A Survey on Mobile Edge Computing

Arif Ahmed
Department of Computer Science & Engineering,
National Institute of Technology, Silchar,
Silchar, India
Email: arifch2009@gmail.com

Ejaz Ahmed
Department of Computer System and Technology,
University of Malaya,
Kuala Lumpur, Malaysia
Email: ejazahmed@ieee.org

***Abstract*—Mobile Edge Computing is an emerging technology that provides cloud and IT services within the close proximity of mobile subscribers. Traditional telecom network operators perform traffic control flow (forwarding and filtering of packets), but in Mobile Edge Computing, cloud servers are also deployed in each base station. Therefore, network operator has a great responsibility in serving mobile subscribers. Mobile Edge Computing platform reduces network latency by enabling computation and storage capacity at the edge network. It also enables application developers and content providers to serve context-aware services (such as collaborative computing) by using real time radio access network information. Mobile and Internet of Things devices perform computation offloading for compute intensive applications, such as image processing, mobile gaming, to leverage the Mobile Edge Computing services. In this paper, some of the promising real time Mobile Edge Computing application scenarios are discussed. Later on, a state-of-the-art research efforts on Mobile Edge Computing domain is presented. The paper also presents taxonomy of Mobile Edge Computing, describing key attributes. Finally, open research challenges in successful deployment of Mobile Edge Computing are identified and discussed.**

Keywords*—*Mobile Edge Computing, Computation Offloading, Collaborative Computing

I. Introduction

Recently, Mobile devices (such as Mobile Phone, Smart Phone, Tablet *etc.*) are emerging as an important tool for learning, entertainment, social networking, updating news and businesses [1]. However, due to resource constraints of mobile devices (processing power, battery lifetime, storage capacity) mobile users do not get the same satisfaction compare to desktop [2]. After the evolution of mobile cloud computing, many cloud computing services such as m-health-care[3], m-learning[4], m-gaming[5] and m-governance[6] are directly accessible from the mobile devices [7]. This also implicates the high network load, and growing demand of network bandwidth since data have to be transmitted and received to and from mobile devices and cloud data centers. It is estimated that demand of bandwidth is expected to be doubled each year[8].

Moreover, new Internet-of-Things (IoT) paradigm enabled the resource-constrained devices to be interconnected through Internet [9]. But, many of these edge devices are embedded with low processor and storage capacity. To overcome above scenario in Mobile Cloud Computing (same scenario as IoT paradigm), many techniques such as cyber foraging [10],[11] or computational offloading[12],[13] have been proposed where edge device offloads some computation to the remote resourceful cloud(Amazon EC2[14] and Windows Azure[15]), thereby saving processing power and energy. However, offloading computation to the public cloud may involve long latency for data exchange between the public clouds and edge device through the Internet. To overcome above problem, cloudlet based offloading is proposed where mobile devices offload computational to the less resourceful server near the user proximity accessible using Wi-Fi access point[10].

However, Cloudlet is less effective in compare to cloud computing characteristic because of the following reason: First, Cloudlet can be accessed only through Wi-Fi access point, which covers only small region, therefore, there is no commitment of ubiquitous computing. Second, Cloudlet is less resourceful in compare to cloud, so, it is not scalable in service and resource provisioning. To overcome above challenging issues, a new paradigm call Mobile Edge Computing is proposed recently [16],[17].

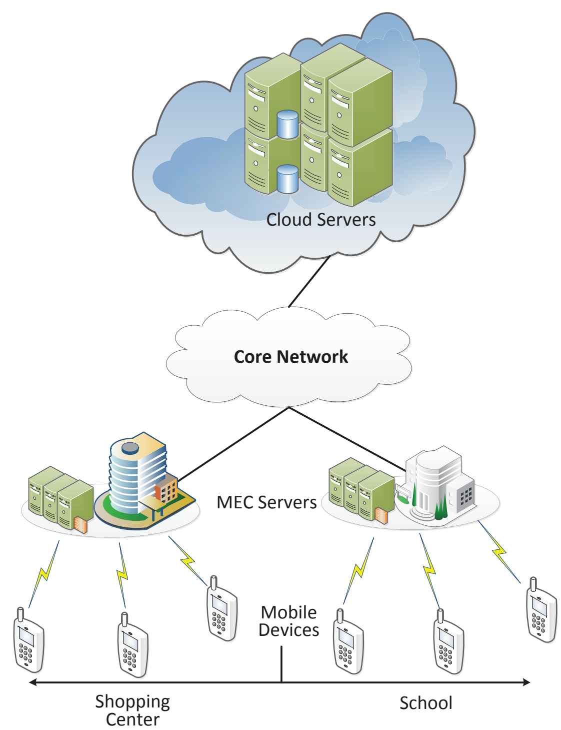

Fig. 1: Mobile Edge Computing Architecture

Mobile Edge Computing enables mobile subscribers to access IT and cloud computing services at the close proximity within the range of Radio Access Network (RAN). The main goal of Mobile Edge Computing to reduce latency by bringing the computation and storage capacity from the core WAN to the edge network. Mobile Edge Computing can be defined as

"Mobile Edge Computing is a model for enabling business oriented, cloud computing platform within the radio access network at the close proximity of mobile subscribers to serve delay sensitive, context aware applications."

