Mobility-Based Resource Allocation and Provisioning in Fog and Edge Computing Paradigms: Review, Challenges, and Future Directions

Sudheer Kumar Battula, Ranesh Kumar Naha, Ujjwal KC, Khizar Hameed, Saurabh Garg, and Muhammad Bilal Amin

Abstract Fog and Edge related computing paradigms promise to deliver exciting services in the Internet of Things (IoT) networks. The devices in such paradigms are highly dynamic and mobile, which presents several challenges to ensure service delivery with the utmost level of quality and guarantee. Achieving effective resource allocation and provisioning in such computing environments is a difficult task. Resource allocation and provisioning are one of the well-studied domains in the Cloud and other distributed paradigms. Lately, there have been several studies that have tried to explore the mobility of end devices in-depth and address the associated challenges in Fog and Edge related computing paradigms. But, the research domain is yet to be explored in detail. As such, this chapter reflects the current state-of-the-art of the methods and technologies used to manage the resources to support mobility in Fog and Edge environments. The chapter also highlights future research directions to efficiently deliver smart services in real-time environments.

Keywords Internet of things · Mobility · Resource allocation · Fog computing and Edge computing

1 Introduction

In recent years, there has been a rapid rise in the Internet of Things (IoT) devices and applications, and the range of the services offered by them. The services include smart transportation [1], disaster-related services [2], smart cities [3] and so on. These devices and applications provide services in a Cloud-like manner

S. K. Battula (✉) · R. K. Naha · U. KC · K. Hameed · S. Garg · M. B. Amin
School of Information and Communication Technology, College of Sciences and Engineering, University of Tasmania, Hobart, TAS, Australia
e-mail: sudheerkumar.battula@utas.edu.au; raneshkumar.naha@utas.edu.au; ujjwal.kc@utas.edu.au; hameed.khizar@utas.edu.au; saurabh.garg@utas.edu.au; bilal.amin@utas.edu.au

251

close to users with the help of end devices [4, 5]. Accordingly, new computing paradigms, such as Fog and Edge have evolved. Fog computing provides users with a decentralized environment where computational resources are brought to the Edge of the network so that they can perform real-time computational work on data without compromising bandwidth and latency issues in Cloud networks. In Fog computing, computational resources, including applications, services, and data, are placed between data and the Cloud. In Edge Computing, several endpoints (powerful IoT devices) are placed near the edges of the network to perform computations on data sources where data is generated, without having to be transferred elsewhere [6, 7].

In paradigms like Edge and Fog, not only the end devices but the users also, are mobile. The allocation and provisioning of the resources should be done in real-time within strict time constraints, even without any prior knowledge on future dynamics of the user mobility to offer and deliver seamless services without any interruption [8]. Additionally, such real-time capabilities also require resources (or a pool of resources) that are capable of completing the task or the application within a given time window. Consequently, Mobile Computing (MC) environments need to choose and provision adequate resources in an optimal manner, which can be non-trivial. For both requirements of effective resource allocation and provisioning, the mobile nature of the Fog, Edge, and end devices invites a number of challenges that can interrupt the execution of the tasks. On top of the mobility, various hardware and OS capabilities of the devices add more complexities to resource management [9]. For example, due to Fog device versatility, hardware and software of a new Fog device that is meant to replace a failed one, may not be compatible or match with the failed device while attempting to migrate the application to a nearby device in a Fog environment and consequently, the execution of the application fails. Due to different operating systems installed on each Fog device and frequent updates to the operating system, the devices can be vulnerable to several kinds of attacks and errors arising from incompatible versions of the devices [10].

