Toward the 6G Network Era: Opportunities and Challenges

Ioannis Tomkos
University of Patras

Dimitrios Klonidis
Ubitech

Evangelos Pikasis
EULAMBIA Advanced Technologies Ltd

Sergios Theodoridis
National and Kapodistrian University of Athens

***Abstract*—The next generation of telecommunication networks will integrate the latest developments and emerging advancements in telecommunications connectivity infrastructures. In this article, we discuss the transformation and convergence of the fifth-generation (5G) mobile network and the internet of things technologies, toward the emergence of the smart sixth-generation (6G) networks which will employ AI to optimize and automate their operation.**

In recent years, we have experienced a tremendous evolution in information and communications technology with major breakthroughs in mobile networks. Early technology developments related with 5G mobile networks targeted mainly the three generic and distinct use cases: "enhanced mobile broadband" (eMBB), "massive machine-type communications" and "ultra-reliable low-latency communications."[1] The associated applications have stringent specifications in terms of critical system parameters like number of connected devices, power consumption, bit-rate, latency and availability, among others. Each family of those use cases is focusing primarily on optimization of one key design parameter (e.g., bit-rate or number of connected devices or latency).

Many applications can be enabled by early 5G networks, such as real-time closed-loop robotic control, video-driven machine–human interaction, and augmented reality/virtual reality (AR/VR) applications with high definition 360° video streaming with high bit rate requirements (in the range of tens of gigabit per second) and low latency (<10 ms). It is foreseen that evolving 5G networks will enable even more demanding

Digital Object Identifier 10.1109/MITP.2019.2963491
Date of current version 12 February 2020.

applications, such as sharing and updating high-resolution maps in real time for control of autonomous vehicles. This is an example of the most demanding use-case among the 5G use-cases family, called ultra-high-speed low-latency communications.[1]

To support the requirements of current and emerging use cases, we need to move beyond current 5G networks, which operate at low (sub-6 GHz) radio frequency bands, toward millimeter (mm)-wave ones that operate at higher carrier frequencies, initially at the high radio frequency window, i.e., above 20 GHz, and then even higher toward optical frequencies in the visible and invisible electromagnetic waves spectrum.[2] Only then, the true capabilities of 5G networks will be gradually revealed. In the subsequent evolution stage, the technologies of both 5G networks and next-generation internet of things (NG-IoT) will move toward some form of convergence, where the use of mm-wave frequencies will become commonplace.[3] The NG-IoT will combine technologies such as AI, distributed edge computing, and end-to-end distributed security supporting high bit rates serving eMBB use cases. Besides the evolution of wireless connectivity at the edge of the network, we anticipate major advances also on the control and management of the edge network that will be supported by the introduction of new technology approaches driven by artificial intelligence (AI) solutions, going well beyond what can be achieved today with software-defined-networking (SDN) and network functions virtualization (NFV). The ultimate goal will be to enhance all the applications and vertical use cases of 5G networks with the capability of cognition offered by cognitive computing systems that use the vast amounts of available data and Machine Learning (ML) algorithms.[1,2]

FROM MOBILE-EDGE-COMPUTING TO AI-AT-THE-EDGE

The advent of the edge computing concept, in combination with the vastly increased number of data processing devices at the edge and the emerging 5G networks, leads to a distributed mobile edge computing (MEC) environment that will enable the efficient implementation of 5G use cases in vertical markets. *MEC* has gained significant momentum recently as it transforms 5G mobile communication networks into distributed cloud computing platforms. Located in close proximity to the edge devices MEC enables low latency service delivery in the order of microseconds, over the radio access network (RAN), to/from mobile users and the IoT.

The anticipated future requirements for introducing intelligence at the edge devices will lead to the evolution of MEC toward an AI-enabled platform that will be capable of offering intelligent services delivered over the fixed or mobile access network to the edge devices. Such devices can include, in the future, *computationally efficient dedicated hardware capable of running locally ML/AI Algorithms at the edge devices, deviating from the* classical concept of ML/AI; the latter focuses mainly on offline and centralized AI/ML and is implemented by a cloud computing model, in which the entire dataset is given *a priori* and is used for the training stage.