Mobile Edge Computing offers real time RAN information (like network load, user's location) to the application developers and content developers. These real time network information are used to provide context aware services to the mobile subscribers, thereby enriching user's satisfaction and improving Quality-of-Experience(QoE). Mobile Edge Computing platform increases the edge responsibility and allows computation and services to be hosted at the edge, which reduces the network latency and bandwidth consumption of the subscribers. Network operators can allow the radio network edge to be handled by third-party partners, this will allow to rapidly deploy new applications and edge services to the mobile subscribers, enterprises.

Fig. 1 shows a novel architecture of Mobile Edge Computing. There are three basic components in the architecture: 1) Edge devices include all type of devices (both mobile phones and IoT devices) connected to the network. 2) Edge cloud is the less resourceful cloud deployed in each of the mobile base station. Edge Cloud have the responsibility of traditional network traffic control (both forwarding and filtering) and hosting various mobile edge applications (edge health care, smart tracking *etc.*) and 3) public cloud is the cloud infrastructure hosted in the Internet.

The prime objectives of Mobile Edge Computing are:
1. Optimization of mobile resources by hosting compute intensive application, such as image processing, m-gaming, at the edge network.
2. Optimization of the large data before sending to the cloud.
3. Enabling cloud services within the close proximity of mobile subscribers.
4. Providing context-aware services with the help of RAN information.

The contributions of this paper are: (a) survey of the state-of-the-art research efforts in Mobile Edge Computation. (b) Devisal of taxonomy based on various parameters such as characteristics, actors, access technologies, applications, objectives, computational platforms, and key enablers. (c) Identification of various open challenges related to the Mobile Edge Computing that impede the successful deployment. The highlighted open research challenges will provide directions to the researchers of the domain.

The paper is structured in the following sections: Section II describes some of the real-life application scenarios that highlight the potential need of Mobile Edge Computing platform. Section 3 presents survey of the current state-of-the-art research efforts in Mobile Edge Computing domain. Section 4 demonstrate the taxonomy of Mobile Edge Computing and defined different attributes involved in the taxonomy. Finally, open research challenges in Mobile Edge Computing are presented in the Section 5 and Section 6 concludes by summarising the paper and intuitions for the readers.

II. Motivation

This Section presents the current application scenarios in the field of Mobile Edge Computing to provide motivation for the researchers in the domain. Mobile Edge Computing platform is suitable for applications such as dynamic content optimization, computation offloading in IoT, mobile big data analytics, and smart transportation. These applications are not suitable to run in mobile or portable devices as applications are computational intensive and require huge storage capacity.

A. Dynamic Content Optimization

Content optimization is performed at the web hosting site to fulfils customers' expectations [18]. For this purpose, traditional content optimization uses user's web surfing history stored in the database [19]. Sometime content optimization is done by asking user's current geographical location and analysing based on that information. Content optimization can be done based on user's context-aware information dynamically [20]. With the help of Mobile Edge Computing, content optimizer can be hosted at the edge server. In this scenario, content optimizer acquires the accurate cell and RAN information (network load, network status *etc.*) dynamically and based on that content optimization is performed. Edge based content optimization enhances network performance, Quality of Experience and new services can be added.

B. Computational Offloading in IoT

As the recent exposure of wearable and low processing power IoT devices, traditional high compute-intensive applications (such as augmentation reality and surveillance system) can not be performed in the device itself [9]. The issue can be resolved by splitting the IoT applications into small tasks and some of the tasks are performed at the cloud (i.e. at core network) provided the latency and accuracy are preserved [21]. The above scenario can be optimized by offloading the tasks at the edge server without transferring the tasks to the core network. Therefore, offloading at the edge server will certainly reduce the latency. Computational offloading in Mobile Edge Computing has couple of challenges: How to split an IoT application? How to identify a task should be offloaded or not? How to synchronize the application when the user is in mobility?

C. Mobile Big Data Analytics

In the recent years, mobile phones are the major way to go online and therefore, making mobile devices the most effective way to understand and analyse to reach companies target markets[18]. Big data is a collection of large volume of data both structured and unstructured and Big data analytics is the process of analysing the big data for better decisions and strategic business moves. In traditionally big data analytics, data collection from the edge devices are transferred to the core network for big data analytics [22], [23]. This process takes high bandwidth and latency. Instead of wasting huge amount of bandwidth, Mobile Edge Computing platform can be used for big data analytics. Big data analytic can be performed at the edge of the network and after the analysis, the results can be sent to the core network. Therefore, the scenario will reduce bandwidth consumption and improve the network latency.

