm (DBMEA) [14].

Returning to the matter of initial population selection, let us summarize that each initial population represents a feasible solution which is then subsequently improved over the course of several iterations through a heuristic (e.g., evolutionary) process [15]. The quality of the initial population of an evolutionary algorithm is rather important as it affects the search for the next (often numerous) generations, and has a significant influence on the quality of the final solution [16][17]. Improvement efforts on the initial population have shown to be effective in reducing the number of generations utilized while also improving the quality of the solution [18-23].

II. THE TRAVELING SALESMAN PROBLEM

The task of the TSP is to find a route through a given set of cities with the shortest possible length (cost). Mathematically, it means to find the shortest Hamiltonian tour in a graph [1].

$$G_{TSP} = (V_{cities}, E_{connections})$$
$$V_{cities} = \{v_1, v_2, \ldots, v_n\}, E_{conn} \subseteq \{(v_i, v_j) \mid i \neq j\}$$
$$C: V_{cities} \times V_{cities} \rightarrow R,\ C = (c_{ij})_{n \times n} \quad (1)$$

Where: C is called the cost matrix, c_{ij} represents the cost of going from city i to city j.

The goal is to find an optimal permutation of vertices (p_1, p_2, p_3, . . ., p_n) that gives the minimum total cost [24].

$$\text{Minimize}\left(\left(\sum_{i=1}^{n-1} Cpi, pi-1\right) + Cpn, p1\right) \quad (2)$$

In general, the TSP can be classified into two different kinds, the Symmetric Travelling Salesman Problem (STSP) and the Asymmetric Travelling Salesman Problem (ATSP).

[1,3] Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary.

[2,3] Department of Information technology, Széchenyi István University, Gyor, Hungary.

DOI: 10.36244/ICJ.2022.3.9

The distance between cities A and B is identical to the distance between cities B and A in the STSP. However, with the ATSP, it is possible to have two different costs or distances between two cities, depending on the direction. This may be realistic if, e.g., the altitude of the two places is different, and thus climbing would take more costs (time, fuel, etc.) than descending on the same route. Hence, the number of tours in the ATSP and STSP on n vertices is (n-1)! and (n-1)! /2, respectively [25].

As mentioned above, due to the combinatorial complexity of the TSP, in practice, for larger instances (graphs), only approximate and/or heuristic procedures are applied in searching for the solution [26].

The TSP can be applied to a wide range of real life discrete optimization problems, especially in logistics, planning, and microchip manufacturing [24].

III. THE DISCRETE BACTERIAL MEMETIC EVOLUTIONARY ALGORITHM

In this section, a heuristic optimization approach will be briefly introduced which has proven rather efficient for a wide family of TSP related problems (mostly, extensions of the original TSP towards more realistic – and more complex – cases), while having two additional advantageous properties: the algorithm is generally applicable with high efficiency, and the runtime is rather predictable in terms of the size of the problem (number of nodes in the graph) [27]. Thus, even though for the basic TSP there is a better heuristic known, but it is very tailor made and not applicable for any other related optimization, and so we intend to investigate a complex approach which will be later extendable for other similar NP-hard problems.

Memetic algorithms extended the idea of using evolutionary algorithms for global search with nested local search methods, originally coming from more traditional mathematics. In each iteration for the individuals a local search step is applied [13]. DBMEA is a memetic algorithm that combines the very efficient Bacterial Evolutionary Algorithm as a global optimization with a simple combinatorial local search technique [25]. The drawbacks of both techniques, namely, the tendency to get stuck in a local optimum typical for the traditional (local) search techniques; and the very slow convergence speed of the outer (global) search cycle, are eliminated with this combined method. The evolutionary algorithms examine the global search space, and thus they only give a quasi-optimal solution because of their relatively slow convergence speed. Local search methods search only a certain neighborhood of the current candidate for solution, so they always converge to the closest local optimum; while their convergence speed is much faster. DBMEA was found to be rather efficient in solving a series of discrete nondeterministic polynomial-time hard optimization problems [25][26]. The DBMEA has four stages: initial population creation, bacterial mutation (coherent segment mutation and loose segment mutation), local search (2-OPT and 3-OPT), and gene transfer which is cyclically repeated.

1) Initial population creation step

Efficiency of an optimization algorithm is judged by its accuracy and speed. So, it is important to reach the (quasi-optimum) as fast as possible. The creation of the initial population can be crucial in reaching acceptably accurate solutions faster. In the literature, the initial population is often created randomly, but there are some deterministic approaches as well. Random creation guarantees the uniform distribution of the population in the search space. In the work of our group, several deterministic approaches were investigated [28]. The following heuristic construction algorithms are worth mentioning:

A. Nearest Neighbor (NN) heuristic

It constructs a tour in which, in the next step, the nearest unvisited city will always be visited. NN is easy to implement and fast to execute.

B. Secondary Nearest Neighbor (SNN) heuristic

It always visits the second nearest unvisited city in the next step of the tour.

C. Alternating Nearest Neighbor (ANN) heuristic

It combines the NN and SNN methods, here, the nearest and the second nearest unvisited cities are visited next in an alternating manner.

Among these three, the best convergence speed in most cases was achieved by the NN approach.

2) Bacterial Mutation Step

During this stage, each bacterium in the population is treated separately. A certain number of (identical) clones are made from the original bacterium (N_{clones}). Then, the bacterium and its clones are subdivided into chromosomes with a fixed length (I_{seg}). There are two semantic types of chromosomes (segments): coherent segments and loose segments. One from the segments of the bacterium is selected randomly and is modified in a clone, while the same gene in the original bacterium remains unaltered. So, it goes on with all the other clones as well. There is a particular clone in DBMEA; it contains the reverse order of the selected segment. Figure 1 shows the process of the clone creation.

The next step is the evaluation of the fitness values. As in the case of the TSP, traditional mathematics offers a possibility to determine the lower bound of the route length, based on the spanning tree of the whole graph, in this case the fitness function is obtained from the difference of the candidate solution individual from the theoretical lower bound, thus, the accuracy of all this way obtained clone bacteria (including the original). If one of the clones is better than the original bacterium, the mutated segment of the better clone is copied back to the original bacterium and to all the other clones. This process is consecutively applied until all the genes of the original bacterium have been mutated.

3) Local Search Step

During this step, individual improvement is carried out. The approach uses the exhaustive investigation of rearranging

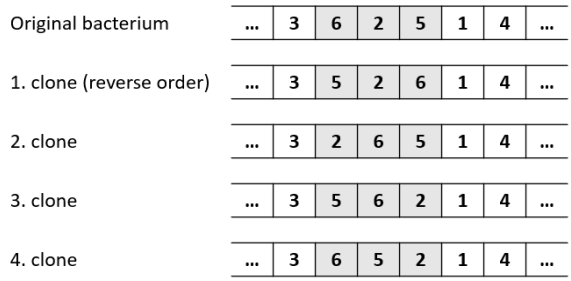


Fig. 1. Clones creation in mutation stage

DBMEA, 2-opt and 3-opt could improve the individual with bounded size sub-graphs, optimizing them locally. Our group has investigated 2-opt and 3-opt local search for reasonable time, and we found that subsequently carried out 2-opt and 3-opt local search cycles are useful, and so, they are applied [28].

A. 2-opt local search

To shorten the TSP tour, in this simple method, two edge pairs in the original graph are exchanged. Assume we have two edge pairs, *AB* and *CD*; then these two will be replaced with *AC* and *BD* edges, resulting in a new potential tour. The truth of the following inequality is examined in the case of the new tour:

$$|AB| + |CD| > |AC| + |BD| \quad (3)$$

If the inequality is true, the edge pairs are swapped; the *AB* and *CD* edges are removed from the graph and replaced with *AC* and *BD*, as illustrated in Figure 2. The 2-opt move requires reversing one of the sub-tours between the initial edges. This iterative process is terminated if no further improvement can be made.

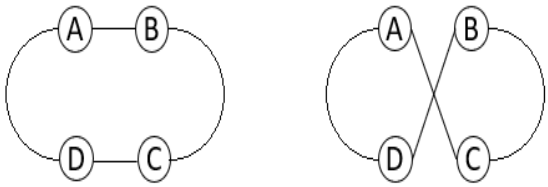


Fig. 2. Example for 2-opt local search

B. 3-opt local search

In this method, three edges are replaced with three others, producing eight alternative ways to reconnect the TSP tour, however, four of them have already been checked as they distort into 2-opt steps, therefore they are not considered here. The possible new replacements in the 3-opt local search are shown in Figure 3.

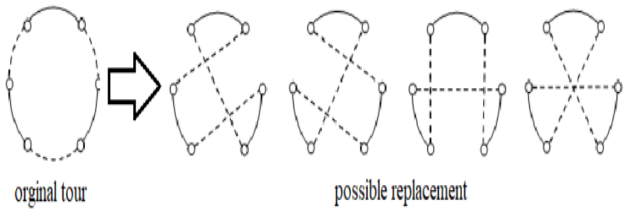


Fig. 3. Possible replacement of 3-opt local search

4) Gene transfer

In this stage, the population is initially sorted in decreasing order according to their fitness values, which are then sorted and separated into two (a superior and an inferior half). The operator repeats the following N_{inf} times: it picks one random bacterium (source bacterium) from the superior half and another random bacterium (destination bacterium) from the inferior part. Then it transfers some randomly picked segments with a pre-defined length ($I_{transfer}$) from the source bacterium to the destination bacterium. The bacterium length will not alter since the duplicate occurrences will be removed. Figure 4 shows the segment transfer in the gene transfer stage.