In Mobile Edge Computing (MEC) paradigms, it is desirable to have an effective resource management mechanism that achieves efficiency and high performance for both systems and networks, while ensuring the fast and efficient distribution and delivery of resources to users and their processes. To ensure efficiency and high performance, some essential factors such as cost-efficient computation, bandwidth utilization, minimum delays, and energy consumption of the mobile devices are important milestones that have drawn researchers' attention to suggesting key resource management techniques in the MC environment [11]. In cost-efficient computation, Mobile Edge devices first meet user process requirements and then define appropriate procedures for managing available and allocated resources to maximize overall system profitability. The quality of services is also an important factor to consider while providing and allocating mobile resources, which necessitates a dynamic need for the resources available to the process to improve the overall quality of service within the network. Moreover, processing IoT applications intermediate computing devices often leads to several other difficulties in device managing systems [12]. For example, inadequate security protocols and frequent

joining and leaving devices can lead to various security attacks and, in some cases, can cause data leakage. On the other hand, due to the limited energy and power of Fog, Edge, or end devices, the framework needs to be designed with a view to energy-saving capabilities as well [13]. Additionally, most of the applications are time-sensitive, and executing the applications within strict deadlines can be difficult [14].

On the other hand, the mobility of the end devices is entirely different from other computing paradigms as it does not follow any single random distribution. The mobility of the devices is direction-oriented as well as it depends on the source and destination of the devices. As a result, modeling the real environment is also a challenging task. Nonetheless, it is undeniable that when it comes down to effective resource allocation and provisioning in MC paradigms, such as Fog and Edge, the mobility of the end devices should be considered along with different other requirements. Several works [15, 16] have already conducted comprehensive sweeps of existing techniques in resource management in Fog and Edge computing paradigms. But, given the fact that the mobility of the end devices in such paradigms are now being explored in depth in several dimensions, a survey that reflects the current state-of-the-art of all the existing techniques, primarily centered around the mobility, is still missing. As such, with all these aspects in hindsight, this chapter first discusses in detail the different requirements and challenges in each layer of Fog, Edge, or other intermediate computing devices to achieving an effective MEC environment. The chapter then provides a comprehensive picture of the current works carried out to handle mobility in the Fog and Edge environment to put light on the future directions where future works can be directed to deliver high-quality IoT services in a computer paradigm characterized by mobility and dynamicity. The specific contributions of the chapter are as follows.

1. A summary of the requirements and challenges in each layer of MC paradigms to effective resource management
2. A comprehensive sketch of existing resource allocation and provisioning mechanisms in Fog and Edge computing centered around the mobility of the end devices
3. A reflection of the techniques used to model the mobility in MC paradigms
4. A summary of the existing mathematical models used for resource allocation in Fog and Edge computing
5. An overview of how future research works should address the challenges, including the mobility-related ones to ensure effective resource allocation in MEC

The work is organized as follows. Section 2 sheds light on the existing mobile-based resource provisioning and allocation mechanisms in Edge computing, while Sect. 3 follows the same summary for Fog computing. Section 4 explains in detail the existing modeling techniques used to add/support mobility in Fog and Edge computing environment. Section 5 briefly discusses several mathematical models considered in mobility-based resource allocation, while Sect. 6 presents various case studies with applications and use cases. Section 7 highlights the future direction

for the research on mobility-based resource allocation and provisioning in MC paradigms, while Sect. 8 concludes the work.

2   Existing Mobile Based Resource Provisioning and Allocation Mechanisms in Edge

Through the age of IoT, Fog/Edge computing is the exciting option to meet the latency requirements of IoT applications by shifting the service delivery from the Cloud to the Edges. This also allows lightweight IoT devices to improve their scalability and energy consumption, provides contextual information processing, and mitigates the backbone network traffic burden [17]. Computation offloading is an important feature for Edge level processing where application offloads to the mobile devices that are resource-constrained to meet process requirements with moderate computation demands. The mobility feature of such an environment makes resource allocation and provisioning a challenging task. This section discusses resource allocation and provisioning in Mobile Edge related computing environments.

