Scientific and Practical Cyber Security Journal (SPCSJ) 8(4): 54 – 65 ISSN 2587-4667 Scientific Cyber Security Association (SCSA)

OVERVIEW OF RESEARCH TRENDS AND CHALLENGES IN 6G MOBILE NETWORKS AND THE COMPUTING CONTINUUM

Stojan Kitanov[1], Dragi Kimovski[2], Fisnik Doko[1], Kurt Horvath[2], Shpresa Tuda[1], Blerta Idrizi[1]
[1] Mother Teresa University, Faculty of Information Sciences, Skopje, Republic of North Macedonia
[2] Alpen Audria University of Klagenfurt, Department of Information Technology, Klagenfurt, Republic of Austria

ABSTRACT: The rapid proliferation of IoT devices, coupled with the generated exponential growth of data, has necessitated the development of advanced network architectures. As a result, 5G mobile networks have already begun to face challenges such as network congestion, latency, and scalability limitations. Therefore, the need for a robust and future-proof solution becomes increasingly evident. In this direction, many research initiatives and industrial communities started to work on the development of 6G mobile networks. On the other hand, the emerging concept of Computing Continuum encompasses the seamless integration of edge, fog, and cloud computing resources to provide a unified and distributed computing environment, and it aims to enable real-time data processing, low-latency response, and intelligent decision-making at the network edge. The primary objective of this research paper is to address the shortcomings of existing network infrastructures and to overcome these shortcomings by integrating advanced AI capabilities in 6G mobile networks with the Computing Continuum. Moreover, it would be proposed a Computing Continuum Middleware for Artificial Intelligence over 6G networks, which would offer high-level and well-defined ("standardized") interfaces which would create an automated, sustainable loop for managing IoT applications utilizing AI approaches over 6G networks.

KEYWORDS: *AI (Artificial Intelligence), Cloud Computing, Computing Continuum, Distributed AI, Fog Computing, Mobile Edge Computing, 5G, B5G, 6G.*

1. INTRODUCTION

Recently we have witnessed a significant increase in the use of distributed network-sensitive applications following the industrial revolution of Artificial Intelligence (AI) and the Internet of Things (IoT) (Rashid, Adib Bin, and Md Ashfakul Karim Kausik, 2024 & Seng, Kah Phooi, Li Minn Ang, and Ericmoore Ngharamike, 2022). An important pillar in this revolution represents the evolution in mobile network technologies across five generations, roughly ten years long each: 1G in the 1980s (analog voice), 2G in the 1990s (digital voice), 3G in the 2000s (mobile data), 4G LTE-A, and WiMAX 802.16 in the 2010s (mobile broadband internet), and finally reaching 5G in the 2020s (ultra-low latency, mm-waves, gigabit throughput), to be extended to 6G (blockchain and terahertz bandwidth) by the end of this decade (Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al, 2023).

Currently, multiple innovative applications that rely on artificial intelligence approaches (e.g., smart city, virtual reality, robotics, swarms) exploit the radio communication improvements and architectural changes in the edge and core networks. Such distributed applications attract mobile participants who connect through smartphones, virtual reality glasses, drones, robots, or any special IoT end device. Processing these interactions in real-time and meeting the requirements of the utilized AI algorithms is essential for their wider integration into our daily lives and for ensuring a responsive and fluent user experience.

The traditional solution to this problem is to offload the workload (i.e., AI model training) to a Cloud data center offering high-performance computing and storage capabilities (Choudhury, Alok, Manojit Ghose, Akhirul Islam, and None Yogita, 2024 & Hao, Tianshu, Jianfeng Zhan, Kai Hwang, Wanling Gao, and Xu Wen, 2021). Unfortunately, the latencies required to reach remote Cloud data centers are often unacceptably high, although the 5G technology promises to keep the user latency between 1 and 4 ms for ultra-reliable low latency communications (URLLC) and enhanced mobile broadband (eMBB) network slices.

Therefore, many industrial and research initiatives started to focus on next 6th generation of mobile networks, or simply 6G (Letaief, Khaled B., Wei Chen, Yuanming Shi, Jun Zhang, and Ying-Jun Angela Zhang, 2019 & Saad, Walid, Mehdi Bennis, and Mingzhe Chen, 2019 & Tariq, Faisal, Muhammad R. A. Khandaker, Kai-Kit Wong, Muhammad A. Imran, Mehdi Bennis, and Merouane Debbah, 2020). The main driving force in designing and optimizing 6G architectures, protocols, and operations is Distributed Artificial Intelligence in the core, radio access, and the network edge, which would provide support of ubiquitous and mobiquitous smart services from the core to the end devices of the network, which would exceed the mobile data traffic used today. It is expected that AI would transform the wireless evolution of future Internet of Everything (IoE) from “connected things or objects” to “connected intelligence.”

Moreover, despite the scarcity of networking technologies that support modern IoT applications, the provisioning of low-end wireless Edge and high-end Cloud resources to the use case applications is still a tedious and time-consuming manual process.

Therefore, recently, efforts have been made to consolidate these resources across a unified cloud and edge ecosystem, named Computing Continuum, that brings the Cloud services closer to the end users. The Computing Continuum encompasses the seamless integration of edge, fog, and cloud computing resources to provide a unified and distributed computing environment, and it aims to enable real-time data processing, low-latency response, and intelligent decision-making at the network edge (Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al, 2023 & Tarneberg, William, Emma Fitzgerald, Monowar Bhuyan, Paul Townend, Karl-Erik Arzen, Per-Olov Ostberg, Erik Elmroth, Johan Eker, Fredrik Tufvesson, and Maria Kihl, 2022). For example, smart city IoT applications require clear differentiation between latency-critical tasks, such as AI-based routing of cars or pedestrians, and latency-permissive tasks, like localization of free, or occupied parking spaces. Furthermore, moving critical workloads onto edge devices (e.g., single-board computers, smartphones, routers) enables important data preprocessing, which reduces the traffic on the radio interface.