D. Smart Transportation

Smart transportation is to overcome the key issues faced by the city dwellers related to transportation such as poor traffic network, poor road conditions, inadequate places for parking, inadequate capacity of public transportation, road safety[24]. For example, traffic control can be automated from the edge network by collecting real time data from the camera and sensor devices installed at the road side. Sensor device can detect approaching objects (such as pedestrian and vehicles) and can measure the distance and speed of the object. Based on the collected data, traffic control can re-route vehicles by proper signaling to smart traffic light. Similarly, smart parking system can be modeled by collecting user-context information and analysis available spaces nearby the user devices from the edge network.

III. Mobile Edge Computing: State-of-the-art Research Efforts

Mobile Edge Computing platform is to facilitate IT and cloud services at the edge of the RAN within the close proximity of the edge devices. In this Section, we describe state-of-the-art research efforts in Mobile Edge Computing.

A. FemtoClouds

Habak *et al.* proposed FemtoClouds [25] system that provides a dynamic and self-configuring "multiple device mobile cloud system to scale the computation of Cloudlet by coordinating multiple mobile devices. FemtoClouds leverage the nearby unutilized mobile devices to serve "compute as a service" at the network edge, therefore reducing the network latency during computational offloading to the traditional cloud data center. Devices in FemtoClouds can be classified in three ways: cloudlet which enable to create a Wi-Fi access point and acts as a control device, compute cluster is a group of mobile devices willing to share hardware and resources, and mobile devices trying to use the compute as a service for offloading.

Initially, a mobile device sends the device information (computation availability of the mobile device, computation available to share, utilization history) and sharing policy to the cloudlet by using the Wi-Fi network to join the compute cluster. Cloudlet can refuse to include the new device to the compute cluster based on device's computational availability and battery level. A mobile device offloads an intensive task by sending the computational code, input and output data size to the cloudlet. The input task is scheduled to the compute cluster by calculating the required computational time (if available otherwise uses Mantis system [26]) at available mobile devices. FemtoClouds control uses greedy heuristic approach optimization model during the scheduling of incoming tasks to mobile devices.

FemtoClouds system provides a community-based computing service by minimizing the dependent on corporate infrastructure and scaling the computation capability. Performance of computation offloading in system depends on the number of mobile devices available in the compute cluster and unused hardware resources of the devices.

B. REPLISOM

Abdelwahab *et al.* proposed REPLISOM[21], a mobile edge cloud architecture to reduce cloud responsiveness when multiple IoT devices replicate memory objects to the edge cloud through LTE environment. REPLISOM architecture augments the evolved NodeB (eNB) with cloud computing resources at the edge that provide clone virtual machine, storage and network resource for specific IoT application. In REPLISON, instead of pushing the new updated memory object, the edge cloud pulls the memory replica to the respective clone virtual machine. The LTE-optimized memory replication protocol in REPLISOM uses sparsity in memory replication to deduce the number of communication. When multiple IoT devices (suppose ith number of devices out of total n devices) try to update memory objects, instead of sending them to the edge cloud, each device sends the new updated memory object to the neighbor devices using Device-to-Device communication technology. The receiving device (supposed jth) compressed the receiving memory replicas into a single replica. The edge-cloud periodically sends pulling request (using pre-scheduled uplink grants) to the respective IoT device. When device J receives the pulling request from the edge cloud, in response it pushes the compressed replica to the edge cloud. Finally, the edge cloud recover the memory object by using compressed sampling construction algorithms and stored to the respective virtual machine.

REPLISOM architecture reduces the latency and cost during offloading when multiple IoT device replicates their updated object to the nearby edge cloud.

C. ME-VOLTE

Beck *et al.* proposed Mobile Edge Computing enabled Voice over LTE (ME-VoLTE) architecture to reduce battery consumption of mobile devices during video call and provide a communication protocol for negotiating the offloading strategy[27]. Here, the process of video encoding during video call procession is offloaded at the MEC edge server. VoLTE is based on the IP Multimedia system and the main two components in VoLTE architecture are 1. Proxy/Serving Call State Control Function or (P-/S-CSCF) whose task is to send signal between mobile devices (UE) and VoLTE network. 2. Media Resource Function (MRF) is a part of VoLTE network responsible for media mixing, playback of stored media, and media transcoding. When a mobile device tries to make a video call, the request (along with set of codec names) is send to MRFs by using the proxy (P/S-CSCF). MRF negotiates the type of encoding based on the current available computing resource, and uplink strength. After the negotiation, mobile device encode the video by choosing one of the codecs that were (sent by the P/S-CSCF) and send the media to a ME-MRF. The VoLTE protocol uses session initiation procedure during video call procession.

ME-VoLTE is Mobile Edge Computing based on video telephony system to reduce energy consumption during video call. .

