An Overview of Mobile Edge Computing: Architecture, Technology and Direction

Arslan Rasheed[1], Peter Han Joo Chong[1*], Ivan Wang-Hei Ho[2], Xue Jun Li[1], and William Liu[3]

[1]Department of Electrical and Electronic Engineering, Auckland University of Technology, New Zealand
[2]Department of Electrical and Information Engineering, The Hong Kong Polytechnic University, Hong Kong
[3]Department of Information Technology and Software Engineering, Auckland University of Technology, New Zealand
[email: {[1]peter.chong, arslan.rasheed, xuejin.li and [3]william.liu}@aut.ac.nz & [2]ivanwh.ho@polyu.edu.hk]
*Corresponding author: Peter Han Joo Chong

Abstract

Modern applications such as augmented reality, connected vehicles, video streaming and gaming have stringent requirements on latency, bandwidth and computation resources. The explosion in data generation by mobile devices has further exacerbated the situation. Mobile Edge Computing (MEC) is a recent addition to the edge computing paradigm that amalgamates the cloud computing capabilities with cellular communications. The concept of MEC is to relocate the cloud capabilities to the edge of the network for yielding ultra-low latency, high computation, high bandwidth, low burden on the core network, enhanced quality of experience (QoE), and efficient resource utilization. In this paper, we provide a comprehensive overview on different traits of MEC including its use cases, architecture, computation offloading, security, economic aspects, research challenges, and potential future directions.

1.   Introduction

With the increased usage of handheld devices, an explosion in data generation has been witnessed during the last decade. According to Cisco [1], the mobile data being generated is 4000 times more than the last decade and it is estimated that it will increase to 49 exabytes monthly by 2021. It is estimated that about 63 percent of the data will be offloaded to the core network, thus posing great challenges to network operators. This huge amount of data not only requires high bandwidth but also involves complex computation resulting in excessive delay. Hence, it is recommended to perform computation at the edge of the network to mitigate the burden on the core, enhance bandwidth, reduce latency, and improve quality of experience (QoE). For this purpose, edge computing is proposed to bring the computation near to the edge of the network, closer to users [2]. The most popular paradigm of edge computing is fog computing, which introduces an abstraction layer of resources to integrate the computation resources of participating devices [3]. Another concept of edge computing is cloudlet，in which micro datacentres are deployed at the edge of the network to provide efficient computation offloading，hence reducing the computation load of the core network as well as the communication latency [4].

Although fog and cloudlet computing paradigms are efficient in high-rate data dissemination and bringing the cloud capabilities at the edge of the network, there is still a lack of coverage and proximity, and they cannot cope with the context-awareness and latency requirements for some applications. The main drawback of cloudlet is its short range as it is accessed by Wi-Fi only, while the concept of fog computing should support multiple access technologies. In addition, the coordination among different devices and the number of hops between customers and the computing machine could further degrade the services. To overcome these challenges, European Telecommunications Standards Institute (ETSI) proposed Mobile Edge Computing (MEC), which features a new paradigm of edge computing that combines the wireless communication standard with the IT infrastructure [5]. MEC is a potential key player in 5G framework that brings cloud computing capability to the edge of cellular network, hence resulting in ultra-low latency, close proximity, context-awareness, and high throughput.

MEC can also provide cloud capabilities at the edge of the network to leverage Network Function Virtualization (NFV) so as to support the virtualization of physical resources for efficient management. Many operators across the world have been putting their efforts towards the realization of MEC and building use case scenarios for MEC. These include the Industry Specification Group (ISG) of ETSI, Vodafone, DOCOMO Japan, Telecom Italia, etc. Besides telecom operators, numerous manufacturers such as IBM, Huawei, and Intel are also contributing towards the seamless integration of cloud computing with cellular networks and standardization of MEC.

A tremendous contribution from the research community has been witnessed in many surveys on MEC, for example, Mach and Becvar [6] discussed the use case scenarios and corresponding computation offloading techniques. However, they did not discuss security concerns and the economic perspective. In [7], the authors have discussed the security issues and future trends, but use cases and applications are not covered. Another survey presented in [8] discussed the security threats without much emphasis on the architecture of MEC.

The research on MEC is amateur yet, and efforts are being put together by academia and industry to enlighten the prospect of MEC. In this paper, we aim to provide a comprehensive survey of the state-of-the-art research in different traits of MEC, including its use cases and applications, architecture, use cases, security, computation offloading, economic and future perspectives.

The rest of the paper is organized as follows. Section 2 describes the architecture of MEC with use cases. Section 3 overviews the computation offloading techniques. Section 4 discusses the security perspectives. Section 5 covers economic models. Section 6 points out the potential research directions following the conclusion in section 7.