To obtain accurate and reliable inference, a central controller divides the training dataset into mini-batches and allocates them to multiple processing devices. The central controller iteratively collects and aggregates their local training results, until a cost function for the training converges. However, such a one-time training process is vulnerable to initially unmodeled phenomena.The developed ML/AI algorithms implemented over such cloud-computing model come with a number of drawbacks, e.g., their need for large training datasets, their difficulty in coping satisfactorily with dynamically changing environments, their poor performance in generalizing their “knowledge” to new datasets without forgetting the previous one, and, most importantly, their huge computational costs that make them unsuitable for edge computing.[4]

DISTRIBUTED AND FEDERATED AI

In contrast to cloud-based architectures, recent research investigations have sparked a huge interest in collaborative, distributed, low-latency, and reliable ML calling for a major departure from cloud-based and centralized training and inference, toward a novel system

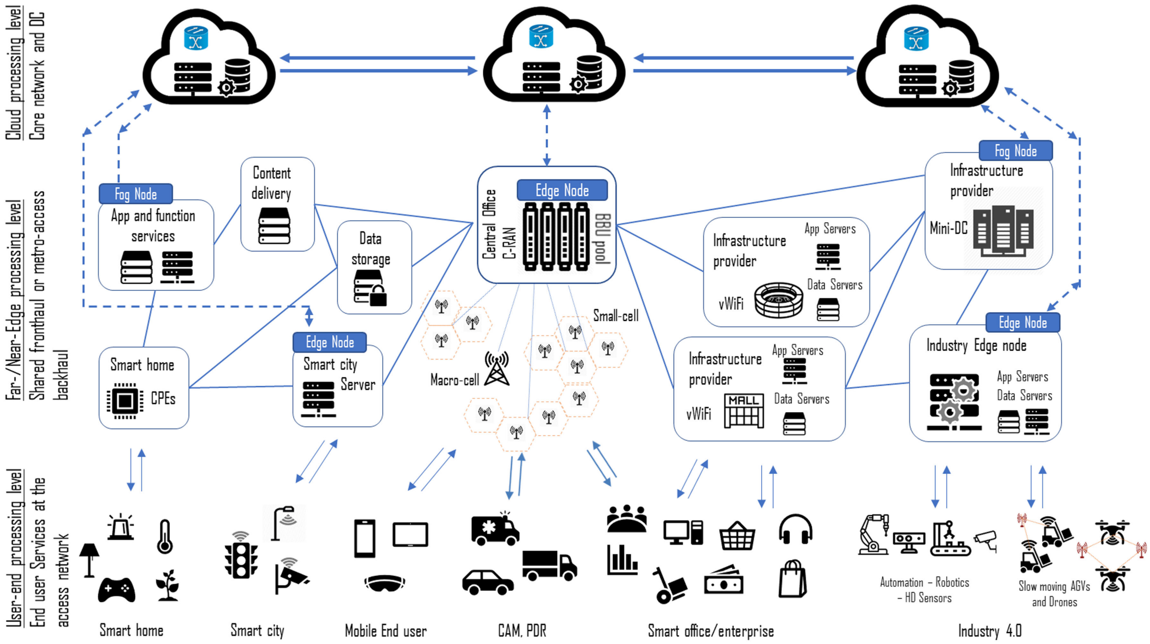

Figure. 1. Diverse set of service provisioning fields addressed by the edge computing concept in a generic 6G network model at the access with interconnection to the legacy centralized edge processing node and cloud.

design coined the term "AI at the Edge," in which we have the following: i) Training data is unevenly distributed over a large number of edge devices, such as network base stations (BSs) and/or mobile devices, including phones, cameras, vehicles, drones, etc.; ii) every edge device has access to a fraction of the data and a limited computation and storage power, where edge devices exchange their locally trained models, instead of exchanging their private data and wherein training and inference are carried out collectively; and finally, iii) data abstraction, cleansing, and dimensionality reduction of network data become vital, as the massive amount of monitored data cannot be stored.