D. Multi-User Computation Offloading

Chen *et al.* proposed a distributed computational offloading model for Mobile Edge Computing[28]. The proposed model utilizes game theoretical approach to achieve the Nash equilibrium of the multi-user computation offloading game. When multiple devices select to offload to the telecoms cloud simultaneously using same wireless channel, then task is offloaded only if computation time is reduced after offloading reduces and saves energy consumption. The author formalizes the offloading decision problem for mobile-edge cloud computing in a multi-channel wireless environment as a multi-user computation offloading game. In the multi-user computation game, offloading decision depends on the performance value of the total cloud resource users. Here, a device decides to offload to the cloud if there is no violation on the Nash Equilibrium.

The advantage of the above model is that the a model solves the NP-hard problem of multi-user computational offloading.

E. Computation Offloading Among the Peers of Mobile Devices

Gao proposed a probabilistic computational offloading framework that offloads mobile computation among the peers of mobile devices within the tactical edge[29]. In war-zone area, application such as processing in-situ sensory data about the nearby environment takes huge computation. The proposed framework offloads some of part of application to the nearby mobile nodes to reduce computational time and energy consumption. The offloading decision by a node is depend on the computational power, energy level of the neighbor node and probable future contact between them. The two nodes can predict their future connection by applying the properties of inter-contact time (ICT) distribution among mobile nodes. According to the proposed framework, before offloading the mobile node calculates the computational time of the task at the new node, energy consumption and their future contact. The new node offloads the task if the time and energy consumption is reduced after offloading and if the new node ensures to complete the task within the contact period. If opportunity of task offloading is available using the above technique, the task is offloaded recursively to the nearby nodes.

The proposed model efficiently distributes the workload of a mobile node (in Tactical Edge) to the neighboring pear nodes recursively. The model reduces the energy consumption and high task throughput. Performance of the above model depends on the number of neighboring node in Tactical Edge network. The model does not consider the implication, if a node leaves the network suddenly.

F. Successive Convex Approximation Algorithm Framework for Computational Offloading

Sardellitti *et al.* proposed successive convex approximation (SCA) algorithm framework to optimize radio and computational resources for computational offloading in a multi-cell Mobile Edge Computing environment[30]. The optimization problem is formulated as a joint optimization of both radio and computing resources of the mobile users by minimizing the mobile device energy consumption and latency during offloading under the power budget. In the framework, the edge network has multiple Small Cell enhanced Node B (SCeNB) linked to the common cloud server provider for offloading purpose. Mobile Users within the same cell uses orthogonal channels to communicate with the cloud, and since, this framework is also for multi-user in dense cells, mobile users accessing different cell may interfere each other. Mobile users can offload a task in remote cloud or execute locally based on the computational resource provided by the cloud server and battery level of the device. The authors first formulate the above offloading scenario for a single user where only a mobile user is accessing the cloud resource. In single user case, the problem of optimization is non-convex but author attained a global optimal solution in closed form by casting the problem into convex optimal problem. When cloud resources are accessed by the multi-user in dense cell environment, the centralized and distributed SCA-based algorithm obtained local optimal solutions of the non-convex optimization problem.

The proposed algorithm shows better result in compare to disjoint optimization schemes. The framework considers only static values to perform resource optimization.

G. Edge Accelerated Web Browsing (EAB) Prototype

Takahashi *et al.* presented Edge Accelerated Web Browsing (EAB) prototype for Mobile Edge Computing to fasten web application execution[31]. In EAB, edge server is deployed in between the mobile client and server near mobile proximity. When a mobile web browser sends a request for a URL page, the response from the server is first intercepted at the edge server. In response, edge server excludes some of the contents. The common tasks of edge server are fetching web contents, evaluation of web contents, layout of the contents component, and task rendering. The proposed prototype outperforms then the normal web browsing.

H. Collaborative Context-aware Real Time Application

Nunna *et al.* proposed Mobile Edge Computing architecture based on 5G technologies for collaborative context-aware real time application[32]. Collaborative application in critical time is not favorable due to high latency of wireless communication. In the above architecture, MEC server is deployed in each eNodeB. Authors utilize the properties of 5G technology such as proximity service and context aware computing to achieve the collaboration. Middleware MEC collaborative platform in MEC server collect important information such as users fine-grained location, radio level *etc.*(expected in 5G) through standard APIs. The above architecture will benefit in Road Accident Scenario and Remote Robotic Telesurgery scenario.

The above model is low latency (because of 5G) and suitable for collaborative computing as latency and synchronization are important factor for performance in collaborative computing model. Although, the above model is theoretical concept as 5G technology is still in developing state.

I. CloudAware

Gabriel *et al.* proposed CloudAware[33], a programming model for Mobile Edge Computing to develop elastic and scalable Mobile Edge Computing application. CloudAware adopts Jadex[34] middleware framework to gain properties such as distributive, concurrency execution and context aware computing. In CloudAware, Discovery Service takes the responsibility for monitoring available networks, network strength, available surrogate computing resources for offloading and workload of each server. Context Manager collects the user mobility status to estimate the future connectivity status of the network. The Partitioner and solver divides the applications into different components and make offloading strategy at run time based on the optimization problem (i.e. minimize computation time) and status of network connectivity.