2.   Architecture and ETSI Standardization

This section revisits the technical details of different proposed solutions for integrating cloud computing with cellular networks. Furthermore, the architecture proposed by ETSI is briefly explained. We emphasize on the architectural aspects, latency aspects, and integration of cloud capabilities with cellular network. A detailed pictorial representation of different architectures is provided in **Fig. 1**. Different proposed solutions are as follows:
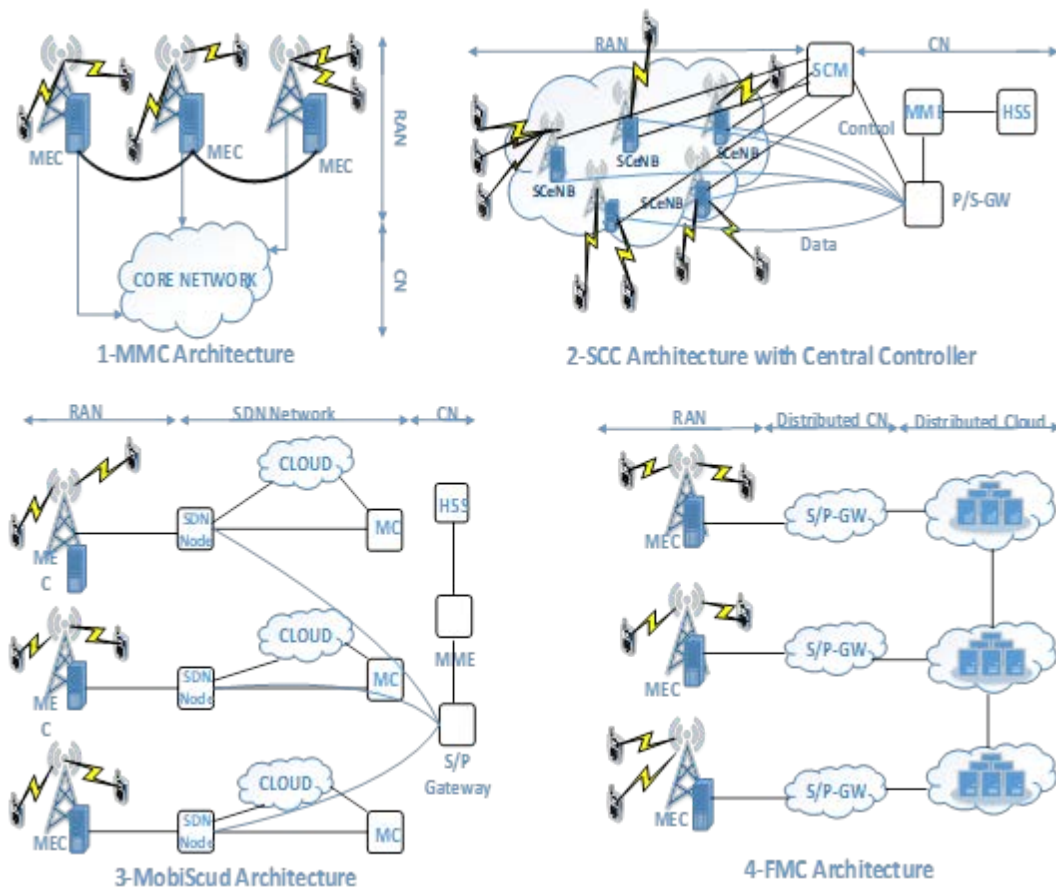
Fig. 1. Comparison of different MEC Architectures.

2.1 Mobile Micro Cell (MMC)

One of the major challenges for contemporary technologies is the stringent latency required for different applications such as Augmented Reality (AR), video streaming, context-awareness, gaming, etc. To provide services with low latency, the first architecture proposed in [9] reduces the access time to resources throughout the network via deploying a server at the base station (BS). The users are connected to the server and hence can access resources with ultra-low latency. To connect the BS with the core network, a backhaul link of high capacity is used while node-to-node communication is also possible to provide mobility management. This simple architecture involves no complexity, and provides ultra-low latency that can meet the requirements of most of the modern applications. The proposed architecture introduces no central entity for controller but instead provides a distributed control in the network. The benefit of MMC lies in the simple architectural design with no overhead for service providers to change the core of the network. However, MMC faces challenges such as single point of failure, low capacity and lack of load balancing. The process of VM migration can be severely affected in case of server outage.

2.2 Small Cell Cloud (SCC)

A European project called TROPIC proposed an architecture that combines cloud computing with mobile network to provide high storage capacity, low latency, and high computation of small cells [10]. The NFV is a process of virtualizing the network services traditionally running on the dedicated hardware such as switches, routers and firewalls [11]. The proposed solution forms a pool of micro, pico, and femtocells' resources with the aid of NFV [12]. The proposed solution forms a pool of micro, pico, and femtocells' resources with the aid of NFV. The proposed architecture is particularly suitable for services with stringent latency requirements. In near future, it is estimated that a large network of small cells will be deployed. As the pool gathers resources from multiple cells, coordination among cells is achieved by a manager called Small Cell Manager (SCM). Other than the coordination among SCeNBs, the SCM also performs virtualization and resource management. Furthermore, SCM integrates the mobile network with IT infrastructure i.e. computation resources. The proposed architecture provides high computational capabilities at the edge of the network, however, mechanism for coordinating eNBs and optimal placement of SMC is yet to be investigated.

Another challenge faced, in case of femtocell particularly, is the availability of dynamic resources as femtocells can be switched off/on. Although SCM provides dynamic resources management however elastic resources make it difficult for SCM to take care of both cloud and mobile network resources.

2.3 MobiScud

Software Defined Network (SDN) brings programmability, ease of management and flexibility into the network by decoupling the control plane from the data plane [13]. To exploit the functionalities of SDN along with NFV in cellular networks, an architecture called MobiScud is proposed in [14]. The integration of cloud services with the mobile network creates opportunity for UEs to access the computation resources with ultra-low latency. It is a distributed architecture, in which resources are not located within the BS but distributed throughout the Radio Access Network (RAN). An entity termed as MobiScud Manager (MC) is responsible for managing the resources, interacting with SDN network and coordinating among different entities. The MC also supports the routing of data among SDN switches and

computation offloading. The backward compatibility with the mobile network is also guaranteed in the proposed architecture for smooth operation. Although the proposed architecture brings the benefits of SDN into mobile network, the selection of optimum node for computation offloading is still an outstanding issue.

