Multi-Access Edge Computing: An Overview and Latency Evaluation

Igor Miladinovic
Computer Science and Digital Communications
University of Applied Sciences Campus Vienna
Vienna, Austria
igor.miladinovic@fh-campuswien.ac.at

Sigrid Schefer-Wenzl
Computer Science and Digital Communications
University of Applied Sciences Campus Vienna
Vienna, Austria
sigrid.schefer-wenzl@fh-campuswien.ac.at

Thomas Burger
Computer Science and Digital Communications
University of Applied Sciences Campus Vienna
Vienna, Austria
thomas.burger@stud.fh-campuswien.ac.at

Heimo Hirner
Computer Science and Digital Communications
University of Applied Sciences Campus Vienna
Vienna, Austria
heimo.hirner@fh-campuswien.ac.at

2021 22nd IEEE International Conference on Industrial Technology (ICIT) | 978-1-7281-5730-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICIT46573.2021.9453495

Abstract—**Multi-Access Edge Computing (MEC) is one of the key emerging technologies currently being deployed by mobile network operators. By introducing cloud application capabilities close to end-users, MEC enables a range of new use cases, in particular those with real-time requirements. This is possible because of low network latency of MEC compared with traditional cloud services. In this paper we evaluate the network latency of one service running in a MEC-like environment and compare it with the latency of the same service running in two geographically different cloud locations. Our goal was to quantify MEC latency, as one of the most important benefits of MEC. The results show an essential advantage of MEC over cloud computing, which makes it a promising technology for real-time services and beyond.**

Index Terms—**Multi-access Edge Computing (MEC), IoT, Latency, Round Trip Time, 5G**

I. Introduction

Communication networks are expected to cope with a rising number of digital services in the near future. An important enabler for these services are connected devices capable to communicate without human interaction, also known as Internet of Things (IoT) devices or smart devices. Forecasts by Gartner [1], IDC [2], or Ericsson [3] show all in the same direction—several tens of billions of IoT devices in the near future. To manage these devices and the associated growth in network traffic, as predicted by Cisco in [4], today's communications networks need to evolve providing a new architecture capable to serve people and things optimally [5]. Data from various sources, such as road sensors, security cameras, vehicles, smart home devices, industry sensors and actuators need to be collected, analyzed, and evaluated in order to gain the current status of these resources and services, to identify actions needed for an improvement or optimization, and to perform these actions. This data must be managed and transmitted efficiently, to avoid an explosion of network traffic.

The new generation of mobile networks, 5G, is supposed to be able to fulfil these requirements. Being more than just another "G", 5G goes beyond higher data rates for mobile Internet [6], [7], and enables new services by supporting massive machine-type communication for IoT, with a very large number of connected devices operating with low power consumption, or ultra-reliable low latency communications [8]. Therefore, 5G is supposed to be capable of optimally serving not only people, but also things, becoming the ultimate access technology of IoT. In particular, IoT for Industry (also called Industry 4.0) and smart cities will benefit from 5G as the enabling technology connecting many sensors and actuators wirelessly. With additional new capabilities, such as Network Slicing [9], [10], 5G is reducing, if not eliminating, the need for deployment of new physical networks dedicated for a single company or industry.

One of the innovative 5G technologies is Multi-Access Edge Computing (MEC) [11], [12]. It is designed to improve Quality of Experience (QoE) and to reduce network traffic. By moving applications from the cloud near to end-users, MEC is capable of serving time-critical applications, where a cloud solution—due to relatively high latency—is not eligible. At the same time, MEC also reduces the network traffic, because not all communication traffic has to go into the cloud. However, MEC is not a 5G-only technology as it can be introduced in existing 4G networks—which are expected to remain successful in the next years. Moreover, the MEC reference architecture [13] is agnostic towards mobile network generations, so that it can be deployed in 4G now, and reused in 5G later on [14].

In our previous work we mainly focused on network traffic optimization by MEC. We have designed a layered IoT architecture for smart cities and considered network traffic optimization by dynamically selecting the optimal application allocation between MEC and the cloud. A Machine Learning (ML) module continuously monitors potential traffic savings of each application, predicting the saving in the near future, and taking decisions which applications to move to MEC and when [15], [16].