CloudAware framework is an abstract programming model, transparent and adopted context feature. The framework is not yet to implement/simulate to know the performance and reliability of the model.

IV. Taxonomy

This Section describes the taxonomy of Mobile Edge Computing with the short description of each component. The Fig. 2 shows the taxonomy of Mobile Edge Computing. The taxonomy of Mobile Edge Computing is based on following parameters: a) Characteristics, b) Actors, c) Access Technologies, d) Applications, e) Objectives, f) Computation Platforms, and g) Key Enablers.

A. Characteristics

Mobile Edge Computing can be categorized by the following properties:

1) Proximity: In Mobile Edge Computing, edge network is accessed by the mobile devices using RAN. Mobile or portable devices can also connect to the nearby devices through device to device (D2D) communication and simultaneously mobile devices can access edge server located at the mobile base station. Since, edge server is nearby to devices; it can extract device information and analyze users behavior to improve services.

2) Dense Geographical Distribution: Mobile Edge Computing host IT and cloud computing services at the edge network which sits at numerous locations. Dense geographical dispersed infrastructure contributes in many ways. Services can be provided based on user mobility without traversing the entire WAN.

3) Low Latency: One of the goals of Mobile Edge Computing is to reduce latency when accessing the core cloud. In Mobile Edge Computing, applications are hosted at the Mobile Edge sever or cloud located at the edge network. Since, the available bandwidth within the edge network is high in compare to the core network, average network latency is reduced.

4) Location Awareness: As the mobile devices are at the close proximity of the edge network, base station collects user's mobility pattern and predict the future network status. Application developers uses user location to provide context-aware services to the user.

5) Network Context Information: Real time RAN information (such as subscriber location, radio condition, network load *etc.*) are used to provide context related services to the mobile subscriber. RAN information are used by the application developers and content provider to service providers of services, thus improving user satisfaction and Quality-of-Experience(QoE).

B. Actors

Mobile Edge Computing environment comprises of many individuals and organizations each having different roles, contributing a platform to provide context aware, low latency, on-demand cloud services within the range of RAN. The overall goal of Mobile Edge Computing is to bring sustainable business model for all the actors and to enable global market growth. Some of the actors are applications developers, content provider, mobile subscribers, mobile edge service provider, software vendors and over-the-top content(OTT) players.

C. Applications

There are huge potential of offering a range of applications by the Mobile Edge Computing. The recent applications in Mobile Edge Computing can be classified as computational offloading, collaborative computing, memory replication in IoT and content delivery. These applications perform computation at the edge network thereby improving network latency, utilizing high bandwidth. Above applications use network context information to improve user's satisfaction by providing different services even when the user is in moving condition.

1) Computation Offloading: Many mobile applications are compute intensive such as face recognition, speech processing applications, m-gaming *etc.* But, running compute intensive application in resource constrained devices consumes lots of resources and energy. Instead of running in mobile host, part of the computation is transferred to the cloud data center and returns the result after successfully executing the task. Since, communication between edge device and core cloud takes long latency, in Mobile Edge Computing, low resourceful server is deployed at the edge of the network. Thereby, compute intensive tasks are offloaded.

2) Collaborative computing: Collaborative computing brings many individuals and organizations to collaborate each other in a distributed system. Applications of collaborative computing in current scenario are ranging from simple sensors devices to remote robotic telesurgery. In such type of applications, location of the devices and communication latency are key role during communication. Augmenting real time collaborative application in mobile edge environment provides a strong real time context aware collaborative system within the edge network.

3) Memory replication in IoT: In the recent year, LTE is becoming the dominant connectivity technology for the devices. IoT devices are less capable of computing and storage. These devices collect data from the surrounding and offload them as a memory object to the scalable cloud infrastructure for further computation. Since, number of IoT devices are growing, simultaneously replicating memory objects creates network bottleneck due to high latency. The edge network in Mobile Edge Computing hosts multiple clone cloud for each devices, bringing the computing capacity at the proximity of the IoT device. This reduce the network latency .