Therefore, without advanced middleware services offering high-level and well-defined (“standardized”) interfaces, every IoT application provider must design and deploy its own ad-hoc solutions to such problems. This makes the entire research and development effort in B5G and 6G technologies fragmented, often redundant, inefficient, and costly. Obviously, such manual and isolated approaches without integrated IoT application modelling and automated service orchestration do not scale with the long-term vision of millions of interconnected IoT devices relying on distributed AI approaches.

The primary objective of this research paper is to address the shortcomings of existing network infrastructures and to overcome these shortcomings by integrating advanced AI capabilities in 6G mobile networks with the Computing Continuum. It is organized in the following. Section 2 provides an overview of 6G Mobile Networks and its features. Section 3 explains the meaning and the importance of the Computing Continuum. Section 4 provides an explanation about our proposed Computing Continuum Middleware for Artificial Intelligence over 6G networks, which would offer high-level and well-defined (“standardized”) interfaces which would create an automated, sustainable loop for managing IoT applications utilizing AI approaches over 6G networks. Finally, Section 5 concludes the paper and provides directions for future research.

2. OVERVIEW OF 6G MOBILE NETWORKS

The 2030 UN Agenda for sustainable development goals, adopted by all UN members in 2015, contains a set of 17 objectives for “shared blueprint for peace and prosperity for the people and the planet,” which should be accomplished by 2030 (“Transforming Our World: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs.”). The rapid advancements of mobile networks may play a major role in fulfilling these sustainable development goals. The deployment of 5G mobile networks have marked the beginning of a true digital society, and have achieved significant improvements in terms of latency, data rates, spectral efficiency, mobility and number of connected smart mobile devices (Jiang, Dajie, and Guangyi Liu, 2016 & Kitanov, Stojan, Edmundo Monteiro, and Toni Janevski, 2016).

However, the rapid development of artificial intelligence (AI), virtual reality (VR), three-dimensional (3D) media, and the internet of everything (IoE), has led to an exponential traffic growth and a massive volume of traffic. The growth of the global mobile traffic according to ITU predictions in 2030 is predicted to be around 5000 EB/month ("Report ITU-R M.2370-0." IMT Traffic Estimates for the Years 2020 to 2030, 2015).

As a result, 5G network would not be able to cope with these rapid increased demands of data traffic (Nawaz, Syed Junaid, Shree Krishna Sharma, Shurjeel Wyne, Mohammad N. Patwary, and Md. Asaduzzaman, 2019). For example, the holographic communication would require a data rate up to terabits per second (Tb/s), which is three times higher than the 5G's data rate and massive low latency (hundreds of microseconds), which is three time less than 5G's latency (Clemm, Alexander, Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck, 2020 & Du, Jun, Chunxiao Jiang, Jian Wang, Yong Ren, and Merouane Debbah, 2020 & Strinati, Emilio Calvanese, Sergio Barbarossa, Jose Luis Gonzalez-Jimenez, Dimitri Ktenas, Nicolas Cassiau, Luc Maret, and Cedric Dehos, 2019). In addition, due to the rapid deployment of Internet of Things (IoT) and future Internet of Everything (IoE) devices, it would be necessary to enhance the connection density and coverage of 5G enabled IoT networks (Tang, Fengxiao, Yuichi Kawamoto, Nei Kato, and Jiajia Liu, 2019 & Zhang, Shangwei, Jiajia Liu, Hongzhi Guo, Mingping Qi, and Nei Kato, 2020). Moreover, the new emerging services of Internet of Everything (IoE) such as extended reality (XR), telemedicine systems, mind-machine interface (MMI), and flying cars would demand very high transmission rates, high reliability, and low latency, which significantly surpass the original goals of 5G networks (Huang, Yakun, Boyuan Bai, Yuanwei Zhu, Xiuquan Qiao, Xiang Su, Lei Yang, and Ping Zhang, 2023 & Khan, Latif U., Ibrar Yaqoob, Muhammad Imran, Zhu Han, and Choong Seon Hong, 2020). In addition, the future mobile networks are expected to be ultra-large-scale, highly dynamic, and incredibly complex system, where the manual optimization and configuration tasks used in the existing mobile networks would be no longer suitable for the next generation mobile networks (Zhang, Shangwei, Jiajia Liu, Hongzhi Guo, Mingping Qi, and Nei Kato, 2020). As a result, 5G networks have already begun to face challenges such as network congestion, latency, and scalability limitations.