2.4 Follow Me Cloud (FMC)

The idea of FMC is same as that of MobiScud, in which the resources are distributed throughout the network rather than a collocated server. The cloud services are running at distributed data centres that follow the trajectory of user equipment (UE). This is the reason for its unique name. Likewise, the core of the network is also distributed and architecture is similar to MobiScud [15]. The proposed architecture is dynamic in a sense that it supports the increasing number of UEs which is very lucrative for mobile operators. The main difference between MobiScud and FMC is the exploitation of SDN in MobiScud [16]. Due to the distributed nature of the network, a controller called FMC Controller (FMCC) is introduced that manages resources and makes decisions regarding the optimal datacentre for computation offloading. However, the distributed DCs can manage themselves in self-organizing way without the help of FMCC which is a strong aspect of FMC. The FMCC can either be placed hierarchically, distributed, or locally. FMC is an extremely distributed architecture that provides computation at the edge of the network resulting in ultra-low latency and high bandwidth, however, the management of such a distributed architecture is a challenging task. A comprehensive comparison table is provided in **Table 1** below.

Table 1. Comparison table for MEC proposed architectures.

Characteristics	**MMC**	**SCC**	**MobiScud**	**FMC**
Hops	1	Variable	Variable	1
Architecture type	Central MEC	Distributed	Distributed	Dynamic
Latency	Ultra-low	High	Low	Ultra-low
Bandwidth	Low	High	High	High
Node-Node Communication	Yes	Yes	Yes	No
Resources Availability	Readily	Outages may occur	Vary	Readily
Complexity	Low	High	Moderate	High
Capacity	Lower	Higher	Moderate	Moderate
Load Balancing	No	Yes	Yes	Yes

2.5 ETSI Reference Architecture

For the introduction and standardization of MEC, a key role is played by ISG ETSI. We briefly discuss here the efforts made by ISG to integrate the MEC concept with 5G. Although a comprehensive architecture of MEC is yet to be finalized but the conceptual draft of MEC is presented in [17]. The proposed draft mentions all the related terminologies so that they could be referenced in future proposed models. In addition, a detailed framework for proof of concept (PoC) of MEC is also presented in [17], which includes a reference architecture with the details of different fundamental elements, reference points and their interactions.

The reference architecture consists of three layers - the mobile edge system level, the mobile edge host level, and the core network. It consists of a virtualization infrastructure, in which multiple functional and management blocks represent physical resources that perform different tasks. From the reference architecture as shown in **Fig. 2**, MEC resides in between the users and the core network. The mobile edge host level consists of mobile edge applications, the mobile edge platform, virtualization infrastructure, and management plane.
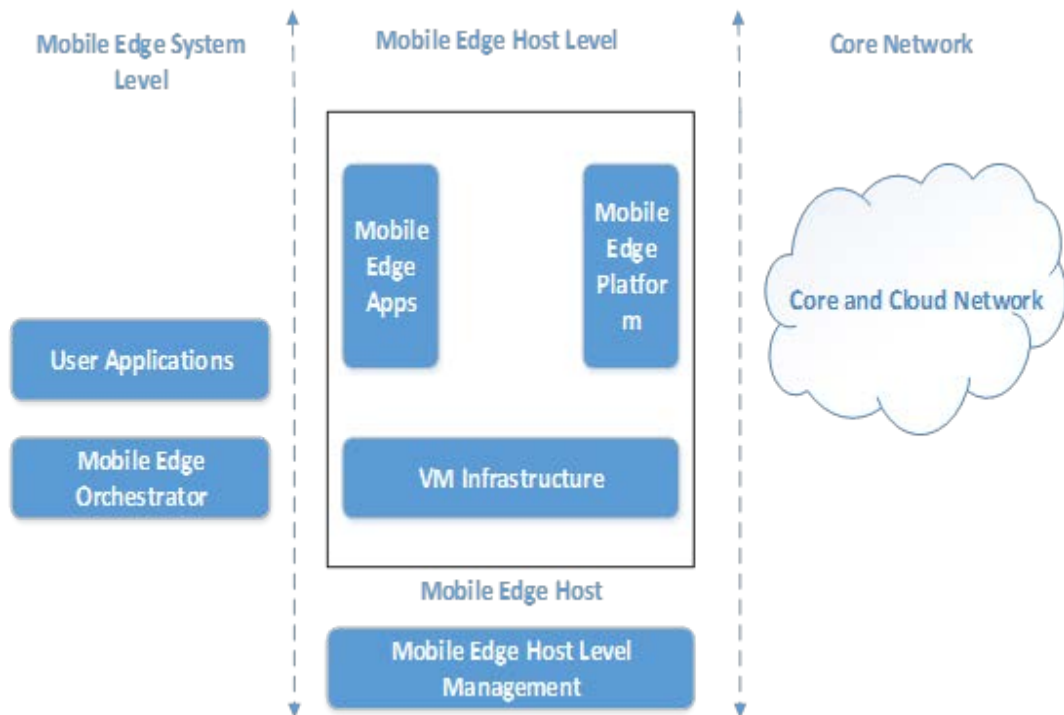
Fig. 2. MEC Architecture proposed by ISG ETSI.

The mobile edge applications are responsible for running VMs to perform different tasks (such as radio network optimization, location information, computation offloading, job execution, and bandwidth management) with the assistance of the virtualization infrastructure. The mobile edge platform is responsible for interacting with mobile edge applications to provide services such as Domain Name System (DNS), security, analysing service requirements and operational rules. The mobile edge system level consists of a mobile edge orchestrator and UE applications.