It becomes evident that AI at the edge is a nascent research field, whose system design is entangled with secure and reliable communications and on-device resource constraints (i.e., energy, memory, and computing power). This architecture allows the deployment of an evolved 5G/6G network that can operate as a distributed computer with the goal to offer AI at the edge and which will be deployed between the cloud processing level (the network core) and the connected end users/devices (at the wireless access). It will be a platform composed by diverse types of nodes and infrastructures with different processing capabilities that are either edge nodes (e.g., industrial or enterprise nodes, smart city servers, smart home servers, customer premises equipment (CPEs0, and private infrastructure gateways) or network services nodes (e.g., mini-data center (DCs), application servers, content delivery, and data storage nodes).

The concept relies on the dynamic formation of collaborative edge computing domains (ECDs) at or close to the end user that are smartly interconnected and managed over the metro-access wireless-optical infrastructure. The ECDs can be composed by one or several edge nodes, with a diverse set of processing capabilities (i.e., from simple Raspberry Pis to nano-/mini-DCs) that can provide the required computational resources and data pools to handle a set of running services. As the service requirements vary, additional resources, data, or applications may be needed from other ECDs (i.e., service expansion). The overall architecture is depicted in Figure 1.

Enabling ML at the network edge introduces novel fundamental research problems in terms of jointly optimizing training, communication, and control under end-to-end (E2E) latency, reliability, security, privacy, and trustworthiness specifications, as well as edge devices' hardware requirements. For practical edge ML, online decentralized training can simplify the implementation and preserve privacy by exchanging not only the entire dataset but the model-related parameters with (or without) a simple parameter aggregator, thereby reflecting a huge volume of user-generated data samples in real time.

A major effort on the "AI at the edge" front is to develop efficient schemes for distributed optimization among a large number of devices that share a common model to be trained, via data that are distributed in a large number of interconnected devices and by utilizing certain models.[5] Such an interesting ML training architecture following a decentralized processing rationale is the federated learning approach, in which mobile devices periodically exchange their neural network (NN) weights and gradients during local training.

UTILIZING COMMUNICATIONS FOR ML (CML) AND ML FOR COMMUNICATIONS (MLC) IN THE FUTURE 6G SMART NETWORKS

From the aforementioned discussions, it becomes clear that high speed, low latency, and reliable communications are essential for supporting ML/AI at the edge; giving rise to the research field entitled *CML*.[1] Going a step further from what has been discussed in the previous section, we note that the use of the next generation of wireless connectivity at the edge of the network can enable far more advanced concepts than the described AI at the edge.

Future generations of wireless communication systems (allocating in excess of 10 GHz RF channels in the THz and optical frequencies regime, while exploiting modulation methods capable of 10 b/symbol), can achieve data rates of about 100 Tb/s.[2] With such assumptions in mind, one can envision as plausible in the future the realization of the concept of wireless cognition, that is, to provide a wireless communications link that can enable massive computations (matching the capabilities of the human brain) to be conducted remotely from the device or machine that is undertaking real-time intelligent actions at the edge of the network. If that vision indeed becomes a reality, then, in the future, robots, autonomous vehicles, IoT devices, and other machines may be designed to exploit cognitive processing performed remotely from them, at a platform/server that will be collocated with a 6G BS and which will be in wireless connection with each other.[2]

Furthermore, exploiting edge ML for improving the performance of communication networks is another research area that epitomizes the research direction of *MLC*,[1] wherein ML can be used to optimize the operation of communication systems and networks.[5] An intelligent AI-enabled future 6G mobile network infrastructure will make the best use of the physical and computing resources at the edge of the network to offer intelligent systems supporting an entirely new generation of services. The ML/AI algorithms may also assist the software-defined control and management plane of the communication network in order to, e.g., i) predict the required resources, based on a number of intrinsic networking parameter or external (e.g., environmental, social, etc.) factors, ii) identify potential physical failures, device malfunctions, and even malicious attacks and usage, and iii) dynamically vary the smart service level agreement (SLAs), or any other functionalities that would benefit from AI for future use. The research on CML and MLC are closely related and interlinked. These two research directions (i.e., MLC and CML) will be instrumental in spearheading the vision of truly intelligent next generation communications systems (i.e., the 6G mobile networks).[1]