Table.1. *Comparison between 5G and 6G Network*

KPI	**5G**	**6G**
Traffic Capacity	10 Mbps/m^2	~ 1-10 Gbps/m^3
Peak Data rate of the device	10 to 20 Gbps	1 Tbps
End-to-End Latency	5 - 10 ms	10 - 100 µs
Uniform user experience	50 Mbps 2D everywhere	10 Gbps 3D everywhere
Localization precision	10 cm on 2D	1 cm on 3D
Mobility	500 km/h	Up to 1000 km/h
Application Types	eMBB, URLLC, mMTCP	mBRLLC, mURLLC, meMBB
Spectral Efficiency	30 bps/Hz	100 bps/Hz
Energy Efficiency	1x	10 - 100x of 5G
Processing Delay	100 up to 50ns	10ns
Frequency Bands	Sub-6 GHz; MmWave for fixed access at 26 GHz and 28 GHz.	Sub-6 GHz; MmWave for mobile access; THz band above 300 GHz; Non RF

Therefore, the need for a robust and future-proof solution becomes increasingly evident. At the moment the main network infrastructure problem is that currently exists no cloud data center, with a dedicated connection to a telecommunication network offering sufficiently low latency and high throughput to support the execution of novel low-latency AI-based IoT applications and bring us into the 6G era by the beginning of the next decade.

As a result, sixth generation (6G) networks have been proposed as a way to enhance the 5G solutions (Kitanov, Stojan, Petrov, Ivan and Toni Janevski, 2021). Moreover, their potential is claimed to facilitate further development of smart IoT solutions. In this direction machine learning (ML) and generally Artificial Intelligence (AI) are becoming necessity for further expansion of the beyond 5G mobile world.

A comparison between 5G and 6G network is given in Table 1. It can be noticed that all parameters such as traffic capacity, data rate, end-to-end delay, processing delay, spectral and energy efficiency, etc. are improved several times over the value provided by 5G.

6G network should provide support for greater number of massive machine type connected devices, than 5G network. In that direction 6G should provide scalability, efficient connectivity with connectivity and reliability, coverage improvement, as well as, QoS and QoE provisioning. In order to manage such complex network, artificial intelligence (AI) with machine learning (ML) techniques are used to support the network autonomy, as well as, to get a knowledge of the surrounding environment in with 6G network would operate. 6G network would possess self-healing, self-organization, self-reconfiguration, self-optimization, self-aggregation, and self-protection capabilities.

6G network and 6G devices would require much higher energy consumption than 5G, increased by a factor of 1000 or even 10000, since they would operate on terahertz frequency bands (from 0.1 THz to 10 THz), and they would should provide support of a very high traffic capacity (10 Gbps/m^3), and ultra large number of ubiquitous connected wireless nodes. Therefore, 6G network should have improved energy efficiency than 5G network.

With the technological innovations such as artificial intelligence, augmented reality, virtual reality, and holographic telepresence, 6G would offer many new possibilities in the e-health, since it would provide very high reliability (99,99999 %), very high precision, and very ultra-low latency, less than 1 ms (Shahraki, Amin, Mahmoud Abbasi, Md. Jalil Piran, and Amir Taherkordi, 2021).

New services such as would be introduced in 6G. In addition, 6G network should require higher frequency bands in the terahertz spectrum, i.e., millimeter waves. In addition, a very high and opportunistic data rate is required to support new emerging applications, such as immersive multimedia. Also, 6G network would require much end-to-end delay of less than 1 millisecond, in order to support some 6G services such as augmented reality, and telepresence. Furthermore, 6G network must require too high reliability, in order to enable mission and safety-critical applications.

As it is already known, there are 3 main key services introduced by 5G: Enhanced Mobile Broadband (eMBB), massive Machine Type Communication (mMTC) and Ultra-Reliable Low-Latency Communication (URLLC) (Shahraki, Amin, Mahmoud Abbasi, Md. Jalil Piran, and Amir Taherkordi, 2021). In 6G network, all services would require low latency, high reliability, high data rate, massive connectivity, and full mobility. Therefore, the following potential 6G services could be: massive URLLC (mURLLC), enhanced mobile broadband URLLC (eURLLC), and massive eMBB (meMBB). The mURLLC is the 5G URLLC service type increased to a massive scale, and combines the URLLC with mMTC (mMTC + URLLC). One potential application in this class could be autonomous intelligent driving. The eURLLC service combines both eMBB and URLLC classes (eMBB + URLLC). AR, VR and holographic meetings are some of the applications that fit into this service type. The meMBB service type combines the mMTC and eMBB types (mMTC + eMBB). The tactile Internet fits into this class which would be used to improve the operations and functions in industrial IoT devices (IIoT) in Industry 4.0.

6G network would provide many new use cases, which cannot be completely supported by 5G (Clemm, Alexander, Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck, 2020). Some of them are holographic telepresence, industrial automation (industry 4.0 transform), e-health, tactile internet, augmented, and virtual reality.

An overview of the 6G network architecture together with the artificial intelligence is given on Fig. 1. Artificial intelligence together with Machine Learning (ML) is introduced at any horizontal or vertical level, at all TCP/IP layers, at any slice configuration and cloud-based network resource (edge computing) (Kitanov, Stojan, and Vladimir Nikolikj, 2022).

The 6G network design would introduce descriptive, diagnostic, predictive and prescriptive AI data analytics (Balali, Farhad, Jessie Nouri, Adel Nasiri, and Tian Zhao, 2020 & Kibria, Mirza Golam, Kien Nguyen, Gabriel Porto Villardi, Ou Zhao, Kentaro Ishizu, and Fumihide Kojima, 2018). The network analytics would analyze the collected historical data in order to get insights of the network status especially of the physical, data link, network and transport layer. Artificial intelligence (AI) together machine learning (ML) techniques are used for network autonomy, as well as, to get a knowledge of the surrounding environment in which 6G network would operate. AI would provide network status and utilization opportunities. Work data which is obtained as an output of the network analytics processes would be used by Core data analytics for detecting and predicting the network anomalies in order to improve reliability and security of the network. The obtained data would be used to detect future faults based at historical and current information and network behavior. Predictive analytics would use data to forecast future resource availability based at user mobility prediction, traffic patterns and overload. Finally, 6G would possess self-capabilities in healing, organization, reconfiguration, optimization, aggregation, and protection.