The mobile edge orchestrator plays a key role in assuring coordination among the UEs and network, authentication, accounting, maintaining record of available resources, establishing connection with virtualization infrastructure, and application management. The UE applications are simple applications that run on host devices. MEC enjoys its close proximity to the end user, thus results in ultra-low latency, high bandwidth, interactive communications, as well as context-aware computation. The **Fig. 3** shows the application and potential use case scenarios of MEC. MEC has the potential to offer efficient services to all the stakeholders. In

this section, we divide the use cases of MEC into the user's and the operator's points of view.

Fig. 3. MEC Use Cases and Applications.

- **Operator's-End Use Cases**

To relieve the backhaul and lessen the burden of the core network, MEC can cache high demand applications such as viral videos and geographical information. Sometimes it is not efficient to send all the information to the core such as sensing data, monitoring information, etc. Hence, MEC server can apply data analytics and aggregation to save the bandwidth. Video streaming requires high bandwidth and low latency. The traditional method of using TCP over HTTP is inadequate due to the high latency of the handshake and acknowledgement process.

Hence, video analytics can be employed in MEC to avoid such a delay. As MEC can monitor the traffic at the RAN level, integration of resources can be carried out more efficiently to achieve optimized utilization, which will significantly enhance the performance of the network [5].

- **Customer's-End Use Cases**

One of the major hindrances of auto-pilot vehicles is latency involved in communications. MEC provides ultra-low latency which fulfils the requirements of connected vehicles. In [9], the authors have investigated different models to facilitate the design and optimization of Vehicular Ad-hoc Network (VANET). Modern applications such as augmented reality gaming and interactive communication requires high computation and ultra-low latency, which could not be satisfied by existing networks [18]. MEC makes it possible by offloading the computation at MEC server. As MEC involves low latency and high computation power, so it could be used for various feature recognition tasks, e.g., face recognition. Some scenarios such

as concerts or local events require video broadcasting in the local vicinity, which do not require data transmission to the backbone. In traditional mechanism, the video is sent to the core and every video request from the user at the edge of the system poses latency and wastes the resources of the core network. To save the resources and reduce latency, video orchestration can be done at MEC server [5].

3.   Computation Offloading

Mobile devices have limited battery and computing resources, so the computation task can be migrated to the network end. This migration is called computation offloading, which involves the shifting of resource-hungry applications such as 3D gaming, video editing, vehicular network infotainment services, and AR to the network. The above mentioned applications have stringent latency requirements, hence offloading at the MEC server is more suitable as it is closer to the UE. The concept of computation offloading looks fancy, but it is very complicated and can be primarily divided into three types; complete offloading, partial offloading, and local computation. The complete and no offloading techniques are straightforward, however partial offloading involves a complex decision depending on the quality of the backhaul link and its capacity, UE capability, battery status, availability of resources, and UE preferences.
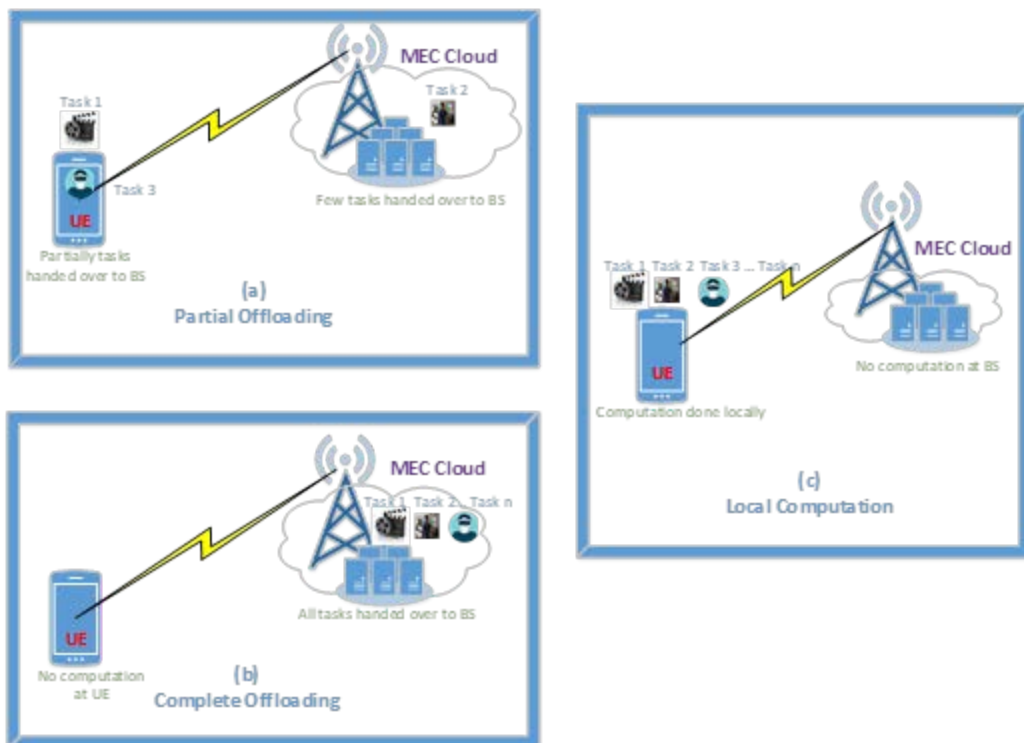
Fig. 4. The Concept of Computation Offloading.

The full computation offloading process involves the handover of all computation to the MEC server, while local computation is carried out at the end users. The partial offloading technique

transfers the computation tasks partially to the network. A conceptual representation of computational offloding is shown in **Fig. 4** where (a) represents the partial offloading, (b) represents complete offloading and (c) represents local computation. In the first section, few tasks are being computed at the BS while rest computation is being carried out at the UE. Similalrly, the second section represents the complete offloading in which all the tasks are being performed at BS. The BS is responsible for all computation and UE is saving its battery. The last section represents a case in which all the tasks are carried out by the UE and BS does not perform computation.