CONCLUSION

The smart 6G networks of the future will integrate intelligent 5G and NG-IoT connectivity infrastructures, with advanced edge-computing hardware that will support the computational-resources' heavy execution of the AI algorithms. The introduction of AI at the network edge faces many challenges related to the fact that its targeted applications must operate in real-time dynamic environments, which should comply with strict end-to-end bit-rate, latency, reliability, privacy/security requirements, as well as edge devices' limitations like energy availability, battery lifetime, and memory size. In this 6G of mobile network, the communications infrastructure connecting the foreseen billions of smart edge devices (collecting and distributing data) should be seen as its "nervous system," while the edge computing hardware executing AI and other algorithms (like, e.g., blockchain for trustworthiness and security) over an SDN/NFV software control and management platform should be seen as its "brain." Such evolution, combined with the proliferation of smart (and wearable) wireless devices at the edge, will make the internet of everything (IoE) a reality. This highlights our vision for the evolution of current 5G and IoT infrastructures toward the smart converged 6G infrastructure that is expected to emerge after 2025, supporting in parallel the actual deployment of the fourth industrial revolution (Industry 4.0), autonomous driving, smart city/building services, AR services, and more that will be able to reshape our way of living.

REFERENCES

1. J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless network intelligence at the edge," 2018, *arXiv:1812.02858*.
2. T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.
3. M. H. Mazaheri, S. Ameli, A. Abedi, and O. Abari, "A millimeter wave network for billions of things." in *Proc. ACM Special Interest Group Data Commun.*, 2019.
4. S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. 2nd ed. San Francisco, CA, USA: Academic, 2020.
5. J. Plata-Chaves, A. Bertrand, M. Moonen, S. Theodoridis, and A. M. Zoubir, "Heterogeneous and multitask wireless sensor networks—Algorithms, applications, and challenges," *IEEE J. Sel. Topics* , vol. 11, no. 3, pp. 450–465, Apr. 2017.
6. S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning-based communication over the air," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018.

Ioannis Tomkos is currently a Professor of Optical Communications with the Department of Electrical and Computer Engineering, University of Patras, Patras, Greece. In the past, he was Professor and Research Director of Next Generation Networks with AIT (Athens, Greece). His "High Speed Networks and Optical Communication (NOC)" Research Group was/is involved in more than 25 EU-funded research projects with a consortium-wide leading role. He is a Fellow of the IEEE, a Fellow of the IET and a Fellow of the "Optical Society." Together with his colleagues and students, he has authored more than 650 peer-reviewed archival articles. Contact him at itom@ait.gr.

Dimitrios Klonid is currently the Head of the Network Softwarization and IoT Unit with Ubitech. He received the Ph.D. degree in optical communications from Essex University, Colchester, U.K., in 2006. He has been involved in numerous EU-funded research projects, gaining expertise in various areas, including optical network planning control and performance evaluation. He has coauthored more than 150 publications in international journals and conferences. Contact him at dklonidis@ubitech.eu.

Evangelos Pikasis is currently a DSP and Telecom Engineer with EULAMBIA Advanced Technologies Ltd., Greece. He received the Ph.D. degree in the area of optical communications from the National and Kapodistrian University of Athens, Athens, Greece, in 2014. His area of expertise includes optical communication systems, analog radio over fiber, advanced modulation schemes, digital signal processing, FPGAs, and cryptography at the physical layer. Contact him at evangelos.pikasis@eulambia.com.

Sergios Theodoridis is currently a Professor of Signal Processing and Machine Learning at National and Kapodistrian University of Athens, Greece. He is the author of the book *Machine Learning: A Bayesian and Optimization Perspective* (Academic, 2nd ed., 2020). He received the 2017 EURASIP Athanasios Papoulis Award, the 2014 IEEE Signal Processing Society Education Award, and the 2014 EURASIP Meritorious Service Award. He currently serves as Vice President of the IEEE Signal Processing Society. He is a Fellow of IET, a Corresponding Fellow of the Royal Society of Edinburgh, a Fellow of EURASIP, and a Fellow of IEEE. Contact him at stheodor@di.uoa.gr.