The conventional centralized ML algorithms need the availability of a large amount of centralized data and training on a central server (e.g., cloud server or centralized machine). However, this would result with a bottleneck in the future ultra-large scale mobile networks. Therefore, 6G networks would adopt ubiquitous AI solutions from the network core to the edge devices, i.e., Edge Intelligence (EI), or edge AI located at the edge of the network would be introduced in 6G. As a result, there would be a significant of big data sources from the cloud data centers to the increased number of smart edge devices (smartphones and IoT devices). Because of this data shift, these edge devices would move the AI solutions to the edge of the network in order to exploit completely the available potential of the edge big data sources. One possible emerging distributed ML technique is federated learning (FL) which would realize ubiquitous AI in the 6G networks. FL does not rely on storing all data to a central server where model training can occur. Instead, the innovative idea of FL is to train an ML model at each device (participant or data owner) where data is generated, or a data source has resided, and then let the participants send their individual models to a server (or aggregation server) and like that to achieve an agreement for a global model. However, despite the considerable potential advantages of FL for the 6G networks, FL is still in its infancy and encounter various challenges for fully operationalize in the 6G networks (Kitanov, Stojan, and Vladimir Nikolikj, 2022).

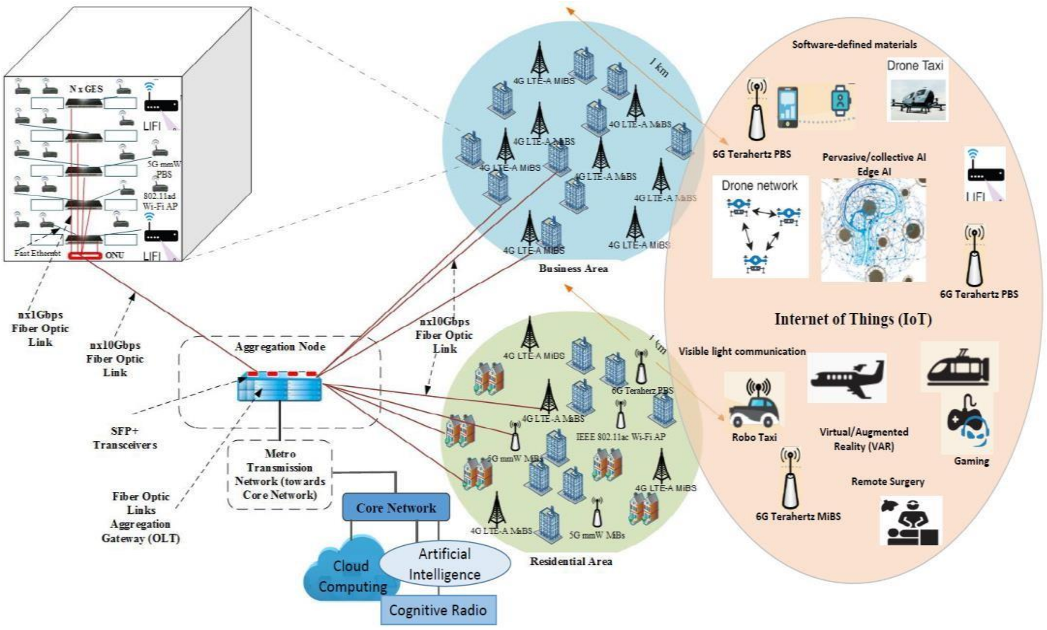

Fig.1. *6G Network Architecture*

Regarding the security aspects, the security vision of 6G would integrate AI to produce security automation which would ensure user/data privacy, would ensure trust, would provide prediction, detection and prevention of attacks, and would limit vulnerability propagation (Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al, 2023). Due to the new technologies in 6G network architecture, 6G new applications, and 6G new services, and the new policy regulations, the "Adversaries" also work non-stop to produce new kinds of security risks. Therefore, the identification and detection of zero-day attacks wouldn't be easy. As a result, the best practical defense would be to stop and prevent the zero-day attacks. In order to assess the value of network security in 6G, new set of Key Performance Indicators (KPIs) and Key Value Indicators (KVIs) such as level of protection, response time, network coverage, autonomicity level, AI robustness, Secure AI model convergence time, security function chain round-trip-time, and deployment cost for security functions, should be defined. In addition, several factors, including network information security, and security-related to AI/ML, should be taken into account when characterizing security (Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al, 2023).

3. OVERVIEW OF COMPUTING CONTINUUM

The idea of cloud computing is based on a very fundamental principal of reusability of IT capabilities. Cloud computing is a computing paradigm, where a large pool of systems is connected in private or public networks, in order to provide dynamically scalable infrastructure for application, data and file storage (Khan, Atta Ur Rehman, Mazliza Othman, Sajjad Ahmad Madani, and Samee Ullah Khan, 2014). The beginning of 21st century has been characterized by a general trend in which "business solutions migrate to the cloud", due to the following benefits:

- At the same time, the shared cloud resources (networks, servers, data warehouses, applications and services) can be rapidly provisioned and managed with minimal interaction by service providers;
- The cloud computing users may use these resources for development, hosting and running of services and applications on demand in a flexible way at any device, at any time and at any place in the cloud; and
- With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly.