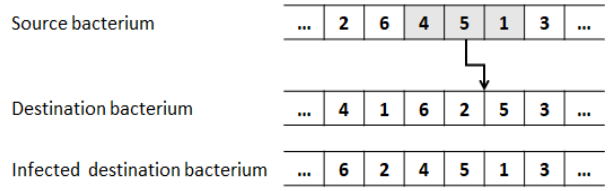


Fig. 4. Gene transfer

IV. THE GENETIC ALGORITHM

Genetic Algorithm (GA) is the most prototypical evolutionary algorithm, which is very widely used in the solution of simpler heuristic optimization problems, and which has many standardized toolbox type implementations. It uses a stochastic search algorithm imitating in a simplified way the natural selection process of living organisms and copying natural genetics. The original Bacterial Evolutionary Algorithm was created by enhancing and modifying some ideas within the GA [29], this way essentially speeding up the convergence. Its continuous memetic extension and its discrete versions were proposed by our group and were applied for a variety of optimization problems rather successfully. The original GA has five steps: Initial population creation, Candidate selection, Crossover, Mutation, and Fitness function evaluation [30-33].

1) Initial population creation

The size of the population varies depending on the problem, but it usually encompasses several hundred or thousands of potential solutions. The starting population is frequently created at random, providing for a wide variety of possible solutions (however, lacking the opportunity of a deterministic improvement approach already applied in this step).

2) Candidate selection

The best offspring solutions must be chosen to be parents in the new parental population in order to facilitate convergence towards optimum solutions. Because of this, an excess of offspring solutions is developed, and the best are chosen in order to make progress toward the optimum. This selection method is based on the fitness values in the population.

3) Crossover

Crossover is a function that permits the genetic material of

two or more solutions to be combined. The reason for such an operator is that both strings might represent successful components of solutions that, when combined, outperform their parents. This operator may easily be expanded to more points, where the solutions are alternately separated and rebuilt. This is not unlike the Gene transfer step in the DBMEA.

4) Mutation

Mutation is the second main character in Genetic Algorithms. Mutation operators change a solution by disturbing them. Random alterations are the foundation of mutation. Mutation is the part of the GA which is related to the "exploration" of the search space. It has been discovered that mutation is required for GA convergence. There are different operators for mutation such as: bit flip mutation, random resetting, swap mutation, scramble mutation, and inversion mutation. Again, here, the Mutation step of the DBMEA has its "ancestor".

5) Fitness function evaluation

In this step, the phenotype of a solution is evaluated. The fitness function measures the quality of the solutions that is generated by the GA. The proper design of the fitness function is part of the overall modeling process of the overall optimization approach. The practitioner may have an influence Genetic Algorithms by designing choices of the fitness function and thus guiding the search.

V. THE PROPOSED TOUR CONSTRUCTION HEURISTIC

We proposed a novel approach that quasi-optimizes, but definitely improves, the initial population for the TSP, from the point of view of applying the DBMEA on this quasi-optimized initial population; by introducing the novel idea of applying the concept of "neighborhood circle" (NC). The NC has a pre-specified radius which will limit and speed up the search for the best possible initial population candidate. The new heuristic method is called Circle Group Heuristic (CGH). CGH is built into the Discrete Bacterial Memetic Evolutionary Algorithm (DBMEA), as its first step, this way increasing the efficiency of this memetic meta-heuristic algorithm that has already proven rather efficient in handling the optimization of TSP type tasks. Next, the CGH method will be explained in detail.

1) The CGH construction step

Starting at the initial node, City 0, which represents the center of the first circle, a circle is drawn with a given radius R. In the first step, the closest unvisited city within the circle will be marked for visit. The tour continues at the next unvisited city within the circle, until all nodes within the circle have been included in the tour. In the subsequent step, the node/city outside the circle, which is the closest to the last visited node/city, is marked as the next city in the tour. This new city on the outside of the circle will become the center of a new circle, and the algorithm starts again as in the case of the first circle. So, on it goes until all the cities have been visited exactly once. A simple tour created by the CGH is shown in Figure 5.

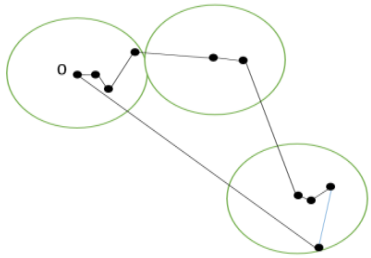


Fig. 5. Example for a simple CGH tour

A. Brute force optimization of the radius

We tested the new method on benchmarks of national TSP instances up to 10639 nodes [34]. In every case, an exhaustive search for the optimal radius was done in discrete steps by testing range of integer numbers [1-100] with brute force method. With these series of tests we have established that the most effective radius of the circle in the CGH generating the best initial population in the sense that the subsequent steps of the DBMEA result in the best approximation in the shortest runtime, is different from graph to graph. Examining the road network graphs of number of countries in this benchmark, with the respective city and road networks assigned with costs obtained from the road section lengths of the respective benchmark data, we found that these graphs differ from each other essentially in the behavior, because of different topologies and sizes of these road networks.

B. GA optimization of the radius

Later, we attempted to find the optimal radius for the CGH by applying genetic algorithm, thus allowing a continuous range for the radius. The GA is a standard toolbox, which is used one independent variables for the fitness function to return the optimal radius, and the test the range of rational numbers [1 – 100]. All the other parameters are defaults of the toolbox [35].

In the next Section, the results obtained by GA optimization of the radius method is presented and evaluated.

VI. RESULTS

The novel CGH tour construction was tested on more than 25 national TSP instance benchmarks up to 10639 nodes. In all cases, the initial population individuals obtained by the above mentioned three heuristics in the previous investigations, NN, SNN, and ANN, were compared with the ones got by the new CGH approach, the radius of CGH is calculated in both methods: simple exhaustive search and GA. In the case of all thus optimized initial individuals, the DBMEA method was applied for solving the respective TSP optimization task. Table 1 shows the tour costs of the optimal tours in the case of the initial populations generated by the previous three older approaches and both sub-versions of the CGH.

TABLE I
THE LENGTH OF THE DETERMINISTIC INITIAL INDIVIDUAL.

Country	Number of cities	NN	SNN	ANN	Exhaustive search		Genetic Algorithm	
					R	CGH	R	CGH
Djibouti	38	9748.946	13509.088	10474.948	78-100	8306.575	79.834	8306.575
China	70	2570.329	4148.675	3565.305	18	2267.889	21.137	2263.575
Burma	80	5477.026	8674.245	6079.910	98-100	4526.302	98.272	4526.302
Qatar	194	11892.888	18980.443	17199.801	9	11649.869	56.280	11255.296
Uruguay	734	102594.358	165796.643	130793.380	57-59	95536.209	54.832	95461.998
Zimbabwe	929	117733.696	200063.995	160430.512	10	114813.039	48.955	114484.256
Luxembourg	980	14212.721	26240.107	20397.885	8	13995.032	16.254	13958.023
Rwanda	1621	32276.665	68487.437	45630.149	10	31596.476	9.768	31596.476
Oman	1979	120542.129	204249.064	152503.643	13	110747.729	12.688	110029.190
Nicaragua	3496	122412.147	229749.240	179481.992	16	118141.497	15.935	118141.497
Canada	4663	1668707.230	2852242.400	2320011.780	71-74	1603709.500	73.834	1603709.500
Tanzania	6117	501427.829	852834.843	696141.743	16	499513.302	16.567	499513.302
Egypt	7146	222335.231	391416.939	306664.393	4	217487.431	3.981	217487.431
Yemen	7663	298953.459	523144.599	417814.532	5	298150.565	5.185	297972.031
Panama	8079	146660.520	277856.032	210793.510	16	142277.699	15.984	141949.651
Ireland	8246	259165.057	421610.120	350446.111	12	255167.585	11.971	255167.585
Argentina	9152	1034964.600	1951034.190	1527346.320	16	1034084.030	15.635	1034084.030
Japan	9847	625031.710	1104954.110	909941.924	2	624849.337	2.215	624849.337
Greece	9882	391415.926	637638.279	523514.157	14	384948.298	13.158	384900.881
Kazakhstan	9976	1346903.560	2320848.050	1863893.670	13	1325094.710	13.058	1325094.710
Finland	10639	657774.773	1081658.770	889702.220	20	649477.672	20.158	649477.672

In the initial population level, the CGH produced on average 4%, 44%, and 30% shorter tours than NN heuristic, SNN heuristic, and ANN heuristic respectively. While the best tour for CGH was 17% shorter than NN heuristic in Burma (80 cities), 54% shorter than SNN heuristic in Rwanda (1621 cities), and 37% shorter than ANN heuristic in China (70 cities).

We made comparisons for the deterministic initial population individuals with the known absolute optimum after each stage of the DBMEA on the China (70 cities) and Oman (1979 cities) instances. The optimal tour for China70 is 2023, while the one for Oman1979 is 86891. The CGH generated for the initial population roughly 20% longer tours than the optimum solutions, making it a useful starting point for a memetic evolutionary algorithm.

The goal of this investigation was to show that introducing a few promising deterministic individuals would improve the efficiency of the evolutionary/memetic algorithm. The convergence speed would be unambiguously faster when using better deterministic initial individuals.

Table 2 illustrates the tour construction heuristics run times for China70 and Oman1979. It shows that the CGH not only provides better tours, but also does it faster than the NN, SNN, and ANN heuristics previously utilized. Tour costs of initial individuals following each stage of DBMEA are illustrated in Table 3.