Zhang et al. [18] implemented a network-slicing 5G system conceptual architecture and proposed convergence management mechanisms between separate access networks and shared power/subchannel allocation schemes in two-tier network slicing-based spectrum sharing structures. In their proposed method, conflict with the co-interference, and cross-tier interference are taken into account. The core network is developed into a historically distributed architecture to decrease monitoring and data transfer delays by control plane separation from the user plane. The core Cloud offers some essential control features such as virtualized resource management, mobility management, and intrusion management. In the Edge Cloud, which is a centralized body of virtualized services, we can find servers and other applications in the radio access network. The Edge Cloud conducts data transfer primarily, and control functions including the encoding of the basebands. To optimize uplink efficiency on each subchannel, research modeled the uplink resource assignment issue for small cells considering the four limitations. First and foremost, each small cell consumer transmits its full power. Secondly, each user of ultra-Reliable and Low Latency Communication (uRLLC) is required to have the minimum data rate requirement. Thirdly, the minimum interference level is obtained from small cell users by the macrocell. Finally, in one transmission cycle, a subchannel can be assigned to a maximum of one consumer in each small cells. Computer simulations were used to show the promising performance of network-based 5G networks.

Ren et al. [17] explored the implementation of Edge computing to construct transparent IoT systems by implementing an IoT architecture focused on computing in transparency. Edge Computing architecture attempts to solve two big IoT applications issues—the first being how to manage data processing requests in real-time and context, and the second being how can IoT tools, the core component of transparent calculation, be distributed on-demand applications/services dynami-

cally. The work also identified benefits and related challenges for the architecture. The architecture provides a scalable IoT platform to deliver intended services for lightweight IoT devices on time, to respond to changing user needs. The IoT architecture, based on transparent computing, can also offer several advantages: (1) reduced response delay (2) enhanced functional scalability (3) centralized resource management (4) on-demand and cross-platform service provisioning and (5) context-aware service support. The projects still face several challenges to their execution, which remarkably hamper the development of associated applications even though the architecture could utilize the advantage of transparent computing for the construction of IoT platforms with scalability. These challenges are: (1) unified resource management platform for heterogeneous IoT devices (2) provision service dynamically on lightweight IoT devices and (3) allocation of computation between the Edge and end devices.

In the Edge network Tasiopoulos et al. [19] built an auction-based allocation and resource provision which maps the application instances known as Edge-MAP. Edge-MAP takes account of the versatility and restricted computing resources available in the Edge micro clouds for the allocation of resources for device bidding applications. The provision of geo-distributed cloudlet services to low latency applications differs in that the allocation of resources must bear in mind the effect of network conditions, such as latency between end-users and Cloud presence points, on the QoS applications. Besides, for the mobile users, as users switch the connectivity to another base station, the latency to an assigned VM changes. Even ideal VM allocations for user requests are thus redundant over time, i.e. user mobility/handoffs are accompanied by VM reassignments if necessary. When a fixed number of VM's are allocated for a long time to an application, VM handoffs lead to available VM's that can be utilized by the users of other apps instead. In this context, to prevent the misuse of VMs, we argue about the need to include low-latency on-demand applications, in which a VM would be instanced on application request for the period of the engagement of the end-user. The provision of VMs takes place through periodical/discrete operation of the auction mechanisms, where the minimum length of the auction and VM configuration for each length/time slice is limited by the added overhead. However, the work did not demonstrate how the provisioning method will work in a complex heterogeneous IoT and edge/Fog environment.

Liu et al. [20] proposed a blockchain platform for MEC video sharing with the adaptive block size. First of all, they developed an incentive system for the partnership between producers of content, video transcoders, and consumers. Besides, the work revealed a block-size video-streaming adjustment system. Also, the work considered two offload modes to avoid the risk of overloading MEC nodes. These modes are—transfer to neighboring MEC nodes and transfer to a device-to-device (D2D) user group. The work then formulated the concerns of resource assignment, download schedule, and adaptive block size as an optimization problem. The problem distribution of the work was done using a low complexity alternating direction method in the algorithm based on multipliers. They showed that the optimum policy for maximum average transcoding profit is not uniform

amongst different small cells or transcoders. They did not take account of a variety of factors such as the holding interest, communication, reputation value, and computation ability. The work also failed to demonstrate the intelligent contract as the Alternating Direction Method of Multipliers (ADMM) coordinator, a means of efficiently facilitating distributed optimization among non-trusting nodes.