However, cloud-centric solutions cannot support fast growing sizes of IoT deployment due to the following (Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al, 2023):

- the amount of data generated by the sensors is so large, that networking infrastructure between the sensors and the Cloud may not be able to efficiently transfer them;
- between the sensors and the cloud, within the ecosystem, multiple computing nodes with reasonable capabilities exist and the cloud-centric model does not support their utilization;
- large IoT ecosystems may require extremely time-constraint decision loops, which cannot be realized in cloud-centric deployments; and
- privacy and security of travelling data is put at stake.

The progressive convergence between Cloud Computing and the Internet of Things resulted in the appearance of the Computing Continuum (Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al, 2023 & Tarneberg, William, Emma Fitzgerald, Monowar Bhuyan, Paul Townend, Karl-Erik Arzen, Per-Olov Ostberg, Erik Elmroth, Johan Eker, Fredrik Tufvesson, and Maria Kihl, 2022). Initially, the concept of Edge Computing represented a middle ground between data centers and IoT hyper-local networks of sensors and actuators. Then a much more nuanced paradigm emerged and placed computing infrastructure on a spectrum covering from the Cloud Data Centers to Edge Nodes with many intermediate levels. High-Performance Computing (HPC), Artificial Intelligence, 5G/6G networks are also part of this Continuum for which hardware and software need to be jointly considered.

Moreover, the role that 6G networks can play in the transition from the IoT, to the Edge-Cloud 6G Computing Continuum.

The computing continuum seamlessly combines resources and services at the center (e.g., in Cloud datacenters), at the Edge, and in-transit, along the data path. Typically, data is first generated and preprocessed (e.g., filtering, basic inference) on Edge devices, while Fog nodes further process partially aggregated data. Then, if required, data is transferred to HPC enabled Clouds for Big Data analytics, Artificial Intelligence model training, and global simulations. Fig. 2. illustrates the implementation of a highly distributed infrastructure with resources spanning the entire computing continuum (Kimovski, Dragi, Roland Matha, Josef Hammer, Narges Mehran, Hermann Hellwagner, and Radu Prodan, 2021).

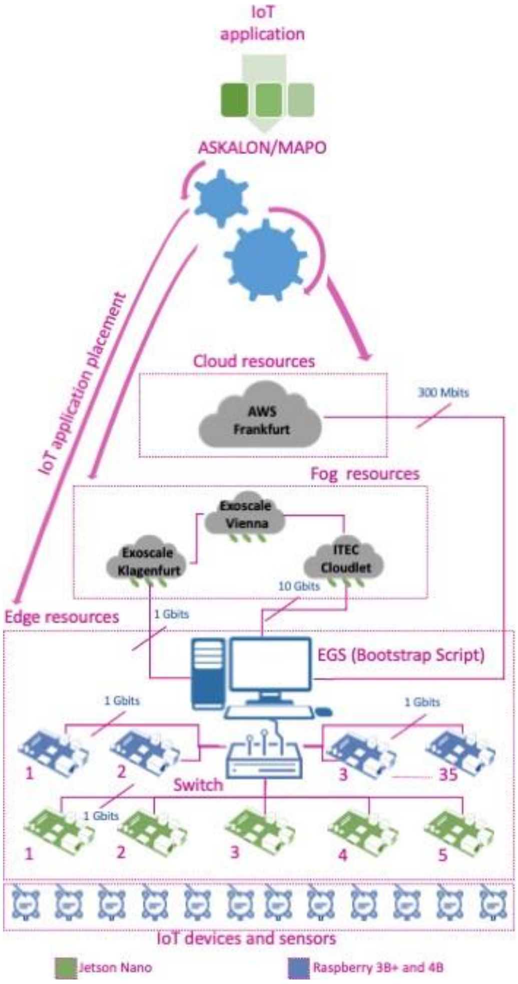

***Fig.2**. Overview of Computing Continuum*

4. THE COMPUTING CONTINUUM MIDDLEWARE FOR ARTIFICAL INTELLIGENCE OVER 6G NETWORKS

Apart from the main network infrastructure problem mentioned in Section 2, there is also an operational problem which needs to be addressed. Actually, the novel IoT applications beyond the 5G networks and 6G networks would require a solid computing infrastructure, such as the computing continuum, which provides a near and rich set of computing devices and smart networking technologies to support their AI computation, storage, and low-latency communication (in 5G and future 6G standards). In addition, the modern AI-based IoT applications suffer from manual orchestration and device communication. They lack advanced middleware support to automate their operation with optimized resource and application modelling, scheduling and performance adaptation in response to low communication latency and throughput constraints.

One possible solution to the network infrastructure problem and the operational problem is our proposed computing continuum middleware for artificial intelligence over 6G mobile networks, that would create an automated, sustainable loop for managing IoT applications utilizing AI approaches over 6G networks, which is depicted in Fig. 3. It consists of Computing Continuum Middleware which is managed and controlled by 6G network management using the distributed edge artificial intelligence.