978-1-7281-5730-6/21/$31.00 ©2021 IEEE

In this paper we evaluate latency of an application running in different cloud locations and in a MEC-like environment. Our aim is to demystify the latency benefit of MEC and to show what is the potential of MEC for real-time applications. Our test application is running in a near and a far cloud location and in a MEC-like environment. The results enable a quantification of latency gains by MEC.

The remainder of this paper is structured as follows. Section II gives an overview of MEC including motivation, purpose and major benefits. Section III introduces our proposed IoT architecture, and demonstrates how MEC is integrated into this architecture. Subsequently, Section IV presents our approach for latency evaluation and discusses the results. Finally, Section V emphasizes the most important findings of this paper and gives an outlook on future work.

II. Multi-Access Edge Computing

A. Limitations of Cloud Computing

The cloud approach enables very high storage and processing capabilities, high scalability, automatic upgrades and updates of applications, potentially high security, and a centralized view for decision making. Besides these advantages, using the cloud for IoT applications results in some drawbacks [17]. The most critical one is the latency. The propagation time in optical fibre depends on the refractive index, which is different for different wavelengths. However, it is slightly below 1.5 resulting in a propagation time in optical fibre of approximately $\frac{2}{3}$ of the light speed in vacuum (more exactly, it amounts to 204,357 km/s for 1310 nm wavelength). Therefore, a distance of 5,000 km to a cloud data center causes a Round Trip Time (RTT) of approximately 45 ms by the propagation time only ($RTT = 2 * \frac{5{,}000\ km}{204{,}357\ km/s}$). There are usually multiple hops towards a cloud data center and each of them adds some processing time, resulting in a latency which is not acceptable for time-critical applications. Moreover, as the processing time and the route are not constant, jitter is also potentially high for applications running in the cloud. Other drawbacks of the cloud include increased network traffic, as the data first has to be transferred to the cloud and after processing back to the IoT device, and a high dependency on the network availability.

B. Fog and Edge Computing vs. MEC

To address this issue, smaller data centers can be introduced and placed between the cloud and IoT devices, usually near the IoT devices. This decentralized computing infrastructure is referred to as fog computing [18]. However, these smaller data centers have limited resources, including computing power and storage capacity [19]. Furthermore, the complexity is increased, as applications running in several data centers need to be synchronized. The distributed architecture of fog computing is potentially more vulnerable, as any of data centers can be the target of a cyber-attack. Edge computing, in contrast, requires more complex IoT devices with some computation power, as the applications are running directly on IoT devices. It provides even lower latency without the need of permanent connectivity.

Inspired by the benefits of fog and edge computing and aiming to improve the latency of applications in 5G networks, the European Telecommunications Standards Institute (ETSI) MEC Industry Specification Group (ISG) [13] specified an emerging technology called Multi-access Edge Computing (MEC) [12]. The concept of MEC is to offer cloud computing functionalities within the Radio Access Network (RAN), very close to end customers and IoT devices [20]. This reduces the latency and the network traffic compared with the cloud approach, and increases availability—as the same application can be deployed several times on different RANs [21]. As MEC enables computing close to end-devices outside the core of the network, it can be seen as a special case of fog computing.

Originally, MEC stood for Mobile Edge Computing, as it was first developed for mobile networks. As MEC development advanced, it was clear that it can also be used with other access technologies. Therefore, ETSI changed its name from Mobile Edge Computing to Multi-access Edge Computing.

C. Major Benefits

MEC can be characterized by several properties, including [22], [23]:

- *Lower network latency* than for cloud computing, due to allocation close to end-users.
- *On-premises* operation, capable of operating independently of the availability of the rest of the network.
- *Proximity* enabling MEC to obtain geographically targeted information from end-users for further analysis, without using any other location service.
- *Location awareness* by utilizing signalling information more accurate location of users can be estimated.