A Successive Complex Approximation (SCA) based algorithm to consider the densely deployed MIMO multi-cell communication system is proposed in [19]. The results show that the proposed algorithm performs efficiently in highly dense environment when it achieves local maxima. The proposed algorithm is more suitable for saving energy for tasks with high computation cost.

The authors in [20] propose vehicular network for computation offloading under stringent latency constraints and proposed a contract based solution. The MEC operator charges the vehicles on "pay-as-you-go" basis. As the number of offloading tasks increases, the cost also increases. The information about charges and computation offloading tasks are broadcasted to the road segments via MEC. Another technique involving game theory is presented in [21], which outperforms the contract-based and web-cached acceleration offloading techniques. The main objective of game theory is to let the users make optimum decisions based on strategic interaction for offloading the task to the server. The strategic interaction makes sure that users do not request for offloading simultaneously so as to avoid interference. It represents the offloading task as a NP-hard problem and achieves Nash equilibrium in case of multiple users with high computation requests. The game theory not only performs efficient offloading but also improves the performance of the network.

There are two main challenges in computation offloading, i.e., computation delay and energy consumption. These two constraints play vital role in the decision whether to offload or not. To make optimal decisions, a one-dimensional search algorithm is proposed in [22]. It takes into account the capabilities of MEC server and UE, link quality, and buffer size. The authors recommend making decision at the UE end on each slot whether to offload or compute locally. The results show that the proposed algorithm outperforms the local computation and full offloading policy. However, UE will continuously receive feedback from the server which involves a large amount of signalling overhead.

To avoid the server failure in case of cloud computing, Berihun F., et al propose dynamic heuristic algorithm using Naive Bayee's classifier combining real-time monitoring data with the historical data [23]. The authors takes into account multiple factors such as CPU load, RAM utilization, data I/O rate and number of users. The idea is to create a hypervisor for each physical server running multiple instances and numerous hypervisors are gathered to form a pool connecting physical servers. As the physical server gets overloaded, the services of the failed server are migrated to other hypervisor. The status of each hypervisor is compared with the previous failure history to estimate the next expected failure time. For the proposed model, the raw data of the status collected through log files is grouped together to create training data for prediction. The status data of each group according to the range of the time intervals are averaged to give more generalized data and are grouped into time sequential order. The prediction based on Naïve Bayee's algorithm uses probabilistic approach to model the

real-time server status against the training data in order to find out the group number. Once the group number is predicted, it is possible know the failure time. This approach ensures a stable cluster as migration of services is carried out within the cluster. The main challenge that needs to be addressed is the false alarm could be triggered.

Another clustering scheme to avoid the MEC failure is proposed in [24]. The authors propose a clustering scheme in which multiple MECs are grouped together in clusters based on the geographic location. As an MEC gets overloaded, it intimates the concerned CH to find an appropriate MEC for offloading. The authors discuss two scenarios; one in which a neighbouring MEC is available in the same cluster and the other case is the situation when there is no MEC within the cluster/communication range to offload the computation. Second situation is more complex and requires special consideration. The authors propose to exploit the neighbouring MEC's UEs as relay nodes (RN) to reach the optimum MEC using Floyd-Warshall algorithm. The disconnected nodes are grouped together and the CH of these nodes find out if the neighbouring RNs are connected with any of the MEC. If found connected, Floyd-Warshall algorithm is used to find out the optimum path to the sever creating a multipath wireless environment.

4.   Security Issues

One of the key players to make a technology successful is security and privacy. To make MEC successful, privacy and network security should be analysed critically. From the security point of view, MEC security can be divided into three main parts; edge security, core network security and virtualization infrastructure security.

4.1 Edge Network Security

The leap of MEC success lies in its server placement at the edge of the network, however, this could be disadvantageous in terms of security threat. Due to closed proximity to the end users, the MEC is more vulnerable to malicious attacks such as DOS, hijacking, eavesdropping, man-in-middle and etc. It becomes easy to jam the server at the edge of the network by sending too many requests. To address this issue, a round-trip time based technique that detects rogue connection is proposed in [26]. The delay between the user and DNS server is measured to find out if the user is connected with the authorised server.

4.2 Core Network Security

The privacy of core is traditionally maintained by third party such as Microsoft, Google and etc. which also includes user's information. The basic concern of customers is sensitive data that cannot be compromised. In the case of MEC, edge devices can share information directly bypassing the core which can be a serious threat. These threats can lead to big disaster if false information are carried to the core network. The effects of such threats can lead to the alteration of network data flow, denial of service and loose access to the network.

4.3 MEC virtualization Security

A virtualized environment is more sensitive to security because if one of the entities is compromised, it can affect the whole virtualized environment. This can ultimately result into the drain of resources, denial of requests for users and even stolen of user's personal data.

Virtualization is one of the key enabler for edge paradigm thus it is necessary to secure it on priority. For this purpose, different policies are introduced in [27] such as network isolation, hypervisor hardening and network abstractions.

5.   Economic Perspective

To make a technology successful, its economic effects need to be critically analysed before launching. In this section, we analyse different economic aspects of MEC and investigate how MEC can play a role in making 5G more profitable both for the operators and users. The main challenge for MEC, like many other technologies, is cost reduction and larger revenue. There are many factors associated with this challenge, from user's perspective, such as good QoE, low latency, Value Added Services (VAS), low rates, data integrity, new applications, and security. Furthermore, the integration of new technologies, e.g., IoT, smart city, VANET, and smart home, with existing cellular network has put a lot of burden on the backbone network and can adversely affect the QoE. Moreover, the operators are in competition for launching new applications and services for commercial, industrial, and educational purposes.