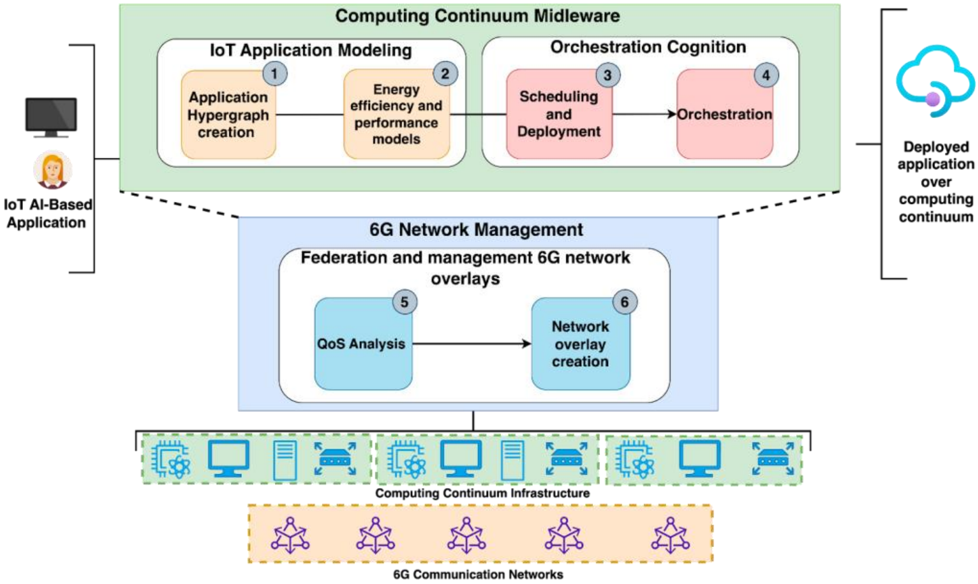

***Fig.3**. Computing Continuum Middleware for Artificial Intelligence over 6G Mobile Networks*

The computing continuum middleware facilitates creating and optimizing the IoT applications' workflows comprising performance and energy models and facilitates the scheduling, deployment and orchestration. It consists of IoT Application Modelling and Orchestration Cognition.

IoT Application Modelling provides the following functionalities: application hypergraph creation and optimization and energy efficiency and performance modelling. The Application hypergraph creation and optimisation analyzes the IoT application with its interactions with the computing continuum, 6G networks and environment and represents them as hypergraphs (HG), which is a powerful higher-order mathematical tool for systems modelling. Unlike classical approaches for managing IoT applications, which model the applications in isolation from the infrastructure and environment, this approach would generalize the applications and the infrastructure as hypergraphs. The Hypergraph models not only the IoT application components but also all multi-dimensional interactions, including the interactions with the environment.

Furthermore, the IoT Application Modelling encompasses energy efficiency analysis by considering multiple sustainability metrics, including energy wastage estimation. It utilizes benchmarks for energy profiling various computing resources available across the computing continuum, including single-board computers, powerful cloud instances and even open-source hardware, including RISC-V-based processors. Moreover, the IoT Application Modelling supports a set of benchmarks for improved hardware performance models, which are essential for improved energy efficiency and sustainable execution.

The Orchestration Cognition would be responsible for the following functionalities: scheduling and deployment and orchestration.

The scheduling and deployment support a preemptive scheduling approach capable of reacting to the changes in the computing infrastructure and the available 6G networks. The scheduling approach utilizes multi-objective heuristics, considering energy efficiency and performance as optimization objectives, to find IoT application scheduling solutions. Besides, it enables transparent deployment of IoT applications using the principle of infrastructure-as-a-code.

The orchestration component enables efficient orchestration and adaptation of the execution of the IoT application concerning the communication performance of the utilized 6G network overlays. It utilizes a Prometheus-based monitoring system to constantly monitor the communication performance of the network and the computing performance of the resources. Based on the gathered monitoring data, the component uses machine learning to detect anomalies in the execution and the resources and adapt the execution in real time. For the given purpose, the component could rely on a Kubernetes orchestrator.

The 6G network management layer enables the federation of multiple network technologies to support IoT application execution with high QoS and dynamic overlay network creation. It consists of QoS analysis and network overlay creation.

To achieve the fundamental goals of 6G, it is necessary to foster higher system capacity, higher data rate, lower latency, and improve the quality of service (QoS). The dynamic QoS analysis through novel, dynamic queue management would imply dynamic queues parameter configuration according to the requirements specified through the service management on the application layer. The QoS traffic management would utilize dynamic queues management and apply mechanisms to make intelligent decisions for max rates of service groups. The QoS analysis would be dynamically performed on the following key performance indicators (KPIs): latency, user data throughput, energy efficiency and power consumption. The KPIs would be dynamically collected from different RATs, and would be analyzed to determine which RAT would offer the best QoS for various profile users. This analysis would be performed by distributed artificial intelligence located at the core and the edge of the 6G network.

Once the QoS analysis is completed, it would be created an automatic network overlay. For that, a virtual or logical network would be created on top of an existing physical network. The overlay creates a new layer where traffic can be programmatically directed through new virtual network routes or paths instead of requiring physical links. Overlays enable administrators to define and manage traffic flows, irrespective of the underlying physical infrastructure. Some of the advantages that network overlay provides over the physical network are scalability, flexibility, management, security, redundancy, and efficiency. The network overlay would provide a better QoS to all users that have different service requirements for IoT applications. The 6G network would be about distributed artificial intelligence at the edge of the 6G network, which would be enabled by using the computing continuum middleware.

4. CONCLUSION

The exponential increase in broadband multimedia wireless communications, as well as the rapid proliferation of smart mobile devices would shape the creation of the future 6G mobile and wireless networks. Terahertz, visible light communication and technologies like compressed sensing theory, new channel coding, large-scale antenna, flexible spectrum usage, AI-based wireless communication and special features as Space-Air-Ground-Sea integrated communication and wireless tactile network are few of the novelties that are expected to become a common network standard of 6G. 6G network should provide scalability, efficient connectivity with connectivity and reliability, coverage improvement, as well as, QoS and QoE provisioning.