TABLE II
THE TOUR CONSTRUCTION HEURISTICS RUN TIMES

Instance	China(70)	Oman(1979)
NN	0.002 sec	1.38 sec
SNN	0.002 sec	2.32 sec
ANN	0.002 sec	1.563 sec
CGH	**0.001 sec**	**0.839 sec**

TABLE III
TOUR COSTS OF INITIAL INDIVIDUALS FOLLOWING EACH STAGE OF DBMEA

DBMEA stage	Instance	NN	SNN	ANN	CGH	Time (Sec)
Mutation stage	China	2471.1391	4021.18307	3052.31702	**2221.05428**	0.008
	Oman	120207.162	195634.363	149871.306	**108658.897**	3.873
Local search	China	2038.36554	2081.70514	2059.382	**2024.74914**	0.106
	Oman	91178.9462	93608.4377	91659.2708	**90324.38**	5018.752
Improved version of local search	China	2118.85481	2126.6373	2125.59152	**2085.91453**	0.357
	Oman	**93182.4458**	102739.247	100512.507	94981.3311	197.564

Figures 6, 7, and 8 show the graphs of the deterministic individuals after the initial populations stage, bacterial mutation stage, and local search stage respectively for China70.

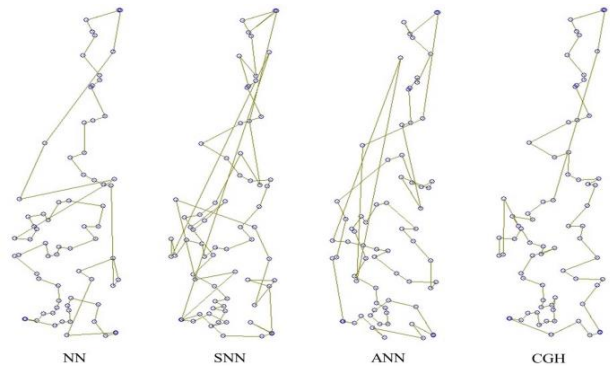


Fig. 6. Initial individual tours of China70

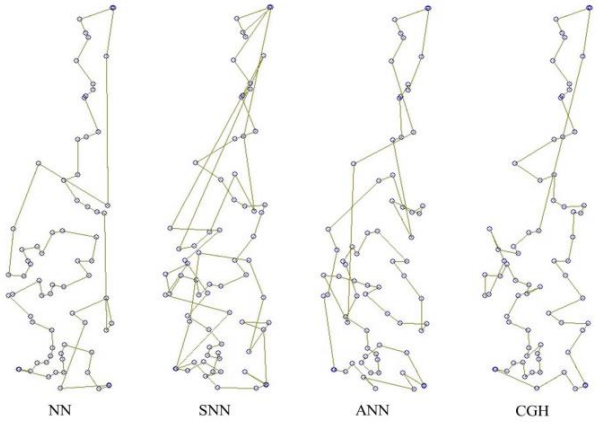


Fig. 7. Tours of the predefined individuals for China70 after bacterial mutation stage

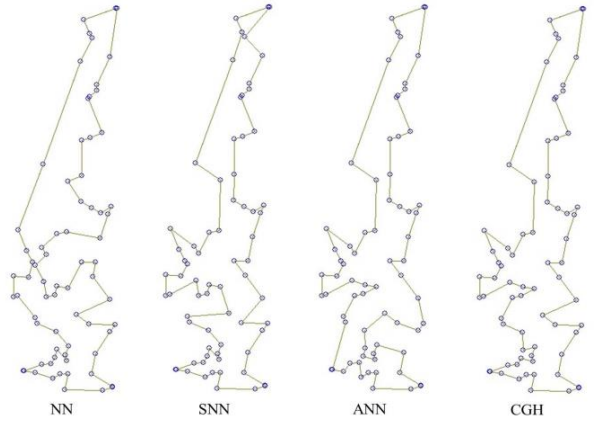


Fig. 8. Tours of the deterministic individuals for China70 after local search stage

VII. Conclusion

In this paper, we proposed a new and efficient initial tour construction method for DBMEA to solve TSP. We had studied and compared the other three known deterministic tour construction heuristics of DBMEA with the here proposed CGH algorithm, and we found that CGH gives better results in all DBMEA stages. During the investigations of the population's behaviors in bacterial mutation and the local search stages, we conclude that in almost all cases, better initial population individuals will lead to faster convergence speed and better approximation of the optimal tour length. Based on our experiments, we suggest the use of our novel proposed CGH in the initial population creation stage of the DBMEA and very likely, other heuristic optimization algorithms for solving the TSP.

References

[1] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, "The Traveling Salesman Problem: a computational study," Princeton University Press, Princeton, Princeton University Press, 2006, PP 1–8. DOI: 10.1515/9781400841103.541.

[2] Y. Bartal, L.-A. Gottlieb, R. Krauthgamer, "The traveling salesman problem: low-dimensionality implies a polynomial time approximation scheme," SIAM J. Comput, Vol. 45, No. 4, pp. 1563–1581, 2016. DOI: 10.1145/2213977.2214038.

[3] A. V. Aho, J. E. Hopcroft, J. D. Ullman, "The Design and Analysis of Computer Algorithms," Addison-Wesley, 1974.

[4] A. J. Ibada, B. Tüű-Szabó, L. T. Kóczy, "A new efficient tour construction heuristic for the Traveling Salesman Problem," In *2021 5th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI 2021)*, Association for Computing Machinery, New York, NY, USA, 2021. pp. 71–76. DOI: 10.1145/3461598.3461610.

[5] L. T. Kóczy, P. Földesi, B. Tüű-Szabó, R. Almahasneh, "Modeling of Fuzzy Rule-base Algorithm for the Time Dependent Traveling Salesman Problem," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019, pp. 1–6 DOI: 10.1109/fuzz-ieee.2019.8858853.

[6] N. Christofides, "Worst-case analysis of a new heuristic for the travelling salesman problem, Report 388," Graduate School of Industrial Administration, CMU, 1976.

[7] K. Helsgaun, "An effective implementation of the Lin-Kernighan traveling salesman heuristic," European Journal of Operational Research 126, pp. 106–130, 2000. DOI: 10.1016/s0377-2217(99)00284-2.

[8] P. Larrañaga, C. M. H. Kuijpers, R. H. Murga, I. Inza, and S. Dizdarevic. "Genetic algorithms for the travelling salesman problem: a review of representations and operators," Artificial Intelligence Review, Vol. 13, No. 2, pp. 129–170, 1999.

[9] M. Dorigo, L. M. Gambardella, "Ant colonies for the travelling salesman problem," Biosystems, Vol. 43, No. 2, pp. 73–81, 1997. DOI: 10.1016/s0303-2647(97)01708-5.

[10] N. E. Nawa and T. Furuhashi, "Fuzzy system parameters discovery by bacterial evolutionary algorithm," in *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 5, pp. 608–616, 1999. DOI: 10.1109/91.797983.

[11] K. Wang, L. Huang, C. Zhou, W. Pang, "Particle swarm optimization for traveling salesman problem," In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics*, 2003. DOI: 10.1109/ICMLC.2003.1259748.

[12] L. Li, Y. Cheng, L. Tan, B. Niu, "A Discrete Artificial Bee Colony Algorithm for TSP Problem," In *Proceedings of International Conference on Intelligent Computing*, 2011. DOI: 10.1007/978-3-642-24553-4_75.

[13] P. Moscato, "On evolution, search, optimization, genetic algorithms and martial arts towards memetic algorithms" Technical Report Caltech Concurrent Computation Program, Report. 826. Pasadena: California Institute of Technology, 1989.

[14] L. T. Kóczy, P. Földesi, B. Tüu-Szabó, "An effective discrete bacterial memetic evolutionary algorithm for the traveling salesman problem," International Journal of Intelligent Systems, Vol. 32, No. 8, pp. 862–876, 2017. DOI: 10.1002/int.21893.

[15] H. Razip and M. N. Zakaria, "Combining approximation algorithm with genetic algorithm at the initial population for NP-complete problem," 2017 IEEE 15th Student Conference on Research and Development (SCOReD), pp. 98–103, 2017. DOI: 10.1109/scored.2017.8305413.

[16] G. Zhang, L. Gao and Y. Shi, "An effective genetic algorithm for the flexible job-shop scheduling problem," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3563–3573, 2011. DOI: 10.1016/j.eswa.2010.08.145.

[17] R. R. Hill, "A monte-carlo study of genetic algorithm initial population generation methods," Proceedings of the 31st Conference on Winter Simulation: Simulation-a Bridge to the Future, vol. 1, pp. 543–547, 1999. DOI: 10.1109/wsc.1999.823131.

[18] J. Wang, O. K. Ersoy, X. Chen, F. Wang, "A Method of Initial Population Generation of Intelligent Optimization Algorithms for Constrained Global Optimization," International Journal of Hybrid Information Technology, al of Hybrid Information Technology Vol. 10, No. 6, 2017, pp. 47–56. DOI: 10.14257/ijhit.2017.10.6.05.

[19] H. Maaranen, K. Miettinen, M.M. Mäkelä, "Quasi-random initial population for genetic algorithms," Computers & Mathematics with Applications, Vol. 47, No. 12, pp. 1885–1895, 2004. DOI: 10.1016/j.camwa.2003.07.011.

[20] B. Kazimipour, X. Li, A. K. Qin, "A review of population initialization techniques for evolutionary algorithms," 2014 IEEE Congress on Evolutionary Computation (CEC), pp. 2585–2592, 2014. DOI: 10.1109/cec.2014.6900618.

[21] S. Helwig, R. Wanka, "Theoretical analysis of initial particle swarm behavior," in Parallel Problem Solving from Nature-PPSN X. Springer, pp. 889–898, 2008. DOI: 10.1007/978-3-540-87700-4_88.

[22] B. Kazimipour, X. Li, A. Qin, "Initialization methods for large scale global optimization," in Evolutionary Computation (CEC), 2013 IEEE Congress on. IEEE, pp. 2750–2757, 2013. DOI: 10.1109/cec.2013.6557902.