Besides the challenges, MEC seeks its potential economic perspectives in many applications such as connected vehicles, gaming, Wireless Sensor and Actuator Networks (WSAN), health care, big data, web caching and virtual reality. The main advantage that MEC enjoys is its close proximity that leads to ultra-low latency, high bandwidth, and computational resources to improve the QoE. MEC can generate higher revenue by exploiting its peculiar nature and rich characteristics. From the perspective of the end users, higher data rates, rich applications and enhanced QoE leads to the satisfaction and loyalty. From the operator's perspective, traffic monitoring and optimization at the edge of the network can reduce the operational cost on a large scale. The self-organizing feature as described in case of FMC will reduce the cost of extra equipment. Similarly, in case of MobiScud where SDN is employed to introduce programmability, flexibility and ease of management into the network. These up-to-mark technologies can further benefit the MEC in terms of bringing innovation into the network. Hence, it is concluded that MEC can be a key enabler for 5G as it seeks many business opportunities in different fields.

6.   Potential Research Directions

As MEC is a new concept, it requires serious efforts to become mature and standardized. There are number of open research areas within MEC and potential challenges. In this section, we analyse the potential directions that could be beneficial for researchers to standardize and validate the concept of MEC.

6.1 Economic Models

There is a need of extensive research on the economic models for MEC and it is missing in the literature so far to the best of our knowledge. Hence, a large space of research is available to investigate the economic perspectives of MEC. In particular, the decision of pricing is difficult, for example, in case of computational offloaded applications where it is hard to decide whether to be paid by customers or the application owner. MEC has lot to offer in terms of financial gain [5] and hence serious studies are needed to be carried out to explore the financial potential. Furthermore, as MEC is a strong candidate for 5G, hence its economic prospects

should be critically analysed. In this regard, suitable business models and policies could be developed.

6.2 Integration of SDN

SDN is a new paradigm that brings flexibility, programmability, and scalability into the network by separating data plane from control plane. SDN has proven its applicability into various scenarios such as vehicular networks, datacentres, and mobile networks. Now it is worth investigating that how SDN could be integrated seamlessly with MEC as it has a lot to offer in terms of enhanced performance and improved QoS. SDN may have huge impact on MEC for improving its intractability, performance, scalability and injection of new applications. Few attempts are already made such as MobiScud to integrate the SDN with cloud and mobile network [15]. There is a big opportunity available for researchers to investigate the insights of integration of SDN with MEC. Furthermore, it could also be investigated how SDN be beneficial in computational offloading, the key enabler of MEC.

6.3 Computation Offloading

One of the hottest topics for researchers in wireless communications, nowadays, is computation offloading. The main objective of computation offloading is to save the battery of handheld devices under the strict constraints of latency. So far, there are multiple challenges that need to be addressed such as realistic assumptions, slicing size, and mobility of users. These challenges are needed to be investigated with real environment and experimental testbeds so that the technology can get matured [19]. The decision of computation offloading plays a vital role as it decides whether to offload completely or partially. Current research totally ignores the fact that uplink of data to the MEC requires more bandwidth as large sized files are uploaded. On the other hands, downlink data is not only smaller in quantity but also refined as only results are sent to the customer. Hence, it opens new research horizons for the researchers to reshape the concepts of research in uplink and downlink traffic.

6.4 Connected Vehicles

Due to the close proximity, MEC can play vital role in the dissemination of emergency messages in vehicular networks with low latency. Due to its location in the network with high computational resources, it is foreseen that vehicular networks could be benefited from MEC and fulfil the requirements of context-awareness, high bandwidth, and ultra-low latency for various vehicular infotainment services. A possible future direction is MEC-enabled vehicular software defined networks [28]. The MEC technology has strong potential to satisfy a vehicular network's requirements. The vehicular networks foresee the MEC as key technology to fulfil the requirements of ITS because of its closed proximity to the end users, context-awareness and distributed architecture. The main issue of VANET is stringent requirements on latency. Another challenge of VANET is dynamic change in topology due to high mobility of vehicles. It requires to have global as well as local view of the network so that safety messages and infotainment services could be provided [29]. The limited bandwidth of DSRC channel of 10MHz is insufficient for futuristic application which cause the burden if accessed through 5G network. Hence, an optimized and load sharing architecture is needed to be proposed which meets the requirements of both the networks such as vehicular cloud computing [30]. Many other research areas are still needed to be investigated in context of connected vehicles such as vehicle to vehicle video streaming with the help of MEC and cloud computing.

6.5 Security

One of the major concern of modern technologies is security and privacy. The success of any service lies in the privacy of its customers and it is the primary concern of any operator to protect its customers' privacy. As MEC is more open than traditional mobile networks allowing different vendors, servers and third party applications hence more security threats are associated with it [27]. Furthermore, the computation offloading is performed over wireless networks which is more vulnerable to the security of networks. Hence, application-specific encryption techniques are required to be implemented so that security could be enhanced however it comes at the cost of application low performance. So, encryption technique is required that has balanced performance and overhead.

6.6 Concept Corroboration

The work done so far is validated either through simulations or mathematical models. Basically, the work done is based on simple assumptions which is though good at the initial stage however for mature research real-time assumptions are needed to be considered. For realization of MEC, these concepts need to be validated under hardware constraints or at least real assumptions. Furthermore, hard trials of the proposed architectures needed to be carried out to materialize the concept of MEC.