4) Content Delivery: Content delivery technology optimizes web content at the web server to provide services with high availability, high performance and to reduce network latency. Traditional web content delivery cannot acclimatize user request once optimization is done. Mobile Edge Computing can provide dynamic optimization of web contents based

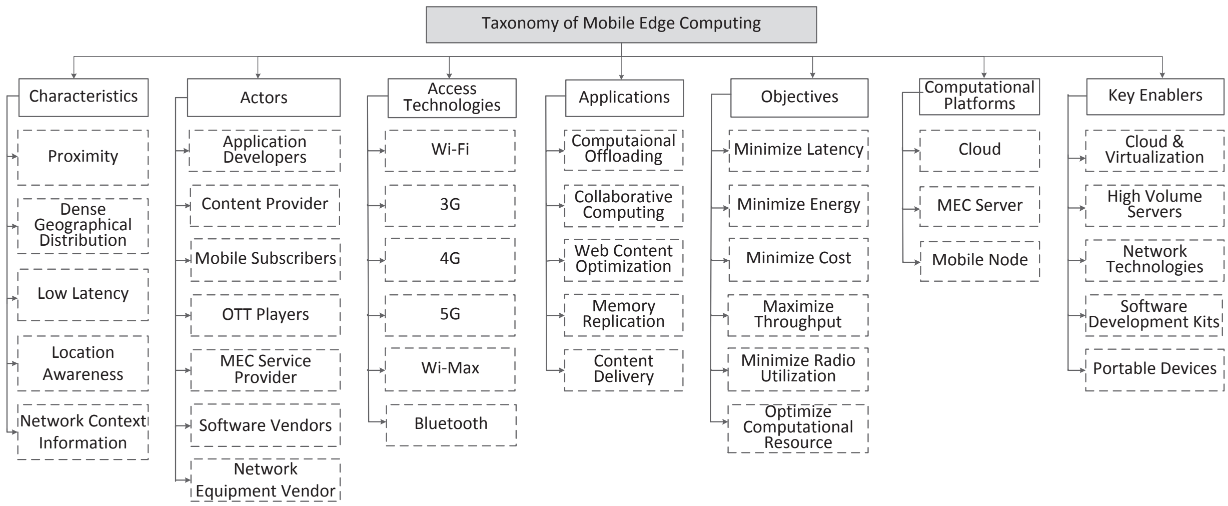

Fig. 2: Taxonomy of Mobile Edge Computing

on the network status and available network load. Since the devices are at the close proximity, edge server can exploit user mobility and service experience to render the content optimization.

D. Access Technologies

In Mobile Edge Computing environment, mobile or portable devices communicate with other devices or with the edge network by using wireless communication such as cellular network (GPRS/CDMA/3G/4G/Wi-MAX) or Wi-Fi-access point. Since, the network is deployed densely , a user can connect to the edge network by switching to any of the available access network .

E. Objectives

The objective attributes define the primary goal of the Mobile Edge Computing. Each component in Mobile Edge Computing such as mobile node or network operator have different objectives. Mobile node tries to minimize communication latency and energy consumption of the mobile devices leveraging the computation and storage at the Mobile Edge Computing infrastructure. The objective of network provider is to minimize the cost of infrastructure and attaining high throughput.

F. Computational Platforms

The computation platform attributes represent different computation hosts in the Mobile Edge Computing platform. In peer-to-peer computing, a task is offloaded to the neighboring mobile devices which are in the close proximity. A task can be offloaded to the nearby cloud deployed at the edge network. In Mobile Edge Computing, mobile edge servers are deployed in each base station.

G. Key Enablers

Realization of Mobile Edge Computing technology in real world is due support by the various key technologies. Key enablers attribute represents different technologies which contribute to provide context-aware, low latency, high bandwidth services to the mobile subscribers at the RAN close proximity.

1) Cloud & Virtualization: Virtualization allows to create variant logical infrastructure in the same physical hardware. Cloud computing platform at the edge of the network creates different virtual machines using virtualization technology to provide different services of cloud computing namely Software-as-a-Service(SaaS), Platform-as-a-Service(PaaS), and Infrastructure-as-a-Service(IaaS).

2) High Volume Servers: Traditional high volume server or Mobile Edge Servers are deployed in each mobile base station of the edge network. Mobile edge server performs traditional network traffic forwarding and filtering and also responsible to execute offloaded task by the edge devices.

3) Network Technologies: Multiple small cells are deployed in Mobile Edge Computing environment. Wi-Fi and cellular networking are the main networking technologies used to connect the mobile devices with the edge sever.

4) Mobile Devices: Portable devices at the edge network compute low intensive task, and hardware related tasks which are non-offloadable to the edge network. Portable devices also perform peer-to-peer computing withing edge network through Device-to-Device communication.

5) Software Development Kit: Software Development Kit (SDK) with standard Application Programing Interface (API) helps in adapting existing services and foster on expediting the development of new elastic edge applications. These standard APIs can be easily integrated in application development process.

V. Open Research Challenges

Mobile Edge Computing is still in infancy state. There are many challenges to overcome in this field. In this Section, we identified some of the open research challenges in Mobile Edge Computing. The description of each open issues will give a glimpse idea for researcher to overcome the challenging problems.

A. Standard Protocol

Standardization is a process to bring an open environment for researchers and industries to work in a single agreed platform. Mobile Edge Computing is a recent technology that has not been implemented properly. So, a standardized open environment is needed to be created for Mobile Edge Computing that will allow seamless and proficiently integration of traditional applications across the Mobile Edge Computing platforms. The standard platform will also accelerate the rapid development of mobile edge applications across the industry, and ultimately increasing the market size. Standard protocols are required to implement standard characteristics of Mobile Edge Computing such as context-aware information, offloading scenario. Once the standard protocols are available, researchers and academician can improve the protocol after implementing or modeling in the real platform.