[23] S. Rahnamayan, H. R. Tizhoosh, and M. Salama, "A novel population initialization method for accelerating evolutionary algorithms," Computers & Mathematics with Applications, vol. 53, no. 10, pp. 1605–1614, 2007. DOI: 10.1016/j.camwa.2006.07.013.

[24] E. L. Lawler, J. K. Lenstra, A. H. G. R. Kan, D. B. Shmoys, "The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization" John Wiley & Sons, New York, 1985.

[25] L. T. Kóczy, P. Földesi and B. Tüű-Szabó, "Adiscrete bacterial memetic evolutionary algorithm for the traveling salesman problem," 2016 IEEE Congress on Evolutionary Computation (CEC), pp. 3261–3267, 2016. DOI: 10.1109/cec.2016.7744202.

[26] H. H. Hoos, T. Stutzle, "Stochastic Local Search: Foundations and Applications," Morgan Kaufmann, San Francisco, 2005.

[27] B. Tüű-Szabó, P. Földesi, L. T. Kóczy, "An Efficient Evolutionary Metaheuristic for the Traveling Repairman (Minimum Latency) Problem," International Journal of Computational Intelligence Systems, Vol. 13, No. 1, Pages 781–793, 2020. DOI: 10.2991/ijcis.d.200529.001.

[28] L. T. Kóczy, P. Földesi, B. Tüű-Szabó,"Enhanced discrete bacterial memetic evolutionary algorithm – An efficacious metaheuristic for the traveling salesman optimization," Information Sciences, Vol. 460–461, pp. 389–400,2018. DOI: 10.1016/j.ins.2017.09.069.

[29] N. E. Nawa, T. Furuhashi, "A study on the effect of transfer of genes for the bacterial evolutionary algorithm," 1998 Second International Conference. Knowledge-Based Intelligent Electronic Systems. Proceedings KES'98 (Cat. No.98EX111), pp. 585–590, 1998. DOI: 10.1109/kes.1998.726026.

[30] O. Kramer, "Genetic Algorithm Essentials," Springer International Publishing, 2017. DOI: 10.1007/978-3-319-52156-5_2.

[31] M. Gen, R. Cheng, L. Lin, "Network Models and Optimization Multiobjective Genetic Algorithm Approach," British Library Cataloguing, Springer, 2008.

[32] L. Nagy, "Classical and quantum genetic optimization applied to coverage optimization for indoor access point networks," Infocommunications Journal, Vol. 4, No. 4, 2012.

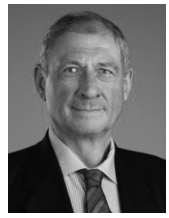
[33] K. Kubícek, J. Novotný, P. Švenda, and M. Ukrop, "New results on reduced-round tiny encryption algorithm using genetic programming," Infocommunications Journal, vol. 8, no. 1, pp. 2–9, 2016.

[34] VLSI TSP dataset, April 2015. http://www.math.uwaterloo.ca/tsp/world/countries. html.

[35] *Genetic Algorithm and Direct Search Toolbox*, User's guide version 1, The MathWorks Inc. 2004.

Ali Jawad Ibada Received his M.Sc. degree in computer engineering from the Middle Technical University, Baghdad, Iraq which was achieved with first-class honors in 2015. Currently, he is a Ph.D. candidate in computer engineering in the Department of Telecommunications and Media Informatics at the Budapest University of Technology and Economics, Budapest, Hungary.
His research interests include evolutionary and memetic algorithms, neural networks, computational intelligence and machine learning algorithms in modeling complex systems.

Boldizsár Tüű-Szabó received the PhD degree in information sciences from the Szechenyi Istvan University, Gyor, Hungary in 2021 and the M.Sc. degree in electrical engineering in 2015. Currenty he is senior lecturer at the Department of Information Technology of Szechenyi Istvan University. His research interests are cover optimization, evolutionary computation and fuzzy modelling.



Laszlo T. Koczy received the M.Sc., M.Phil. and Ph.D. degrees from the Technical University of Budapest (BME) in 1975, 1976, and 1977, respectively; and the D.Sc. degree from the Hungarian Academy of Science in 1998. He spent his career at BME until 2001 and from 2002 at Szechenyi Istvan University (Gyor, SZE). He has been a visiting professor in Australia, Japan, Korea, Austria, Italy, etc. His research interests are fuzzy systems, evolutionary and memetic algorithms, and neural networks, as well as applications in infocommunications, logistics, management, and others. In the last years, he has focused on NP-complete problems, especially route selection and optimization and the application of metaheuristics for approximate solution of such complex tasks. He has published over 775 articles, most of those being refereed papers, and several text books on the subject. His Hirsch-index is 40 by Google Scholar (based on 7300 citations there).

Micro Service based Sensor Integration Efficiency and Feasibility in the Semiconductor Industry

Germar Schneider[1], Paul Patolla[2], Matthias Fehr[1], Dirk Reichelt[2], Feryel Zoghlami[1], and Jerker Delsing[3]

***Abstract*—The semiconductor industry is strongly increasing the production capacities and the product portfolio for a wide range of applications that are needed in the worldwide supply chains e.g. the automotive, computer and security industry. The complex manufacturing processes require more automation, dig- italisation and IoT frameworks, especially for highly automated semiconductor manufacturing plants. Over the last years, this industry spent much effort to control highly sensitive materials in production by product monitoring using advanced process control by various sensors in production. Nevertheless, until today, sensor integration, especially for such sensors that are not supported by the equipment vendors, is time-consuming and complicated. This article aims to use a micro-service-based approach by Eclipse Arrowhead as an open-source microservice architecture and implementation platform [1]. This architecture is an easy and powerful framework that can be used for multiple sensor applications to control the manufacturing material flow in a modern semiconductor plant with a high product mix. The article describes how the engineering process was designed, the architecture of the use case and the main benefits in the operational business are shown.**

***Index Terms*—Eclipse Arrowhead, IoT Frameworks, Sensor-integration, Micro Services, Engineering Process, Digitalisation, Industry4.0**

I. MOTIVATION

Semiconductor manufacturing has become increasingly complex in recent years. A variety of new IC facilities manufacturing products according to More Moore or More than Moore have started production based on 200 and 300mm wafers. Wafer facilities that follow both the one and the other strategy have to adapt their production to the diverse requirements of the respective technologies and customer requirements. The Time2Market factor is enormously important to deliver the products in an optimal quality to the customers at the right time. Monitoring the supply chain is one of the most important points to achieve the mentioned goals. Sensors for monitoring the material flow and hundreds of different process steps play an increasingly important role in highly automated manufacturing plants. Over the last ten years, Infineon Technologies Dresden (IFD) has almost completely automated the 200 mm production line and established the world´s first fully automated 300 mm line for power semiconductors. IFD worked over the last 15 years already on the automation of its 200 mm line with hundreds of different products in the same line. The challenges in the field were described by Heinrich et al. [2]. The influence of automation on the production in the semiconductor industry was already researched [3]–[5] by different teams from IFD in the front end and the wafer test area [6]. The results of this work showed that the controlling and monitoring of a high mix product portfolio requires advanced automation and factory integration concepts compared to high volume production with a low product mix. In 2010, IFD started the first worldwide production line for power semiconductors based on 300 mm wafers and thin wafer technologies. This kind of manufacturing required special manufacturing concepts again. The use of more automation, especially digitalization, has been the main enabler to overcome the challenges by the new requirements of power semiconductor manufacturing described by G. Schneider et al. [7]. Both production lines at IFD follow the More than Moore strategy and produce hundreds of different products in different technology nodes with the highest quality requirements e.g. products in the medical sector or automotive applications. Besides modern production plants, which already have a large number of integrated sensors, additional sensors or IoT devices and external, chemical and physical sensors are increasingly needed for real-time monitoring and controlling hundreds of individual process steps. The number of sensor data per day has now increased to more than 1 billion per day, which stresses the manufacturing facility's server capacities and IT performance. Due to the introduction of automation by means of robots or fully automated, hybrid transportation systems, only a few humans remain in the production area who can sense deviations with their sensory organs. Therefore, it has become increasingly important to online monitor the different machines, processes and, above all, maintain the

[1] Infineon Technologies Dresden GmbH & Co. KG Dresden, Germany (e-mails: {germar.schneider, matthias.fehr, feriel.zoghlami}@infineon.com)
[2] University of Applied Sciences Dresden, Dresden, Germany (e-mails: {paul.patolla, dirk.reichelt}@htw-dresden.de)
[3] Lulea University of Technology, Lulea, Sweden (e-mail: jerker.delsing@LTU.se)

DOI: 10.36244/ICJ.2022.3.10

entire systems such as the different machines, processes and the whole cleanroom. In this context, sensor fusion [8] and IoT [9] play an important role in ensuring the link between the different participant members inside the production line. The IoT offers many new methods to easily integrate these kinds of IoT devices and process and visualize the resulting data while sensor fusion is more focusing on the perception of a dynamic environment including humans with the advantage of a spatial and temporal coverage extension and improvement of the global system resolution [10], [11]. Hence, IIoT can be understood as the connection of smart assets, which are part of a larger system of systems (SoS) in industrial environments to optimize the value of production [12]. A SoS can be described as a set of systems working together to achieve a more complex target or a higher purpose [13], whereas each system can act independently and have its own purpose. Furthermore, the individual systems of the set are organized independently to fulfill their purposes. The combination of systems provides results that cannot be achieved by individual systems [14]. Five characteristics of SoS can be used to define and differentiate it from other complex, but monolithic systems: operational independence of its systems, management independence of the systems, evolutionary development, emergent behaviour, and geographic distribution [15], [16]. In comparison to consumer applications, the specifications and requirements in industrial environments are more restrictive. Their focus is commonly on security requirements, device interoperability, quality of service, and communication technologies, and protocols [17]. The right IoT framework for the entire system is chosen based on the System of System approach and remains a big challenge There are already many different products offering so-called IoT frameworks on the market, but only a few of those frameworks offers real capabilities in terms of SoS and real-time capability. Panigua and Delsing [18] compared different IoT frameworks and emphasized the Arrowhead framework, an open source software that provides various important tools and applications. This article gives a nice overview about the features of the most famous IoT frameworks like AUTOSAR, FIWARE.. which show the advantages of the Arrowhead framework compared to the other frameworks. In general, besides being an open source framework and supporting security and interoperability features, Arrowhead Eclipse Framework has the following big advantages [18]:

- The different functionalities and core services are distributed into the different core systems instead of having a unique middleware that reduces the scalability.
- The orchestration system is capable of computing new orchestration patterns in runtime and providing dynamic orchestration and authorization.
- It supports and manage applications with real-time constraints.
- It supports different data transmission protocols (TCP/UDP, DTLS/TLS) as well as communication protocols (HTTP, CoAP, MQTT, OPC-UA).