Out of the first two properties in particular, we can summarize the major benefits of MEC compared to cloud computing: (1) lower latency, (2) reduction of network traffic, and (3) higher availability. Benefits (2) and (3) can be compensated by the network to some extent, by network dimensioning for more traffic as well as by site- and geo-redundancy of critical network components. In contrast, benefit (1) remains to be out of reach for cloud computing. Therefore, in this paper we further evaluate this benefit, as it can be essential for the success of MEC.

D. The Role of MEC for Network Operators

During the last decades telecommunications network providers have been investing a lot in the networks to ensure connectivity, availability and quality of service. However, the majority of applications are cloud based and can be used independently of the network providers, making them easily replaceable and downgrading them to connectivity providers only—also called bit-pipe.

In the past network providers already started initiatives to increase their value proposition. A prominent example is application enablement [24], where several network provides tried to compete with cloud players by opening the network via defined APIs and creating their own ecosystem for development and management of new applications. However, it was not successful as developers were able to reach more customers over ecosystems of global players, such as Google and Apple.

With MEC, mobile network providers gain an important asset from the architectural point of view. Therefore, they can enhance their customer value, and become a smart-pipe instead of a bit-pipe. In contrast to application enablement, network provides are not directly competing with global players, but enhancing the offerings of global players and extending them with own services. Furthermore, as several major network providers are currently deploying MEC in their networks, network providers not offering MEC will have a competitive disadvantage in the future.

III. Integrating MEC into an IoT Architecture

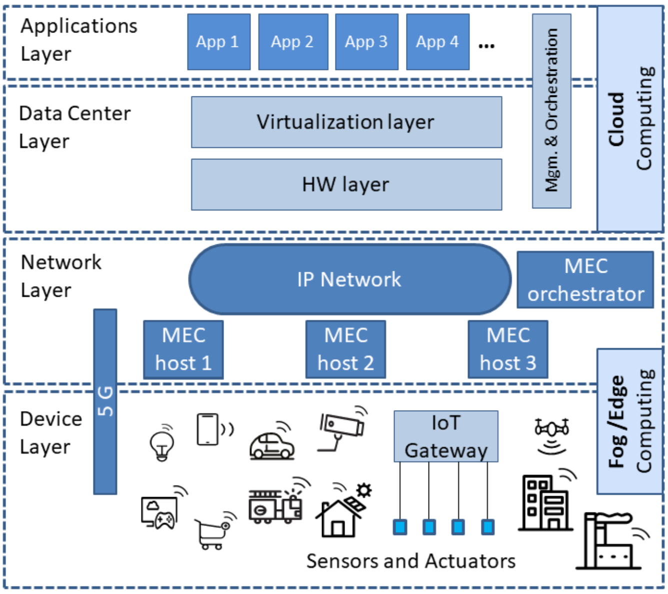

Fig. 1. Proposed IoT architecture with MEC

Fig. 1 shows our proposed IoT architecture with MEC, taking into account the ETSI MEC reference architecture from [13]. The architecture comprises the following four layers:

Device Layer is the layer where IoT end-devices are placed. The connectivity of these devices is either over IP directly or—in the case of less sophisticated devices—over a dedicated IoT gateway. However, the number of constrained devices supporting IP connectivity is growing, also driven by 5G networks, which reduces the need for IoT gateways.

Network Layer is the layer where the network operator's routers and switches are building up an IP network for the connectivity of IoT devices with the Internet. MEC is placed in this layer, including MEC hosts and the MEC orchestrator.

Data Center Layer enables applications running in the cloud. It is composed of three sublayers: (1) HW layer, containing commercial off-the-shelf (COTS) hardware components, including storage, network interfaces and CPUs. It provides a scalable and elastic hardware platform shared among all IoT applications and other network functions; (2) Virtualization layer is an abstraction layer, which provides virtual machines and/or containers towards the IoT applications; (3) Management and orchestration layer is responsible for the lifecycle management of IoT applications and for the coordination of the resources and different IoT applications.

Applications Layer is the place where different IoT applications are running in the cloud.