B. Simulation Platform

Simulation Platform is a process of modeling a real world system with mathematical formula and it can be implemented using a general programming language. The model should have the capacity to reconfigure and experiment in different scenario. The advantage of using simulation is that it helps us to understand the whole system and feasibility (in cheap cost) without implementing the model in real[35]. The cost of developing a Mobile Edge Computing infrastructure for experimentation and testing need a lot of effort and financial investment. Instead of that, developing a simulation platform for Mobile Edge Computing will encourage researchers to experiment various scenario of Mobile Edge Computing.

C. Mobility Management

Mobility is one of the reason for frequent disconnection of link between devices and edge network in Mobile Edge Computing platform. When the device is in moving state, service quality of the application is degraded due to varying properties of network parameters such as delay, bandwidth, jitter *etc.* One of the challenging issues in Mobile Edge Computing is implementing a mobility management technique with which user can access edge application without any disconnection. Mobility management technique should take care of both horizontal and vertical mobility.

D. Heterogeneity

Edge network in Mobile Edge Computing is highly heterogeneous in terms of wireless network interfaces. Edge devices can access services through different available radio access technologies such as Wi-Fi, 3G, 4G, WiMAX and 5G. Due to high probability of signal interference in wireless networks, how to manage the switching of network without degrading the fundamental property of Mobile Edge Computing (i.e. low latency, high bandwidth, and user location awareness) is one of the challenging issues.

E. Pricing Model

User mobility is common in Mobile Edge Computing where the network is heterogeneous. Appropriate pricing model should be made if the user access to edge services from roaming base station. Pricing model should be dynamic based on the network parameters of the access technologies such as latency, session reestablishment delay, jitter, bandwidth availability, and quality of security.

F. Scalability

Scalability property ensures the availability of the service regardless of the number of client devices in the edge network. In the normal scenario, most of the devices don't access edge application simultaneously. In recent years, the number of the edge devices (such as mobile devices, IoT devices *etc.*) are growing and if huge number of devices access a particle service simultaneously, this will implicate network bottleneck and ultimately the service may interrupt. The edge server should ensure scalability of the service by applying load balancing mechanism, Server clusters.

G. Security

Security is one of the challenging issues in Mobile Edge Computing platform where applications are hosted at the edge network. In contrast to traditional computing, application developers have no role in challenging the security policies enacted by the radio network. Before the application is hosted in mobile edge server, following security challenges must be resolved. The application hosted at the edge server should authenticate a user accessing the application resources whether the one it claims to be. Mobile edge server must protect applications and data store at the edge server from intrusion. On the other hand, mobile device needs to authenticate the edge application accessing from the edge server. At last Mobile Edge Computing platform must give assurance of data integrity.

VI. Conclusion

Mobile Edge Computing brings computation and storage capacity of traditional core network within the range of the radio of access network. In this new architecture, traditional base station not only perform traffic control, but also deploy less resourceful edge server/cloud to provide context-aware services towards mobile subscribers within the close proximity. The primary objective of Mobile Edge Computing is to provide application and services with less latency and minimum bandwidth.

The paper investigates real time application scenarios which are fit in Mobile Edge Computing platform. Then, a survey on the current state-of-the-art research effort in Mobile Edge Computing are described in detail. Based on the survey, a taxonomy on Mobile Edge Computing is presented with clear description of each attributes. Later on, major open research challenges in successful deployment of Mobile Edge Computing are identified.

REFERENCES

[1] E. Ahmed, A. Gani, M. Sookhak, S. H. Ab Hamid, and F. Xia, "Application optimization in mobile cloud computing: Motivation, taxonomies, and open challenges," *Journal of Network and Computer Applications*, vol. 52, pp. 52–68, 2015.

[2] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches," *Wireless communications and mobile computing*, vol. 13, no. 18, pp. 1587–1611, 2013.

[3] D. B. Hoang and L. Chen, "Mobile cloud for assistive healthcare (mocash)," in *Services Computing Conference (APSCC), 2010 IEEE Asia-Pacific*. IEEE, 2010, pp. 325–332.

[4] G. Sun and J. Shen, "Facilitating social collaboration in mobile cloud-based learning: a teamwork as a service (taas) approach," *Learning Technologies, IEEE Transactions on*, vol. 7, no. 3, pp. 207–220, 2014.

[5] J. Wu, C. Yuen, N.-M. Cheung, J. Chen, and C. Chen, "Enabling adaptive high-frame-rate video streaming in mobile cloud gaming applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[6] K. Sabarish and R. Shaji, "A scalable cloud enabled mobile governance framework," in *Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS), 2014 IEEE*, Sept 2014, pp. 25–34.