This contribution shows a new approach using the framework of Eclipse Arrowhead, which can overcome the challenges for easy sensor integration and the interoperability of many different use cases in semiconductor wafer facilities. The framework is based on a service-oriented architecture and provides automation capabilities [19]. While enabling device interoperability and IIoT at a service level, it meets the demands in terms of real-time control, engineering simplicity, security and scalability. Furthermore, the framework enables simple sensor integration and consists of the following three core services: authorization system, a service registry unit, and the orchestration system (see Figure 1 below).

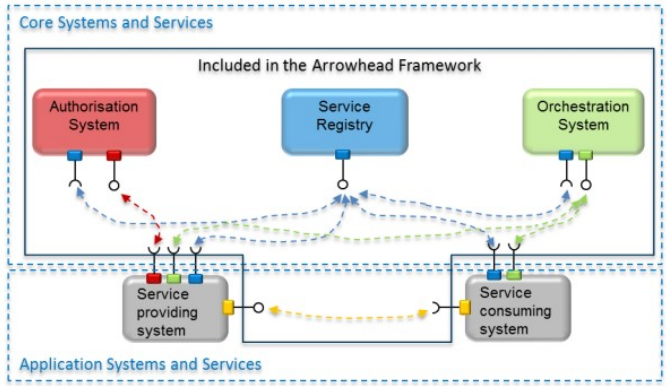


Fig. 1. Core System and Services

In the different applications, the provided information is sent to a service consuming system that can be used in many different use cases. This architecture offers a simple method of interoperability for hundreds of sensors that can be than easily connected to internal clouds in the IT environment of the wafer facilities. The Arrowhead framework has been used in various IoT automation scenarios [19], e.g. for programmable logic controller (PLC) device monitoring, energy optimization, replacement devices, maintenance, and as for the efficient deployment of a large number of of IoT sensors [18]. A first use case of the implementation of sensors in the cleanroom of IFD based on the architecture for automatic integration of sensor with the IO-link standard into a system of systems was already published by Patolla [20]

II. DESCRIPTION

Infineon Dresden has been working for over 20 years to improve process monitoring using statistical process control (SPC) and advanced process control (APC) to better monitor increasingly complex manufacturing processes. In addition to the classical fault detection and control system (FDC), internal and external sensor data has been used for many years at different points in the process, the system, or the cleanroom where the respective process data cannot be generated the wafer producing tool itself. The integration of suchlike sensor data, which are installed outside of the equipment and not supported by the vendors, is very time consuming and requires extremely resource-intensive engineering. The integration of

the sensors in the respective manufacturing execution system of the fab requires coordination of different departments in the fields of IT, individual process technology and production. This is associated with a high expenditure of time and money. Furthermore, the responsibility of such sensors is not always given especially if they are not directly assigned to a tool owner. Another problem is that the respective sensor data can be found and orchestrated in the system, and the certainty must be given that it is really sensor data and that it can be managed in a safe factory environment. Today, various software companies offer countless software solutions to manage sensors and visualize sensor data in so-called IoT frameworks. Therefore, it is not easy for the respective users to select the right application not to have to go from one software solution to the next. This article provides an overview of how IFD found a way to integrate external sensors using the Eclipse Arrowhead framework easily.

III. INNOVATION

The innovation of the project lies in the integration, management and visualization of external sensors in a semiconductor facility using a state-of-the-art IoT framework based on Eclipse Arrowhead: the Arrowhead-Framework shows the potential of how to easily implement any sensor into the manufacturing execution system (MES) of existing semiconductor production in the future and what advantages such integration and orchestration of the data means for the future semiconductor manufacturing. The respective architectures and interoperability are discussed in particular. The use of the Eclipse Arrowhead Framework offers a lot of opportunities in case of interoperability and security. If one use case is implemented, many other use cases can be easily implemented using similar processes. This open-source application offers the possibility to the industry to enhance the automation and digitalisation level with the lowest engineering efforts.

IV. RESULTS

For the implementation of sensors in the cleanroom, the sensor team of IFD worked to get out the main requirements for the future integration of sensors to enable them to be compatible with a variety of different applications in the factories. The main goal was to reduce the engineering efforts and the engineering costs. In a first step, a possible architecture was described based on the requirements, which should comply with the Eclipse Arrowhead IoT Framework. The first use cases are implemented after the functional design was finished and the applications tested and validated, which were important steps before starting the operation of the sensors in the manufacturing area. Figure 2 shows the engineering process with all interfaces needed to obtain compliance. This engineering process was defined by G. Urgese et al. [21]. IFD strongly followed this engineering process in implementing the different use cases in production. In a second step, the architecture with all interfaces of the most important stakeholders are established, see Figure 3 underneath Figure 2.

The architecture results show the most important interfaces of the individual areas and stakeholders, which are required for later implementation. The stakeholders are engineers from the field of the IT department, of the process and product engineering groups and by specialist working on big data applications. As shown in this 3, seven main stakeholder functions were needed in the specific use cases involved to establish and run the operational concepts. The end-users, typical process or manufacturing engineers, needed support from different IT experts with deep knowledge of the IT infrastructure and the IT network. In addition, other IT specialists are needed to maintain the sensor applications in the MES afterwards. For data controlling, stakeholders with knowledge on advanced process control and data scientists for visualization of complex data evaluation must also be involved in the process. An overview is provided in Figures 4 and 5 of how the IoT framework from Eclipse Arrowhead will enable significantly simplified sensor integration in the future while drastically reducing engineering efforts. As already mentioned, the Arrowhead Framework provides a variety of tools that can be important not only for the semiconductor industry but also for a variety of other industries such as the automotive, energy, building or building or a variety of other industries such as the automotive, energy, building or building metal industry to achieve overall compliance. Important attributes of this framework are security characteristics, a high level of interoperability and high saving potentials to reduce engineering costs. Furthermore, such IoT solutions can greatly improve current automation and digitalization levels in the respective fabs and, therefore, enhance the plant's KPIs, time to market, and the overall competitiveness of the company. We can show in this article, based on the example in the Infineon Dresden fab, how sensor integration can be realized with Eclipse Arrowhead and what benefits can be achieved now and in the future in the various applications.

Figure 4 shows how many stakeholders had been involved in a classical sensor integration in the fab without using an IoT framework. Six main stakeholders were needed in the past, which had to configure multiple applications. Three different specialists from the IT department in the fields of network applications, MES and factory integration were needed to implement hardware and software in the IT system of the fab. Furthermore, engineers in the production and unit product development (UPD), as well as data scientists, had to work on the implementation and visualization of the data. Figure 5 shows that the stakeholders could be reduced to only three main stakeholders using the Eclipse Arrowhead integration. The main work is saved within the IT department reducing the efforts for different alignments and waiting times from one expert to the other experts. Therefore, the work from three stakeholders could be reduced to only one main stakeholder for the MES integration. Figure 6 illustrates the reduction of time from 3 months down to around one day and the reduction from 6 main stakeholders to at least only two main stakeholders. The engineering efforts for only one sensor application have been estimated at three months. Security