7.   Conclusion

MEC is a promising technology that brings cloud capabilities to the edge of the cellular network so as to improve numerous performance metrics such as latency, bandwidth, QoS, and optimized wireless resources. MEC is a tremendous amalgamation of IT infrastructure and cellular network that brings computation capabilities into the cellular network. The MEC has strong potential to provide futuristic applications, which require high computation and low latency such as video streaming, gaming, AR, connected vehicles, and e-health. However, the MEC concept is not mature for the time being and requires further validations with real-time considerations in practical scenarios.

References

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," USA 2017.

[2] Al Agha, K., Pujolle, G., Ali‐Yahiya, T., "Mobile‐Edge Computing," *Mobile and Wireless Networks*, vol. 2, pp. 283-306, Sep. 2016. Article (CrossRef Link)

[3] Madsen, H., Burtschy, B., Albeanu, G., et.al., "Reliability in the utility computing era: Towards reliable fog computing," in *Proc. of 20th IEEE International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 43-46, 2013. Article (CrossRef Link)

[4] Verbelen, T., Simoens, P., De Turck, F., and Dhoedt, B., "Cloudlets: Bringing the cloud to the mobile user," in *Proc. of the third ACM workshop on Mobile cloud computing and services*, pp. 29-36, 2012. Article (CrossRef Link)

[5] Hu, Y. C., Patel, M., Sabella, D., Sprecher, N., and Young, V., "Mobile edge computing—A key technology towards 5G," *ETSI*, Sophia Antipolis, France, White Paper, vol. 11, pp. 1-16, 2015.

[6] Mach, P., and Becvar, Z., "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628-1656, Mar 2017. Article (CrossRef Link)

[7]   Abbas, N., Zhang, Y., Taherkordi, A., and Skeie, T., "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450-465, 2018. Article (CrossRef Link)

[8]   Roman, R., Lopez, J., and Mambo, M., et al., "Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, vol. 78, pp. 680-698, 2018. Article (CrossRef Link)

[9]   Xie, Y., Ho, I. W-H., and Magsino, E., "The Modeling and Cross-layer Optimization of 802.11p VANET Unicast," *IEEE Access*, vol. 6, pp. 171-186, 2018. Article (CrossRef Link)

[10] Wang, S., Tu, G.-H., Ganti, R., He, T., Leung, K., Tripp, H., et al., "Mobile micro-cloud: Application classification, mapping, and deployment," in *Proc. of Annual Fall Meeting ITA*, pp. 1-7, Oct. 2013.

[11] Chiosi, M., et al., "Network functions virtualization: An introduction, benefits, enablers, challenges & call for action," in *Proc. of SDN and OpenFlow SDN and OpenFlow World Congress*, 2012.

[12] ETSI, "Network function virtualization (NFV): Architectural framework," *ETSI Gs NFV*, vol. 2, p. V1, 2013.

[13] Manzalini, A., "Towards 5G software-defined ecosystem: Technical challenges, business sustainability and policy issues," *IEEE SDN White Paper*, 2016.

[14] Ku, I., Lu, Y., and Gerla, M., "Software-defined mobile cloud: Architecture, services and use cases," in *Proc. of International Wireless Communications and Mobile Computing Conference*, pp. 1-6, 2014. Article (CrossRef Link)

[15] Wang, K., Shen, M., Cho, J., Banerjee, A., Van der Merwe, J., and Webb, K., "MobiScud: A fast moving personal cloud in the mobile network," in *Proc. of Workshop All Things Cellular Oper. Appl. Challenge*, pp. 19-24, 2015. Article (CrossRef Link)

[16] Taleb, T., and Ksentini, A., "Follow me cloud: Interworking federated clouds and distributed mobile networks," *IEEE Network*, vol. 27, no. 5, pp. 12–19, Oct. 2013. Article (CrossRef Link)

[17] ETSI: Mobile edge computing (MEC): "Terminology, Proof of concept Framework, Framework and Reference Architecture," V1.1.1, Mar. 2016.

[18] Yuen, S., Yaoyuneyong, G., and Johnson, E., "Augmented reality: An overview and five directions for ar in education," *Journal of Educational Technology Development and Exchange*, vol. 4, no. 1, pp. 119-140, 2011. Article (CrossRef Link)

[19] Sardellitti, S., Scutari, G., and S. Barbarossa, S., "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 89-103, June 2015. Article (CrossRef Link)

[20] Takahashi, N., Tanaka, H., and Kawamura, R., "Analysis of process assignment in multi-tier mobile cloud computing and application to edge accelerated web browsing," in *Proc. of IEEE 3rd International Conference on Mobile Cloud Computing Services and Engineering (MobileCloud)*, pp. 233–234, 2015. Article (CrossRef Link)

[21] Zhang, K., Mao, Y., Leng, S., Vinel, A., and Zhang Y., "Delay constrained offloading for mobile edge computing in cloud-enabled vehicular networks," in *Proc. of 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, pp. 288–294, 2016. Article (CrossRef Link)

[22] Chen, X., Jiao, L., Li, W., and Fu, X., "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Transactions on Networking*, vol. 24. no. 5. pp. 2795–2808, Oct 2016. Article (CrossRef Link)

[23] Fekade, B., Maksymyuk T., and Jo, M., "Clustering hypervisors to minimize failures in mobile cloud computing," *Wireless Communications and Mobile Computing*, vol 16, no. 18, pp. 3455-3465, Dec 2016. Article (CrossRef Link)

[24] Satria, D., Park D., and Jo, M., "Recovery for overloaded mobile edge computing," *Future Generation Computer Systems*, vol. 70, pp. 138-147, May 2017. Article (CrossRef Link)

[25] Liu, J., Mao, Y., Zhang, J., and Letaief, K. B., "Delay-optimal computation task scheduling for mobile-edge computing systems," in *Proc. of IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1451–1455, 2016. Article (CrossRef Link)