In the network there are multiple MEC hosts, which are placed near to end-users. Typically, they are co-located with mobile base stations, but they can also be deployed in different locations, as stated in [25]. MEC orchestrator provides the core management functionalities of a MEC system, including [13]: (1) monitoring the overall MEC system, such as MEC hosts, topology, and available resources, (2) on-boarding of application packages and their validation, (3) selection of MEC hosts for application installation, (4) initialisation of application installation and termination, and (5) initialisation of application reallocation.

IV. Evaluation of Network Latency

In this section we evaluate one of the most important benefits of MEC—the network delay between a client and a server. The client is always located in Vienna/Austria and the server in three different locations: near to the client simulating MEC, in Frankfurt/Germany (approximately 600 km air-line distance from the client), and New York/USA (approximately 6,730 km air-line distance from the client). To be able to measure transmission time precisely, we use Network Time Protocol (NTP) [26] for the time synchronization between the client and each server. With the exception of the MEC scenario, all servers were running on Amazon Web Services (AWS). In the MEC scenario the server was running on a dedicated machine in the same LAN as the client.

The first phase for all measurements is synchronization of the server time. The client sets up a *GET-Request* and sends it to the REST end point */api/testing/synchronizeTime* of the server. This triggers the server to synchronize its time using NTP. At the end of this first phase, the client receives a response about the success of the synchronization.

Fig. 2 illustrates the steps of the second phase for our measurements in all scenarios. If the client starts the measurements, the client also synchronizes its time via NTP. Afterwards, the client sends a request to the REST end point */api/testing/pingDatabase* of the server. Shortly before the request is sent, we record the first time as Point 1. As soon as the server receives the request, we again record the time as Point 2. Then the server sets up a query to the database and waits for the response. After receiving the response from the database the time is recorded as Point 3. The time stamps from the server (Point 2 and Point 3) are transported as a JSON object in the response to the client. After receiving this response, the current time is recorded as Point 4.

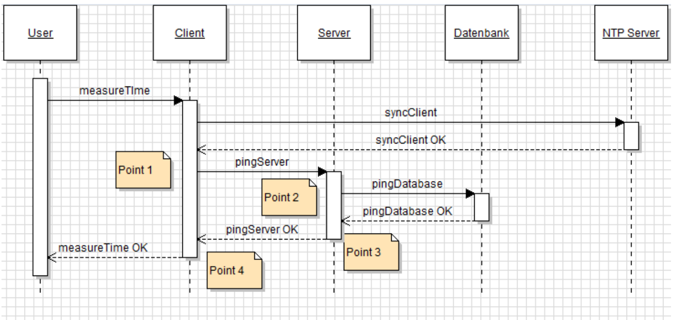

Fig. 2. Measurement sequence for all scenarios

Our objective was to measure the Round Trip Time (RTT) between a client request and a server response. As we focused on the RTT, the processing time on the server should not be considered. Therefore, we excluded this time, represented by $P_3 - P_2$, from the time between a request and the corresponding response at the client ($P_4 - P_1$). So, we calculated the RTT according to Eq. 1, where P_n represents the time measured at Point n (see Fig. 2).

$$\begin{aligned} RTT &= P_4 - P_1 - (P_3 - P_2) \\ &= P_4 - P_3 + P_2 - P_1 \end{aligned} \tag{1}$$

First, we performed 30 measurements for each of the following scenarios:

- Scenario 1 with the server running at AWS in Frankfurt/Germany,
- Scenario 2 with the server running at AWS in New York/USA, and
- Scenario 3 with the server running in a MEC-line environment.

Fig. 3 summarises the results. In Scenario 1, we observed an average RTT of 50.27 ms with a standard deviation of 2.88 ms and in Scenario 2, we measured an average RTT of 217.43 ms with a standard deviation of 4.01 ms. In contrast to that, in the MEC scenario, we measured an average RTT of 10.93 ms with a standard deviation of 2.20 ms.

These results show the potential of MEC, in particular for time-critical applications. Even for a server located in Frankfurt, the measured RTT reduction through MEC amounts in average to 78.25%. Besides that, the standard deviation is lower, so that the variability of the RTT is also lower. This makes MEC an enabler for a range of time-critical applications, which cannot be deployed in the cloud. Examples include augmented reality, connected cars, and intelligent video acceleration [27].