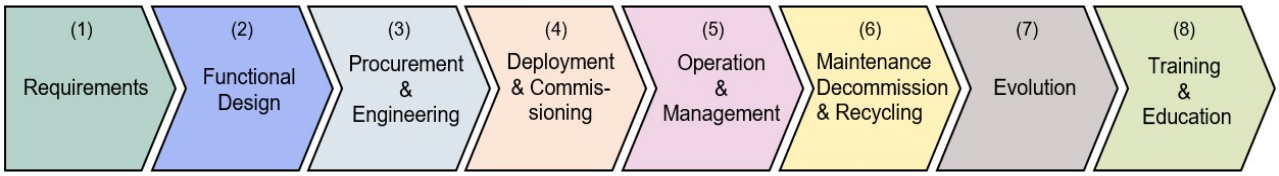


Fig. 2. Engineering Process

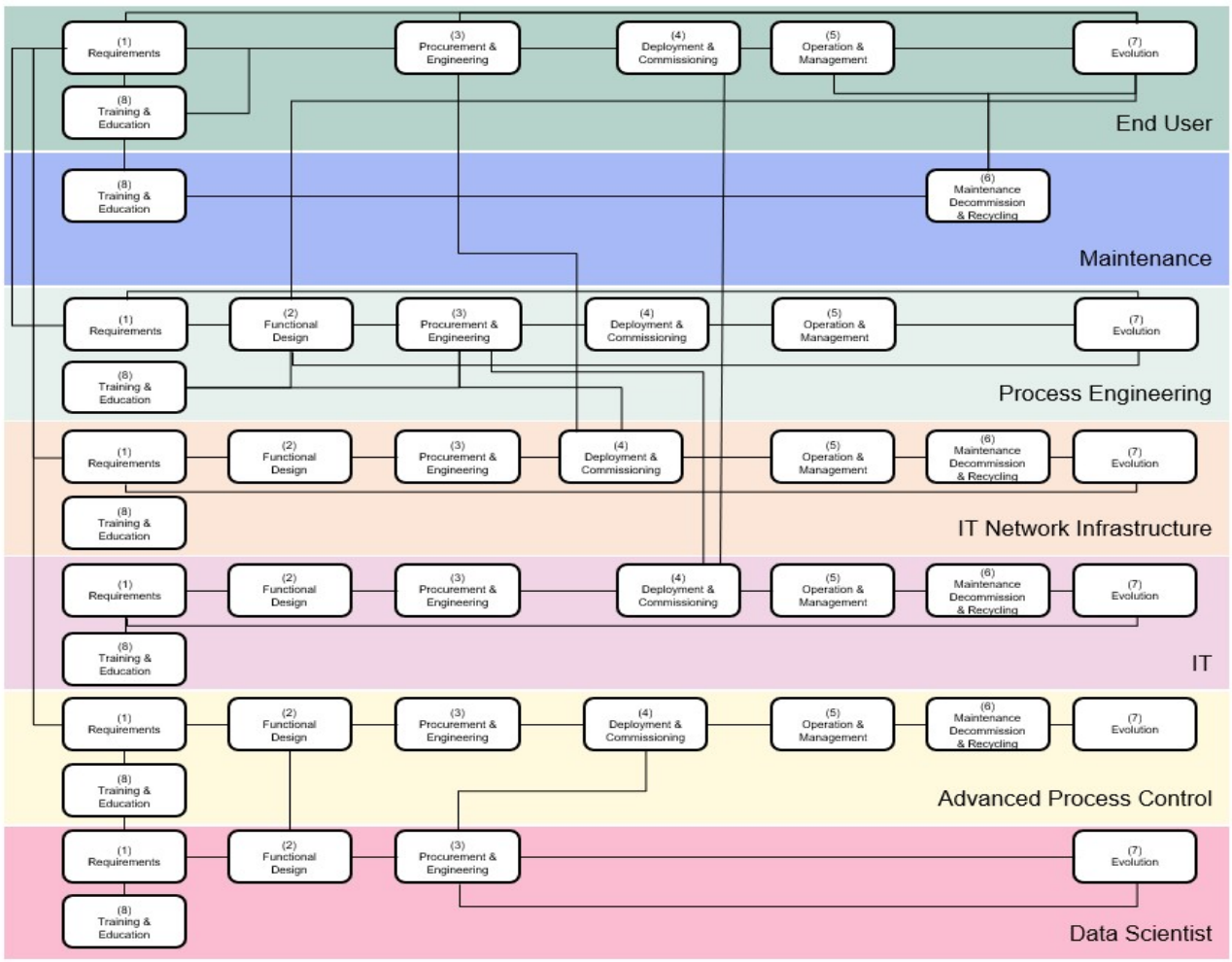


Fig. 3. Ownership of digitalization engineering process including stakeholders

is important in the semiconductor industry and open source applications are in general forbidden due those restrictions. Therefore, the applications were performed on only internal cloud solutions.

We showed that Eclipse Arrowhead Framework improved use cases for applications with one sensor to reduce the engineering time and efforts and sensor integration of external sensors for complex semiconductor equipment will be as easy as to plug and play an USB stick. Furthermore, for processes e.g. external sensor systems e.g. on on automated guided robots need more than one sensor system. For those use cases sensor fusion and the use of a multiple sensor integration is needed. We showed the architecture for such a use case with two different sensors delivering two output signals. Figure 7 below shows an overview of how the signals from two sensor systems, ToF and radar sensors, can be combined and create value out of the fused data for obstacles detection (more general details about the ToF/radar fusion system itself was already described in the papers [22] and [23]). An MQTT broker acts as the main information node.

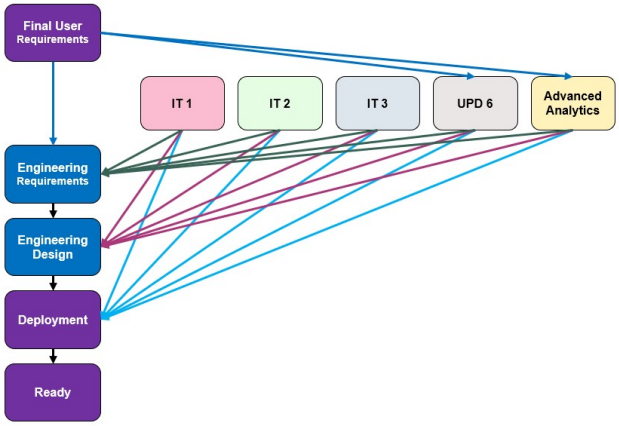


Fig. 4. Required engineering resources before Eclipse Arrowhead

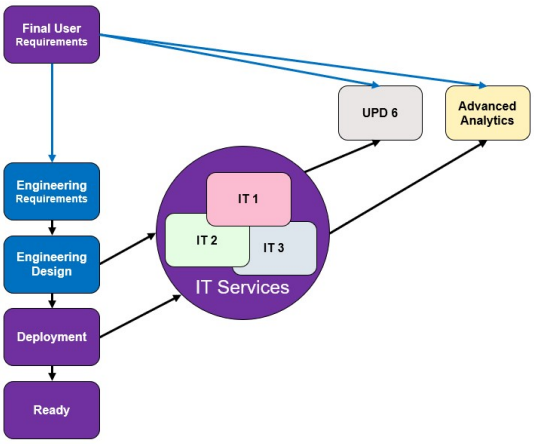


Fig. 5. Reduction of engineering resources after the implementation of Eclipse Arrowhead

The data from the test algorithm is stored in a database at the end to facilitate future data analysis and to enable live monitoring. Regarding how the Eclipse Arrowhead framework works, all required services in the use case must be registered in the service registry. This can be done at the beginning or during runtime and is done only once for one application. A provided metadata description supports the discovery of services, as it can be freely defined as a key-value pair. Subsequently, services that need to consume other services can query them via the Orchestrator service. The Authorization service supports the Orchestrator service in the area of security since only registered systems, which must be activated for other services, can query other services. The validation of the ToF/Radar sensor data fusion use case using the Arrowhead framework opens the door to other possibilities to apply the sensor fusion concept in other use cases by designing similar architectures. In our case, the architecture we choose gives us the possibility to easily

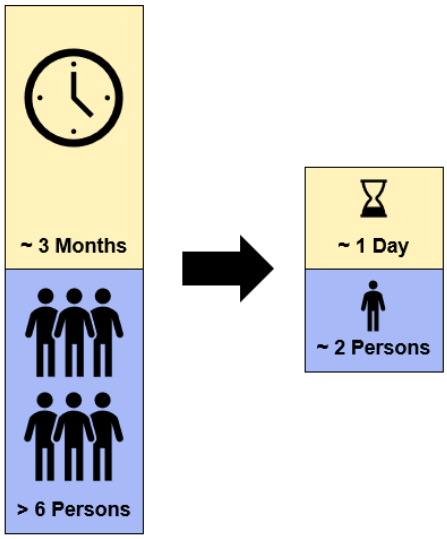


Fig. 6. Engineering resources and time savings before (blue color) and after use of Eclipse Arrowhead framework (yellow color)

integrate more sensors if needed for a better perception of the surrounding environment. Moreover, structuring our process flow into sub-units, each responsible of one part of the whole system gives the user the possibility to apply changes into the used algorithms or even replace some of the used methods (e.g update pre-processing algorithms for the radar raw data , change the fusion architecture, etc..). These changes are automatically considered by the Arrowhead modules, which means no extra effort is needed for the integration. The sending of data between system units is simultaneous, which allows us to meet real-time test requirements.

In the beginning, all participating services are registered in the Service Registry and enabled for communication with each other in the Authorization service. In the provided description is from each service the Topic, to which it sends its data. Then the sensor systems are started, which request the MQTT broker at the Orchestrator service. When an algorithm (”Preprocessing”, ”Fusion”, ”Background substraction”, ”Testing”) is started, it subscribes to its required topic at the MQTT broker and then sends its result back to the broker. The ”Testing” algorithm also stores its test results in a database. This use case can also be transferred to many other applications when different signals are used and combined in a modified or in a new sensor fusion system. Looking into the architecture we propose, which is required to fulfill the eclipse arrowhead requirements, we can integrate new sensors and modify or even replace one or more used algorithms without extra programming effort. Finally, all is designed to ensure real-time tests.