[26] Han, H., Sheng, B., Tan, C. C., Li Q., and Lu, S., "A timing-based scheme for rogue ap detection," *IEEE Transaction on Parallel and Distributed Systems*, vol. 22, no. 11, pp. 1912-1925, Nov 2011. Article (CrossRef Link)

[27] Pek, G., Buttyan, L., and Bencsath, B., "A survey of security issues in hardware virtualization," *ACM Computing Surveys*, vol. 45, no. 3, p. 40, 2013. Article (CrossRef Link)

[28] Liu, J., Wan, J., Jia, D., Zeng, B., Li, D., Hsu, C.H., and Chen, H, "High-efficiency urban traffic management in context-aware computing and 5G communication," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 34-40, 2017. Article (CrossRef Link)

[29] Zhang K., Mao, Y., Leng, S., Zhao, Q., Li, L., Peng, X., Pan, L., Maharjan, S., and Zhang, Y, "Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896-5907, 2016. Article (CrossRef Link)

[30] Gerla, M., "Vehicular Cloud Computing," in *Proc. of Vehicular Communications and Applications Workshop*, pp. 152-155, 2012. Article (CrossRef Link)

Arslan Rasheed received his B.Sc.(Hons) from Bahauddin Zakariya University, Pakistan and M.Sc. with distinction from Comsats IIT, Islamabad in 2011, and 2015 respectively. He has 5 years of professional and teaching experience. He is currently pursuing his Ph.D from the department of Electrical and Electronic Engineering, Auckland University of Technology (AUT), New Zealand. His research interests include wireless communication networks, Vehicular Communication, cloud computing, cognitive radio, routing and so on. He has published some technical papers in these areas.

Peter Han Joo Chong is currently a Professor and Head of Department of Electrical and Electronic Engineering at Auckland University of Technology, Auckland, New Zealand. He received the Ph.D. degree from the University of British Columbia, Canada, in 2000. He is an Adjunct Professor at the Department of Information Engineering, Chinese University of Hong Kong, Hong Kong. He was previously an Associate Professor (tenured) in the School of Electrical and Electronic Engineering at Nanyang Technological University (NTU), Singapore. Between 2011 and 2013, he was an Assistant Head of Division of Communication Engineering. Between 2013 and 2016, he was a Director of Infinitus, Centre for Infocomm Technology. From February 2001 to May 2002, he was a Research Engineer at Nokia Research Center, Helsinki, Finland. Between July 2000 and January 2001, he worked in the Advanced Networks Division at Agilent Technologies Canada Inc., Vancouver, BC, Canada. His research interests are in the areas of mobile communications systems including MANETs/VANETs, multihop cellular networks and Internet of Things/Vehicles.

Ivan Wang-Hei Ho is currently a Research Assistant Professor at the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University. He received the B.Eng. and M.Phil. degrees in Information Engineering from The Chinese University of Hong Kong, and the Ph.D. degree in Electrical and Electronic Engineering from Imperial College London, UK. He worked on the MESSAGE project funded by EPSRC and Department for Transport, UK, and the ITA project funded by the US Army Research Laboratory and UK Ministry of Defence during his Ph.D. studies. In 2007, Ivan spent a summer working at the IBM T. J. Watson Research Center, Hawthorne, NY, USA. After his Ph.D. graduation, he was with the System Engineering Initiative at Imperial College as a Postdoctoral Research Associate. In Sept 2010, he co-founded P2 Mobile Technologies Limited at Hong Kong Science Park and served as the Chief R&D Engineer. He primarily invented the MeshRanger series wireless mesh embedded system, which won the Silver Award in Best Ubiquitous Networking at the Hong Kong ICT Awards 2012. Ivan is an Associate Editor for IEEE Access and IEEE Transactions on Circuits and Systems II, TPC Co-chair for the CoWPER Workshop in IEEE SECON 2018, and TPC member for IEEE conferences (ICC, WCNC, PIMRC, etc.). His research interests are in wireless communications and networking, specifically in vehicular ad-hoc networks (VANET), intelligent transportation systems (ITS), Internet of things (IoT), and mobile physical-layer network coding.

Xue Jun Li received his B.Eng (First Class Honors) and Ph.D in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2004 and 2008, respectively. Between November 2007 and July 2008, he worked as Research Engineer and then as Research Fellow in Network Technology Research Centre, NTU. Between August 2008 and September 2008, he worked as Research Scientist in Temasek Laboratories @ NTU. From September 2008 to May 2011, he was with School of Electrical and Electronic Engineering, NTU as a faculty member. Between June 2011 and January 2013, he worked as Research Scientist in Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. Since January 2013, he has been with Auckland University of Technology (AUT), where he is a Senior Lecturer. His research interests include design/analysis of wireless networking protocols, modelling/design of radio frequency integrate circuits, and system optimizations.

William Liu is currently a Senior Lecturer in the Department of Information Technology and Software Engineering, School of Engineering, Computer and Mathematical Sciences at the Auckland University of Technology, New Zealand. He holds a Masters degree and a PhD degree in Electrical and Computer Engineering, both obtained at the University of Canterbury, New Zealand in 2005 and 2010 respectively. He had been working as a network planner and designer in Beijing Telecom in China for 5 years before he immigrated to New Zealand in 2002. He has co-authored more than 80 papers published in international journals and conferences, and he participates in the Program Committees of several premier IEEE conferences including INFOCOM, GLOBECOM, ICC, GreenCom, DRCN, RNDM and ATNAC. His main research interests are in the design and performance evaluation of the architecture and protocols for packet-oriented networks. He is working especially in the areas of network survivability, sustainability and trustworthy computing.