In our final experiment, we wanted to investigate the dependency of the RTT on the time of day. In particular, the impact of network usage by business and residential customers on the network conditions could influence the RTT in cloud scenarios (Scenario 1 and 2). We performed the same measurements as

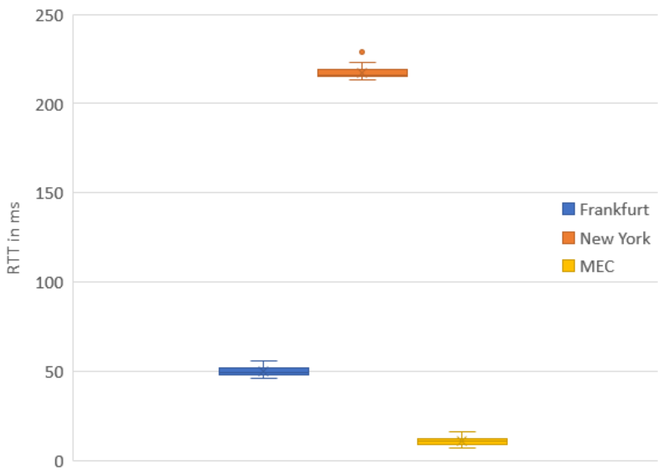

Fig. 3. RTT for the three different scenarios for a user located in Vienna/Austria

before, but at around 6:00 AM, 1:00 PM and 8:00 PM local time in Frankfurt and New York. Fig. 4 depicts the results.

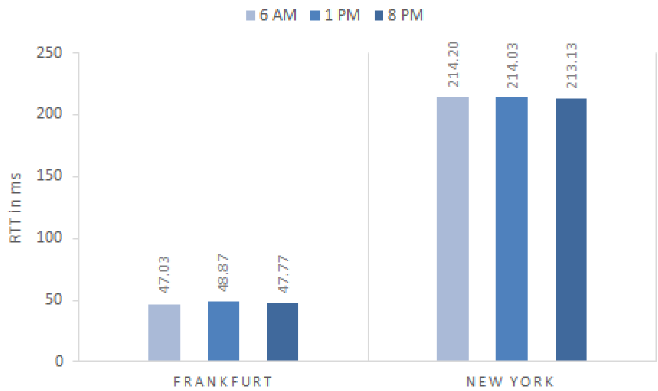

Fig. 4. RTT at different times of day a user located in Vienna/Austria

In our experiment we were not able to measure any remarkable difference in RTT at different times of the day, with 3.75% maximum difference in Frankfurt and 0.5% in New York. This indicates that today's networks are able to manage different degrees of network load and/or that today's Internet traffic is nearly constant during the day—which is concordant with the current Internet traffic report [28]. This is a positive sign for cloud computing, so that the amount of RTT remains the major benefit of MEC, but not its variability due to the time of day.

V. Conclusion

MEC is a promising technology driven by the deployment of 5G networks. There are several benefits of MEC, and one of the key benefits over cloud computing is a lower latency. In this paper we demonstrated the potential of MEC by comparing the network latency, represented by the RTT, of the same application running in a MEC-like environment and in two cloud locations, near and far away from the end-user.

Our results show that MEC clearly outperforms both cloud locations, providing a RTT reduction of almost five times compared to the near cloud location. This enables a range of time-critical applications to be placed in the network, which was not possible with the traditional cloud approach.

In our future work we plan to investigate the impact of lower latency for non time-critical applications, by evaluating QoE of different applications running in MEC and in the cloud. Furthermore, we intend to extend our measurements with a MEC system of a mobile network operator in Austria, as soon as this infrastructure will be available.

References

[1] Gartner Inc., "Gartner Says 5.8 Billion Enterprise and Automotive IoT Endpoints Will Be in Use in 2020," https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io, Aug. 2019.

[2] C. MacGillivray and D. Reinsel, "Worldwide Global DataSphere IoT Device and Data Forecast, 2019–2023," IDC Market Forecast - Doc US45066919, May 2019.