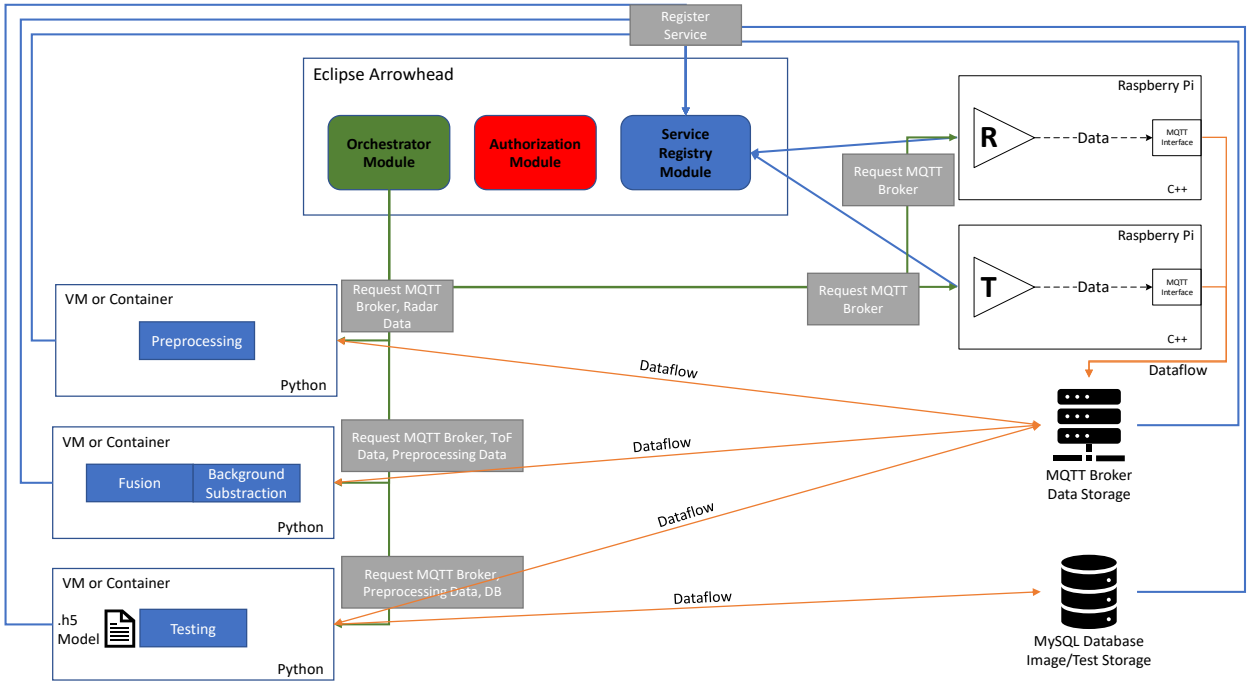


Fig. 7. Architecture using the Arrowhead Framework for more than one sensor which can be applied for use cases with multiple sensor data

V. CONCLUSION

The use cases showed that significantly fewer departments need to be contacted due to the automation of the integration process, which can be performed without needing a tool stop or the configuration of a multitude of different applications. In the future, the only need consists of picking the most suitable solution and connecting and configuring the sensor to start the analytics. Therefore, this application based on Eclipse Arrowhead will be a contribution of high importance to enhance the competitiveness of a semiconductor factory to install in a very easy and not time-consuming way sensors for advanced process monitoring and can be also used as an enabler for advanced data analytics, machine learning and artificial intelligence. The result of this work provides clear evidence that substantial engineering savings can be achieved using IoT frameworks like Eclipse Arrowhead. The next steps are to investigate the wider applicability of the approach to many other applications in the semiconductor industry and the overall supply chains of the entire component systems.

ACKNOWLEDGMENT

This concept was developed and published as part of the ”Arrowhead Tools” project. This project is co-financed with tax funds on the legal basis of the budget passed by the Saxon state parliament. This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826452. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and Sweden, Austria, Spain, Poland, Germany, Italy, Czech Republic, Netherlands, Belgium, Latvia, Romania, France, Hungary, Portugal, Finland, Turkey, Norway, Switzerland.

REFERENCES

[1] Eclipse Arrowhead. www.eclipse.org/arrowhead

[2] H. Heinrich, G. Schneider, F. Heinlein, S. Keil, A. Deutschlander, and R. Lasch, “Pursuing the Increase of Factory Automation in 200mm Fron-tend Manufacturing to Manage the Changes Imposed by the Transition from High-Volume Low-Mix to High-Mix Low-Volume Production,” in 2008 IEEE/SEMI Advanced Semiconductor Manufacturing Conference, May 2008, pp. 148–155. doi: 10.1109/ASMC.2008.4529020.

[3] S. Keil, A. Deutschla ¨nder, R. Lasch, H. Heinrich, G. Schneider: “Flow Production in Semiconductor Industry – a Paradigm Shift in IC-Manufacturing”, The 18th International Symposium Research-Education-Technology, June 2008, Danzig.

[4] S. Keil et al., “Establishing continuous flow manufacturing in a Wafertest-environment via value stream design,” in 2011 IEEE/SEMI Advanced Semiconductor Manufacturing Conference, May 2011, pp. 1–7. doi: 10.1109/ASMC.2011.5898196.

[5] S.Keiletal.,“Innovationandmanufacturingexcellenceinmaturemulti-product semiconductor fabrication facilities via design for flow by 3,” in 2011 Semiconductor Conference Dresden, Sep. 2011, pp. 1–5. doi: 10.1109/SCD.2011.6068762.

[6] D.Eberts,R.Rottnick,G.Schneider,S.Keil,R.Lasch,andO.Buhmann, “Managing variability within wafertest production by combining lean and six sigma,” in 2012 SEMI Advanced Semiconductor Manufacturing Conference, May 2012, pp. 33–38. doi: 10.1109/ASMC.2012.6212864.

[7] G. Schneider, S. Keil, and G. Luhn, “Opportunities, challenges and use cases of digitization within the semiconductor industry,” in 2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Apr. 2018, pp. 307–312. doi: 10.1109/ASMC.2018.8373173.

[8] J. Z. Sasiadek, "Sensor fusion," Annual Reviews in Control, vol. 26, no. 2, pp. 203–228, Jan. 2002, doi: 10.1016/S1367-5788(02)00045-7.

[9] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A Literature Review," Journal of Computer and Communications, vol. 3, no. 5, Art. no. 5, May 2015, doi: 10.4236/jcc.2015.35021.

[10] W. Elmenreich. "An introduction to sensor fusion. Vienna University of Technology", Austria, 502, 1-28. 2002

[11] M. L. Fung, M. Z. Q. Chen, and Y. H. Chen, "Sensorfusion: A review of methods and applications," in 2017 29th Chinese Control And Decision Conference (CCDC), Chongqing, China, May 2017, pp. 3853–3860. doi: 10.1109/CCDC.2017.7979175.

[12] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial internet of things (IIoT): An analysis framework," Comput. Ind., vol. 101, pp. 1–12, Oct. 2018, doi: 10.1016/j.compind.2018.04.015.

[13] P. Varga et al., "Making system of systems interoperable – The core components of the arrowhead framework," J. Netw. Comput. Appl., vol. 81, pp. 85–95, Mar. 2017, doi: 10.1016/j.jnca.2016.08.028.

[14] A. Albers, C. Mandel, S. Yan, and M. Behrendt, "SYSTEM OF SYS- TEMS APPROACH FOR THE DESCRIPTION AND CHARACTERI- ZATION OF VALIDATION ENVIRONMENTS," 2018, pp. 2799–2810. doi: 10.21278/idc.2018.0460.

[15] F. Blomstedt et al., "The Arrowhead Approach for SOA Application Development and Documentation," Oct. 2014. doi: 10.1109/IECON.2014.704887.

[16] M. W. Maier, "Architecting principles for systems-of-systems," Syst. Eng., vol. 1, no. 4, pp. 267–284, 1998, doi: 10.1002/(SICI)1520-6858(1998)1:4¡267::AID-SYS3¿3.0.CO;2-D.

[17] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, Opportunities, and Directions," IEEE Trans. Ind. Inform., vol. 14, no. 11, pp. 4724–4734, Nov. 2018, doi: 10.1109/TII.2018.2852491.

[18] C. Paniagua and J. Delsing, "Industrial Frameworks for Internet of Things: A Survey," IEEE Syst. J., vol. 15, no. 1, pp. 1149–1159, Mar. 2021, doi: 10.1109/JSYST.2020.2993323.

[19] J. Delsing, IoT Automation: Arrowhead Framework. Boca Raton: Taylor Francis Inc, 2017.

[20] P. Patolla, D. Reichelt, D. Mothes, and G. Schneider, "Anarchitecture for an automatic integration of IO-Link sensors into a system of systems," in IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada, Oct. 2021, pp. 1–6. doi: 10.1109/IECON48115.2021.9589686.

[21] G. Urgese, P. Azzoni, J. v. Deventer, J. Delsing and E. Macii, "An Engineering Process model for managing a digitalised life-cycle of products in the Industry 4.0," NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, 2020, pp. 1-6, doi: 10.1109/NOMS47738.2020.9110365.

[22] F. Zoghlami et al., "ToF/Radar early feature-based fusion system for human detection and tracking," in 2021 22nd IEEE International Con- ference on Industrial Technology (ICIT), Mar. 2021, vol. 1, pp. 942–949. doi: 10.1109/ICIT46573.2021.9453703.

[23] F. Zoghlami, M. Kaden, T. Villmann, G. Schneider, H. Heinrich. "AI- Based Multi Sensor Fusion for Smart Decision Making: A Bi-Functional System for Single Sensor Evaluation in a Classification Task". Sensors, 21(13), 4405, March 2021.

Germar Schneider holds a Diploma and a PhD in analytical chemistry from the University of Ulm, Germany in 1995. He joined the Siemens AG in Essonnes in France in 1995 as a process engineer in the wet department. In 1998 in Dresden, he became the section manager for the 200 mm wet department. From 2004 to 2008, he built up a team that was important for new factory automation and integration projects. Between 2008 and 2012, as a manager in the new wafer test department, he was responsible for production equipment engineering. With 27 years of experience combining the know-how of process engineering, production, maintenance, automation and the work in very large EU projects e.g. EPT300, EPPL, SemI40, Productive40, iDev40 and arrowhead tools, Germar Schneider is working on new factory integration concepts to improve the manufacturing of the semiconductor industry.

Paul Patolla holds a Diploma in business informatics from the University of Applied Science of Dresden (HTW) in 2019. Paul worked already as a student member at IFD from 2016 to 2019 in the material management department. He is a professional in database design and software development. Meanwhile, Paul is a member of smart production system with a focus on IIoT, automation and digitization and he has been working on the arrowhead tools project since 2019.

Matthias Fehr is a state-certified technician specialized in electrical engineering and data processing. Matthias has been working for more than 25 years at Infineon Dresden in chemical mechanical polishing and for more than ten years, he has been leading the sensor group of Infineon Dresden. Matthias strongly supports the arrowhead tools project for sensor integration in the semiconductor industry.