Naha et al. [14] suggested the allocation and provisioning of resources through the use of hybrid and hierarchical resource rankings and provisioning algorithms for the Fog-Cloud computing paradigm. The resource allocation method assigns Fog-Cloud resources by rating the resources based on the number of resource constraints within Fog devices. The resource provisioning approach offers a hierarchical and flexible mechanism to meet the complex requirements of the users. The work achieves resource distribution and scheduling in the Fog-Cloud environment with three separate measures to reach the deadline by considering complex user requirements. In the first phase, Fog resources are identified to comply with the deadline requirements from the available devices in the Fog environment. The algorithm allocates jobs in Fog devices because the Cloud response time is high relative to the Fog. The second step is to try to accomplish this task using Fog servers and Cloud resources if available resources are not adequate in the Fog. Here it considers computing response time, bandwidth, and power for these resources when selecting Cloud resources. Thirdly, it investigates whether or not services in the Fog environment are unavailable. In such a case, the system would attempt to complete the work with Cloud resources before the resource unavailability message is generated. Dynamic changes of resources and failure to handling, such as resource and communication failures in a Fog computing paradigm, were not considered. The work also did not consider designing a simulation of a complex Fog environment with a bigdata IoT application.

A low-complex heuristic algorithm was introduced by Liao et al. [21] by providing a near-optimal low-time solution. The goal is to maximize the number of battery charged UE tasks successfully offloaded while maintaining a reasonable level of network efficiency. A low complexity heuristic algorithm was developed to resolve the formulated nonlinear mixed-integer programming problem. The work looked at the interference between all of the EU and used hyper graphic approaches in the pre-allocation of channels. Task offload was then evaluated following the estimation and allocation of contact resources. The work assumes the duality uplink and downlink to be accomplished with optimum transmitting capacity. UEs adjust to the discharge assignments upon receiving the feedback from MEC to increase the network performance, and bandwidth will be reassigned. A detailed investigation was carried out to find the relationship between the number of tasks successfully discharged and the number of UEs under different MEC computing tools. However, the research did not concentrate much on the latest communication framework for joint optimization, smart coded caching, and computation in the MEC networks and their unprecedented applications. The use of large-scale enhancement learning for the competitive pricing policy, which can equalize mobile operator cost and UEs' economic impact, has not been considered.

Table 1 provides a detailed comparison of existing techniques for mobile-based resource provisioning and allocation mechanism in Fog/Edge computing

Table 1 Comparison of existing mobile based resource provisioning and allocation mechanisms to support mobility nature edge

Ref	Problem addressed	Techniques used	Performance challenges	Use cases	Experimental setup	Evaluation criteria	Dataset used
[18]	Handover and network slicing issues in 5G networks	Joint power and sub channel allocation scheme	Quality of service	Not defined	Not defined	Capacity of slice, Transmit power, Latency	Not defined
[17]	Transparent and service provisioning issues	Not defined	Lightweight computing, Transparency, Scalability	Transparent computing based wearables	TCwatch and Smart watch	App size, Latency, Energy consumption	Smart Wearable data
[19]	Edge resource provisioning and latency	Auction based resource allocation and provisioning mechanism	Mobility, Quality of service	Vehicle as a mobile users in cellular areas	Cellular infrastructure (1 km radius)	Requests Generation, Execution time, Iterations comparison, Price, Quality of service	Mobility dataset5 (TAPASCologne project)
[20]	Resource allocation, and scheduling issues	Incentive mechanism, Low-complexity alternating direction method	Quality of service, Efficiency, Video Quality	Not defined	Network coverage radius of the MBS as 1000 m	Transcoding revenue, Bandwidth consumption, tolerance	Not defined
[14]	Dynamic changes in the parameters of user requirements	Hybrid and hierarchical based resource ranking and provisioning algorithms	Quality of service	Not defined	CloudSim toolkit	Processing time, Instance cost, Network delay	Not defined
[21]	Network resource scheduling issue in user equipment's	Centralized decision and resource allocation algorithm,	Energy consumption, Quality of service	Not defined	Matlab	Optimal transmission power, computation, Distance	Not defined

environments. Following parameters such as problem addressed, techniques used, performance challenges, use cases, experimental setup, evaluation criteria, and data set, are used for comparison purposes.