[7] E. Ahmed, A. Gani, M. K. Khan, R. Buyya, and S. U. Khan, "Seamless application execution in mobile cloud computing: Motivation, taxonomy, and open challenges," *Journal of Network and Computer Applications*, vol. 52, pp. 154–172, 2015.

[8] *Ericsson Mobility Report, 2013; http://www.ericsson.com/res/docs/2013/ericsson-mobility- report-june-2013.pdf*.

[9] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[10] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for vm-based cloudlets in mobile computing," *Pervasive Computing, IEEE*, vol. 8, no. 4, pp. 14–23, 2009.

[11] M. Satyanarayanan, "A brief history of cloud offload: A personal journey from odyssey through cyber foraging to cloudlets," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 4, pp. 19–23, 2015.

[12] E. Ahmed, A. Akhunzada, M. Whaiduzzaman, A. Gani, S. H. Ab Hamid, and R. Buyya, "Network-centric performance analysis of runtime application migration in mobile cloud computing," *Simulation Modelling Practice and Theory*, vol. 50, pp. 42–56, 2015.

[13] J. Liu, E. Ahmed, M. Shiraz, A. Gani, R. Buyya, and A. Qureshi, "Application partitioning algorithms in mobile cloud computing: Taxonomy, review and future directions," *Journal of Network and Computer Applications*, vol. 48, pp. 99–117, 2015.

[14] *EC2 - Amazon Web Services https://aws.amazon.com/ec2/*.

[15] *Microsoft Azure: Cloud Computing Platform and Services https://azure.microsoft.com/en-in/*.

[16] E. PORTAL, "Mobile-edge computing-introductory technical white paper," 2014.

[17] M. Beck, M. Werner, S. Feld, and S. Schimper, "Mobile edge computing: A taxonomy," in *Proc. of the Sixth International Conference on Advances in Future Internet*, 2014.

[18] R. R. Sarukkai and A. Mendhekar, "Method and apparatus for accessing targeted, personalized voice/audio web content through wireless devices," Apr. 27 2004, uS Patent 6,728,731.

[19] G. Simmons, G. A. Armstrong, and M. G. Durkin, "An exploration of small business website optimization: Enablers, influencers and an assessment approach," *International Small Business Journal*, vol. 29, no. 5, pp. 534–561, 2011.

[20] J. Zhu, D. S. Chan, M. S. Prabhu, P. Natarajan, H. Hu, and F. Bonomi, "Improving web sites performance using edge servers in fog computing architecture," in *Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on*. IEEE, 2013, pp. 320–323.

[21] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Replisom : Disciplined tiny memory replication for massive iot devices in lte edge cloud," *Internet of Things Journal, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.

[22] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[23] D. DEV and R. PATGIRI, "Dr. hadoop: an infinite scalable metadata management for hadoophow the baby elephant becomes immortal!"

[24] S. Madakam and R. Ramaswamy, "The state of art: Smart cities in india: A literature review report," *International Journal of Innovative Research and Development*, vol. 2, no. 12, 2013.

[25] K. Habak, M. Ammar, K. Harras, and E. Zegura, "Femto clouds: Leveraging mobile devices to provide cloud service at the edge," in *Cloud Computing (CLOUD), 2015 IEEE 8th International Conference on*, June 2015, pp. 9–16.

[26] Y. Kwon, S. Lee, H. Yi, D. Kwon, S. Yang, B.-G. Chun, L. Huang, P. Maniatis, M. Naik, and Y. Paek, "Mantis: Automatic performance prediction for smartphone applications," in *Proceedings of the 2013 USENIX conference on Annual Technical Conference*. USENIX Association, 2013, pp. 297–308.

[27] M. T. Beck, S. Feld, A. Fichtner, C. Linnhoff-Popien, and T. Schimper, "Me-volte: Network functions for energy-efficient video transcoding at the mobile edge," in *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on*. IEEE, 2015, pp. 38–44.

[28] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *Networking, IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[29] W. Gao, "Opportunistic peer-to-peer mobile cloud computing at the tactical edge," in *Military Communications Conference (MILCOM), 2014 IEEE*, Oct 2014, pp. 1614–1620.

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

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

[32] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, "Enabling real-time context-aware collaboration through 5g and mobile edge computing," in *Information Technology - New Generations (ITNG), 2015 12th International Conference on*, April 2015, pp. 601–605.

[33] G. Orsini, D. Bade, and W. Lamersdorf, "Computing at the mobile edge: Designing elastic android applications for computation offloading."

[34] A. Pokahr and L. Braubach, "The active components approach for distributed systems development," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 28, no. 4, pp. 321–369, 2013.

[35] A. Ahmed and A. Sabyasachi, "Cloud computing simulators: A detailed survey and future direction," in *Advance Computing Conference (IACC), 2014 IEEE International*, Feb 2014, pp. 866–872.