Dirk Reichelt holds a diploma and a PhD in business informatics from the Technical Univer- sity of Ilmenau. He worked for several years in the semiconductor industry. Since 2010 he has been a full professor for information management at the Dresden University of Applied Sciences. The focus of his team and him are on the research and development of cyber-physical production systems and the use of Industrial IoT solutions to create process innovations.

Feryel Zoghlami holds a Diploma in industrial informatics and automation engineering from the National Institute of Applied Sciences and Technology in Tunis in 2018. Feryel is currently a PhD candidate since 2019 at IFD and at the University of Bielefeld. Her focus is on developing solutions to improve human-robotic collaboration by reference to the sensor fusion concept and different machine learning tools
and artificial intelligence in general. Feryel is part of the Arrowhead tools project since 2019

Jerker Delsing received the M.Sc. degree in engineering physics from the Lund Institute of Technology, Lund, Sweden, in 1982, and the Ph.D. degree in electrical measurement from Lund University, Lund, in 1988. In 1994, he was promoted to Associate Professor in Heat and Power Engineering with Lund University. Early 1995, he was appointed Full Professor in Industrial Electronics with the Lulea University of Technology, where he is currently the Scientific Head of EISLAB. His present research profile can be entitled IoT and SoS automation, with applications to automation in large and complex industry and society systems. Prof. Delsing and the EISLAB group have been a partner and coordinators of several large to very large EU projects in the field, e.g., socrades, IMC-AESOP, arrowhead, FAR-EDGE, productive4.0, and arrowhead tools.

IEEE International Conference on Blockchain and Cryptocurrency (ICBC 2023)

1–5 May 2023 | Dubai, UAE – https://icbc2023.ieee-icbc.org/

Call for Papers

NOTE: *ICBC'23 will be operated with double-blind review process! The conference will also adopt a rebuttal procedure*.

ICBC 2023 will be the 5th edition of the IEEE International Conference on Blockchain and Cryptocurrency, sponsored by the IEEE Communications Society. ICBC 2023 is seeking submissions of technical papers (both full and short), posters, and tutorial proposals in the following areas related to Blockchains and Cryptocurrencies as well as emerging areas of Blockchains.

- Distributed Consensus & Fault Tolerance Algorithms
- Performance, Scalability Issues
- Distributed Database Technologies for Blockchain
- Blockchain Interoperability and Cross-chain mechanisms
- Blockchain Platforms
- Decentralized App Development, DAPPs, and Services
- Smart Contracts and Verification
- Security, Privacy, Attacks, Forensics
- Transaction Monitoring and Analysis
- Regulations & Policies in Cryptocurrency
- Novel Mechanisms for the Creation, Custody, and Exchange of Cryptoassets
- Anonymity and Criminal Activities of the Cryptocurrency
- Managing the Risks of Cryptocurrency
- Distributed Trust
- Decentralized Internet Infrastructure
- Decentralized Financial Services
- Blockchain for Internet of Things/Cyber Physical Services
- Blockchain social media platforms/tokens
- Blockchain and Machine Learning/Artificial Intelligence

ICBC 2023 authors are invited to submit papers describing original, previously unpublished work, not currently under review by another conference, workshop, or journal. Only PDF files will be accepted for the review process, and all submissions must be made electronically through EDAS via the URL: https://edas.info/N29908.
Note that the maximum paper length is 8 pages for full papers, 4 pages for short papers, 2 pages for poster submissions, and 16 pages for Systemization of Knowledge (SoK) papers, all excluding references.

Important Dates

Paper Submission Deadline:	December 4, 2022
Tutorial Proposal and Poster Submission Deadline:	January 15, 2023
Demo Submission Deadline:	January 25, 2023
Acceptance Notifications:	February 12, 2023
Tutorial and Demo Acceptance Notifications:	February 28, 2023
Camera-Ready Deadline:	March 19, 2023

General Co-Chairs

Adel Ben Mnaouer	Canadian University Dubai, UAE
Burkhard Stiller	University of Zürich, Switzerland
Fakhreddine Karray	MBZUAI, UAE

Technical Program Co-Chairs

Moayad Aloqaily	MBZUAI, UAE
Vinayaka Pandit	IBM India Research Lab, India
Laura Ricci	University of Pisa, Italy

WORKSHOP ON SCALABLE AND TRUSTWORTHY AI FOR 6G WIRELESS NETWORKS (6GSTRAIN)

SCOPE

To design and deploy artificial intelligence (AI)-enabled sixth-generation (6G) zero-touch automated wireless networks, scalability, trust and explainability of AI algorithms are required. Recent advances in standardization point out fully decentralized AI as the way to deploy large-scale network automation. In ETSI's zero-touch architecture – for instance – each network domain is endowed with a data collection element that feeds a local AI analytics and decision entity. The central entity plays only the role of a coordinator/model aggregator without necessarily having access to the distributed raw data. Nonetheless automatically managing a massive number of network elements by only increasing the processing resources is not a guarantee for scalability since it also increases the complexity of their management, the degree of contention in the system, as well as suffers from the lack of collaboration between the distributed decision entities. A scalable architecture would therefore achieve a trade-off in i) utilization of shared resources to minimize contention but also to avoid complex management of unnecessary resources; ii) information flow by sharing only compressed parameters instead of raw data and iii) degree of collaboration by enabling the exchange of inferences between decentralized analytics/decision engines while avoiding that they fall in competitive or too cooperative situations.

On the other hand, the practical deployment of AI automation in 6G requires establishing a high level of trust and transparency in the AI black boxes. In this regard, explainable AI (XAI) tools and metrics will play a pivotal role in unveiling the rationale behind AI predictions and decisions. From the state-of-the-art gradient-based attribution methods – such as Integrated Gradients, Saliency Maps and ε-LRP – to perturbation-based attribution methods such as Shapley Value sampling, the XAI framework enables a better understanding of the causality in AI models.

TOPICS OF INTEREST

We seek original completed and unpublished work not currently under review by any other journal/ magazine/conference. Topics of interest include, but are not limited to:

- Federated learning for scalable 6G networks.
- Decentralized reinforcement learning for scalable 6G networks.
- XAI for trustworthy AI in 6G networks.
- AI/XAI for Open RAN in 6G Networks.
- Zero-touch network architectures and protocol design for decentralized AI.
- Decentralized AI schemes with low energy consumption.
- Semantic communications for 6G scalability.
- Distributed Data and knowledge distillation for 6G.
- Wireless communications for AI operation.
- Decentralized resource management and network slicing.
- Decentralized AI for low latency applications.
- Decentralized AI for PHY/MAC operation.
- Decentralized AI and Blockchain for 6G.
- Decentralized AI integration in 6G PoCs.
- Production platforms for decentralized and federated learning.
- New business models for XAI.

IMPORTANT DATES

Paper Submission Deadline: 20 January 2023

Paper Acceptance Notification: 6 March 2023

Camera Ready and Registration for accepted papers: 15 March 2023

KEYNOTES (Confirmed)

Prof. Merouane Debbah, Centrale-Supélec, France/TII, UAE
Prof. Melike Erol-Kantarci, University of Ottawa/Ericsson R&D, Canada

CHAIRS

Dr. Hatim Chergui, CTTC, Spain (Contact: chergui[at]ieee[dot]org)
Dr. Kamel Tourki, Ericsson R&D France
Prof. Mustapha Benjillali, INPT, Morocco
Dr. Bouziane Brik, Bourgogne University, France
Prof. Adlen Ksentini, Eurecom, France

icc2023.ieee-icc.org

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

IFIP Networking 2023

June 12-15, 2023 / Barcelona, Spain

Call for Papers

The International Federation for Information Processing (IFIP) Networking 2023 Conference will be held at the Universitat Politècnica de Catalunya, in the beautiful city of Barcelona, Spain. This is the 22nd event of the series, sponsored by the IFIP Technical Committee on Communication Systems (TC6).

The main objective of Networking 2023 is to bring together members of the networking community, from both academia and industry, to discuss recent advances in the broad and quickly-evolving fields of computer and communication networks, to highlight key issues, identify trends, and develop a vision for future Internet technology, operation, and use. Networking 2023 technical sessions will be structured around the following yet non limitative areas:

- Network Architectures, Applications and Services
- Network Modeling, Analysis and Operation
- Network Security and Privacy
- Wireless Networking

Important dates

Full paper submission: January 30, 2023
Notification of acceptance: March 31, 2023
Camera-ready version: May 2, 2023

All papers should be submitted via EDAS. Full instructions on how to submit papers are provided on the IFIP Networking website:

networking.ifip.org/2023/

ORGANIZING COMMITTEE

General Chairs

Davide Careglio, Universitat Politècnica de Catalunya, Spain
Jordi Domingo-Pascual, Universitat Politècnica de Catalunya, Spain

Program Chairs

Mun Choon Chan, National University of Singapore
Violet R. Syrotiuk, Arizona State University, USA
Xavier Gelabert, Huawei Stockholm Research Centre, Sweden

Steering Committee

Robert Bestak, Czech Technical University in Prague, Czech Republic
Silvia Giordano (Chair), University of Applied Science and Arts - SUPSI in Ticino, Switzerland
Henning Schulzrinne, Columbia University, USA
Burkhard Stiller, University of Zurich, Switzerland
Joerg Widmer, IMDEA Networks, Spain

Local Chair

Josep Solé Pareta, Universitat Politècnica de Catalunya, Spain

Publicity Chairs

Mirosław Klinkowski, National Institute of Telecommunications, Poland
Salvatore Spadaro, Universitat Politècnica de Catalunya, Spain

Web Chair

Albert López Brescó, Universitat Politècnica de Catalunya, Spain

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*