[3] Ericsson, "Ericsson Mobility Report," Report, Jun 2020.

[4] Cisco, "Cisco Annual Internet Report (2018–2023)," Cisco, Tech. Rep., 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf

[5] M. K. Weldon, *The Future X Network, A Bell Labs Perspective*. CRC Press, 2016.

[6] GSMA Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," White Paper, Dec. 2015.

[7] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.

[8] ITU-R, "IMT Vision –Framework and overall objectives of the future development of IMT for 2020 and beyond," Recommendation ITU-R M.2083-0, Sep. 2015.

[9] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.

[10] "NGMN 5G White Paper," San Diego, CA, USA, Feb. 2015.

[11] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[12] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[13] MEC ETSI ISG, "Multi-access Edge Computing (MEC); Framework and Reference Architecture," ETSI GS MEC 003 V2.1.1 (2019-01), 2019. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/02.01.01_60/gs_MEC003v020101p.pdf

[14] F. Giust, G. Verin, K. Antevski, J. Chou, Y. Fang, W. Featherstone, F. Fontes, D. Frydman, A. Li, A. Manzalini, D. Purkayastha, D. Sabella, C. Wehner, K.-W. Wen, and Z. Zhou, "MEC Deployments in 4G and Evolution Towards 5G," ETSI, Tech. Rep., Feb. 2018. [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp24_MEC_deployment_in_4G_5G_FINAL.pdf

[15] I. Miladinovic and S. Schefer-Wenzl, "Towards Intelligent Multi-Access Edge Computing Using Machine Learning," in *International Conference on Interactive Mobile Communication Technologies and Learning*, 2019.

[16] I. Miladinovic, S. Schefer-Wenzl, and H. Hirner, "IoT Architecture for Smart Cities Leveraging Machine Learning and SDN," in *2019 27th Telecommunications Forum (TELFOR)*, 2019, pp. 1–4.

[17] I. Miladinovic and S. Schefer-Wenzl, "A Highly Scalable IoT Architecture through Network Function Virtualization," *Open Journal of Internet Of Things (OJIOT)*, vol. 3, no. 1, pp. 127–135, 2017, special Issue: Proceedings of the International Workshop on Very Large Internet of Things (VLIoT 2017) in conjunction with the VLDB 2017 Conference in Munich, Germany. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:101:1-2017080613543

[18] M. Chiang and T. Zhang, "Fog and IoT: An Overview of Research Opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, Dec 2016.

[19] D. Hanes, G. Salgueiro, P. Grossetete, R. Barton, and J. Henry, *IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things*, ser. Fundamentals. Pearson Education, 2017.

[20] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for internet of things realization," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2961–2991, 2018.

[21] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, "Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 84–91, Oct 2016.

[22] Q. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W. Hwang, and Z. Ding, "A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE Access*, vol. 8, pp. 116 974–117 017, 2020.

[23] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, "Mobile-edge computing introductory technical white paper," ETSI, Tech. Rep., Sep. 2014. [Online]. Available: https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf

[24] K. F. Kocan, K. W. McKiou, N. R. Pandya, J. Hernandez-Herrero, and F. T. Freuler, "The SCF: A framework for application enablement service composition," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 171–193, 2013.

[25] S. Kekki, W. Featherstone, Y. Fang, P. Kuure, A. Li, A. Ranjan, D. Purkayastha, F. Jiangping, D. Frydman, G. Verin, K.-W. Wen, K. Kim, R. Arora, A. Odgers, L. M. Contreras, and S. Scarpina, "MEC in 5G networks," ETSI, Tech. Rep., Jun. 2018. [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf

[26] J. Martin, J. Burbank, W. Kasch, and P. D. L. Mills, "Network Time Protocol Version 4: Protocol and Algorithms Specification," RFC 5905, Jun. 2010. [Online]. Available: https://rfc-editor.org/rfc/rfc5905.txt

[27] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile Edge Computing. A key technology towards 5G," ETSI White Paper No. 11, ETSI, Tech. Rep., Sep. 2015. [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf

[28] Internet Traffic Report, Website. [Online]. Available: http://www.internettrafficreport.com/