The main driving force in designing and optimizing 6G architectures, protocols, and operations is Distributed Artificial Intelligence in the core, radio access, and the network edge, which would provide support of ubiquitous and mobiquitous smart services from the core to the end devices of the network, which would exceed the mobile data traffic used today. It is expected that AI would transform the wireless evolution of future Internet of Everything (IoE) from “connected things or objects” to “connected intelligence.”

The security vision of 6G would integrate AI to produce security automation which would ensure user/data privacy, would ensure trust, would provide prediction, detection and prevention of attacks, and would limit vulnerability propagation. new set of Key Performance Indicators (KPIs) and Key Value Indicators (KVIs) should be defined, in order to assess the value of network security in 6G.

The Computing Continuum encompasses the seamless integration of edge, fog, and cloud computing resources to provide a unified and distributed computing environment, and it aims to enable real-time data processing, low-latency response, and intelligent decision-making at the network edge. The role that 6G networks can play in the transition from the IoT, to the Edge-Cloud 6G Computing Continuum.

The computing continuum middleware for artificial intelligence over 6G mobile networks, that would create an automated, sustainable loop for managing IoT applications utilizing AI approaches over 6G networks.

Although AI learning algorithms provide many benefits in terms of learning and recognition ability, still they require very high level of computational complexity, high power consumption, and many computing and processing resources. Therefore, significant research efforts are needed to advocate the inter-working collaboration and cooperation among hardware components and AI learning algorithms. In addition, scalability, great robustness, and flexibility of learning frameworks are important aspects to provide support of the potential unbounded number of interacting entities and to provide high-quality QoS and QoE services in 6G networks. Thus, the design of scalable, robust and flexible AI learning frameworks for 6G networks is still an open issue.

ACKNOWLEDGEMENT

This work was supported by Shota Rustaveli National Science Foundation of Georgia (SRNSFG) - CG-24-220.

This research paper is a part of the bilateral national research project etitled: 6G Continuum: Computing Continuum Middleware for Artificial Intelligence over 6G Networks, funded by the governments of the Republic of North Macedonia and the Republic of Austria, in which participates the following institutions: Mother Teresa University, Faculty of Information Sciences, Skopje, Republic of North Macedonia and Alpen Audria University of Klagenfurt, Department of Information Technology, Klagenfurt, Republic of Austria.

REFERENCES

1. Balali, Farhad, Jessie Nouri, Adel Nasiri, and Tian Zhao. *Data Intensive Industrial Asset Management*. *Springer eBooks,* 2020. https://doi.org/10.1007/978-3-030-35930-0
2. Choudhury, Alok, Manojit Ghose, Akhirul Islam, and None Yogita. “Machine Learning-based Computation Offloading in Multi-access Edge Computing: A Survey.” *Journal of Systems Architecture* 148 (February 16, 2024): 103090. https://doi.org/10.1016/j.sysarc.2024.103090.
3. Clemm, Alexander, Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck. “Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions.” *IEEE Communications Magazine* 58, no. 1 (January 1, 2020): 93–99. https://doi.org/10.1109/mcom.001.1900272.
4. Du, Jun, Chunxiao Jiang, Jian Wang, Yong Ren, and Merouane Debbah. “Machine Learning for 6G Wireless Networks: Carrying Forward Enhanced Bandwidth, Massive Access, and Ultrareliable/Low-Latency Service.” *IEEE Vehicular Technology Magazine* 15, no. 4 (September 25, 2020): 122–34. https://doi.org/10.1109/mvt.2020.3019650.
5. Hao, Tianshu, Jianfeng Zhan, Kai Hwang, Wanling Gao, and Xu Wen. *AI-oriented Workload Allocation for Cloud-Edge Computing*, 2021. https://doi.org/10.1109/ccgrid51090.2021.00065.
6. Huang, Yakun, Boyuan Bai, Yuanwei Zhu, Xiuquan Qiao, Xiang Su, Lei Yang, and Ping Zhang. “ISCom: Interest-Aware Semantic Communication Scheme for Point Cloud Video Streaming on Metaverse XR Devices.” *IEEE Journal on Selected Areas in Communications* 42, no. 4 (December 22, 2023): 1003–21. https://doi.org/10.1109/jsac.2023.3345430.
7. Jiang, Dajie, and Guangyi Liu. “An Overview of 5G Requirements.” In *Springer eBooks,* 3–26, 2016. https://doi.org/10.1007/978-3-319-34208-5_1.
8. Khan, Atta Ur Rehman, Mazliza Othman, Sajjad Ahmad Madani, and Samee Ullah Khan. “A Survey of Mobile Cloud Computing Application Models.” *IEEE Communications Surveys & Tutorials* 16, no. 1 (July 4, 2013): 393–413. https://doi.org/10.1109/surv.2013.062613.00160.
9. Khan, Latif U., Ibrar Yaqoob, Muhammad Imran, Zhu Han, and Choong Seon Hong. “6G Wireless Systems: A Vision, Architectural Elements, and Future Directions.” *IEEE Access* 8 (January 1, 2020): 147029–44. https://doi.org/10.1109/access.2020.3015289.
10. Kibria, Mirza Golam, Kien Nguyen, Gabriel Porto Villardi, Ou Zhao, Kentaro Ishizu, and Fumihide Kojima. “Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-

Generation Wireless Networks." *IEEE Access* 6 (January 1, 2018): 32328–38. https://doi.org/10.1109/access.2018.2837692 .

11. Kimovski, Dragi, Roland Matha, Josef Hammer, Narges Mehran, Hermann Hellwagner, and Radu Prodan. "Cloud, Fog, or Edge: Where to Compute?" *IEEE Internet Computing* 25, no. 4 (February 9, 2021): 30–36. https://doi.org/10.1109/mic.2021.3050613.
12. Kitanov, Stojan, Edmundo Monteiro, and Toni Janevski. *5G And the Fog — Survey of Related Technologies and Research Directions*, 2016. https://doi.org/10.1109/melcon.2016.7495388.
13. Kitanov, Stojan, Ivan Petrov, and Toni Janevski. "6G MOBILE NETWORKS: RESEARCH TRENDS, CHALLENGES AND POTENTIAL SOLUTIONS." *Journal of Electrical Engineering and Information Technologies* 6, no. 2 (January 1, 2021): 67–77. https://doi.org/10.51466/jeeit2162186067k.
14. Kitanov, Stojan, and Vladimir Nikolikj. "The Role of Edge Artificial Intelligence in 6G Networks." *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*, June 16, 2022, 1–4. https://doi.org/10.1109/icest55168.2022.9828567.
15. Kumar, Jaydip, Jitendra K Samriya, Marek Bolanowski, Andrzej Paszkiewicz, Wiesław Pawłowski, Maria Ganzha, Katarzyna Wasielewska-Michniewska, et al. "Towards 6G-Enabled Edge-Cloud Continuum Computing – Initial Assessment." In *Communications in Computer and Information Science*, 1–15, 2023. https://doi.org/10.1007/978-3-031-25088-0_1.
16. Letaief, Khaled B., Wei Chen, Yuanming Shi, Jun Zhang, and Ying-Jun Angela Zhang. "The Roadmap to 6G: AI Empowered Wireless Networks." *IEEE Communications Magazine* 57, no. 8 (August 1, 2019): 84–90. https://doi.org/10.1109/mcom.2019.1900271.
17. Nawaz, Syed Junaid, Shree Krishna Sharma, Shurjeel Wyne, Mohammad N. Patwary, and Md. Asaduzzaman. "Quantum Machine Learning for 6G Communication Networks: State-of-the-Art and Vision for the Future." *IEEE Access* 7 (January 1, 2019): 46317–50. https://doi.org/10.1109/access.2019.2909490.
18. Rashid, Adib Bin, and Md Ashfakul Karim Kausik. "AI Revolutionizing Industries Worldwide: A Comprehensive Overview of Its Diverse Applications." *Hybrid Advances* 7 (August 23, 2024): 100277. https://doi.org/10.1016/j.hybadv.2024.100277.
19. "Report ITU-R M.2370-0." *IMT Traffic Estimates for the Years 2020 to 2030*, 2015.
20. Saad, Walid, Mehdi Bennis, and Mingzhe Chen. "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems." *IEEE Network* 34, no. 3 (October 15, 2019): 134–42. https://doi.org/10.1109/mnet.001.1900287.
21. Seng, Kah Phooi, Li Minn Ang, and Ericmoore Ngharamike. "Artificial Intelligence Internet of Things: A New Paradigm of Distributed Sensor Networks." *International Journal of Distributed Sensor Networks* 18, no. 3 (March 1, 2022): 155014772110628. https://doi.org/10.1177/15501477211062835.
22. Shahraki, Amin, Mahmoud Abbasi, Md. Jalil Piran, and Amir Taherkordi. "A Comprehensive Survey on 6G Networks:Applications, Core Services, Enabling Technologies, and Future Challenges." arXiv.org, January 29, 2021. https://arxiv.org/abs/2101.12475.
23. Strinati, Emilio Calvanese, Sergio Barbarossa, Jose Luis Gonzalez-Jimenez, Dimitri Ktenas, Nicolas Cassiau, Luc Maret, and Cedric Dehos. "6G: The Next Frontier: From Holographic Messaging to Artificial Intelligence Using Subterahertz and Visible Light Communication." *IEEE Vehicular Technology Magazine* 14, no. 3 (August 8, 2019): 42–50. https://doi.org/10.1109/mvt.2019.2921162.
24. Tang, Fengxiao, Yuichi Kawamoto, Nei Kato, and Jiajia Liu. "Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches." *Proceedings of the IEEE* 108, no. 2 (December 6, 2019): 292–307. https://doi.org/10.1109/jproc.2019.2954595.
25. Tariq, Faisal, Muhammad R. A. Khandaker, Kai-Kit Wong, Muhammad A. Imran, Mehdi Bennis, and Merouane Debbah. "A Speculative Study on 6G." *IEEE Wireless Communications* 27, no. 4 (August 1, 2020): 118–25. https://doi.org/10.1109/mwc.001.1900488.
26. Tarneberg, William, Emma Fitzgerald, Monowar Bhuyan, Paul Townend, Karl-Erik Arzen, Per-Olov Ostberg, Erik Elmroth, Johan Eker, Fredrik Tufvesson, and Maria Kihl. *The 6G*

Computing Continuum (6GCC): Meeting the 6G Computing Challenges, 2022. https://doi.org/10.1109/6gnet54646.2022.9830459.

27. “Transforming Our World: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs,” n.d. https://sdgs.un.org/2030agenda.
28. Zhang, Shangwei, Jiajia Liu, Hongzhi Guo, Mingping Qi, and Nei Kato. “Envisioning Device-to-Device Communications in 6G.” IEEE Network 34, no. 3 (March 27, 2020): 86–91. https://doi.org/10.1109/mnet.001.1900652.