3 Existing Mobile Based Resource Provisioning and Allocation Mechanisms in Fog

Fog environment offers a wide range of interesting services with its highly mobile end devices. With mobility comes several factors that have to be considered while managing the resources during provisioning and allocation. In this section, we describe the existing works to reflect the current state-of-the-art of resource management in mobility management in the Fog environment.

Waqas et al. [22] pointed mobility as an inseparable entity of Fog computing that enables a wide range of interesting services to improve the user experience. In a mobility-aware Fog environment, the user information is constantly updated at Fog servers and nodes to increase the effectiveness of the services on offer. Such constant information updates can invite several challenges, such as prediction of user behaviors, resource constraints, geographical limitations, and so on. Developed as an extension to Cloud computing, Fog computing is agile, highly mobile, and has low latency. While offering different computing services, Fog servers and devices can be mobile and such mobility should be handled properly to ensure low latency for the services offered. Mobility management should consider not just mobile devices, but also resource management including allocation and scheduling mechanisms. In this section, we present the works that have introduced innovative measures to tackle mobility while provisioning and allocating resources in the Fog environment.

Gosh et al. [23] proposed a mobility and delay-aware framework called LOCA-TOR that offers efficient location-based resource provisioning in an intelligent transportation system. The framework uses an optimal matching algorithm of the MapReduce paradigm to minimize the time required for service-provisioning and service waiting. LOCATOR was implemented on Google Cloud Platform using different realistic datasets that characterize mobility and shown to offer less execution time than that of the baseline methods.

Babu et al. [24] proposed architecture for Fog-based node-to-node communication in 5G network. The node-to-node communication is enabled by Fog servers and data analytics unit. In the proposed architecture, the authors proposed a robust mobility management scheme for mobile users to make such communication possible. Mobility management comprises location management and calls delivery procedures. The location management procedure updates the most recent location information, while the call delivery procedure delivers the call to the target node to enable node-to-node communication. The work offers several advantages of low overhead database update cost, data loss, signal exchange, easier update, and high

security compared to similar networks. The authors also highlighted costly real-time implementation, poor results with a 3G network, limited privacy, and data replication as the major limitations of the proposed work.

Xie et al. [25] proposed a method for offloading tasks in vehicular Fog computing based on the mobility of vehicle nodes to minimize the service time. The mechanism used vehicle-to-vehicle (V2V) links to offload a task decomposed into subtasks in any proportion from user to service-vehicle in parallel. The mechanism used a hidden Markov model to predict V2V links state based on the mobility information of the vehicles collected. A rule was then set to choose the target service-vehicles and the proportion of decomposition of a task into subtasks, based on the predicted results. The mechanism was shown to have a better performance on service time and total finished tasks when compared to single-point and random task offloading.

The authors in [26] proposed URMILA (Ubiquitous Resource Management for Interference and Latency-Aware services) to switch between Fog and Edge resources for IoT services ensuring the latency of those services are met. The authors proposed a novel algorithm to find and choose the most suitable Fog node to serve IoT applications remotely when the application could be served with the Fog resource. The devised method considers the interference caused due to co-located and competing IoT services on multiple Fog nodes and controls the application executions such that the SLOs are met with low latency. The capabilities of the algorithm were tested with the real-world context on an emulated yet realistic IoT testbed.

Wang et al. [27] considered architecture for three-layered Fog computing networks (FCN) and characterized the user equipment mobility with sojourn time in each coverage of FCN. The user formulated the reduction of the probability of migration as a mixed non-linear programming problem that would maximize the revenue of user equipment. In the mixed-integer problem, the first part is task offloading, while the second part is resource allocation. The task-offloading problem was solved by using Gini coefficient-based FCNs selection algorithm (GCFSA) that gave a sub-optimal strategy. The resource allocation problem was solved by a genetic algorithm based distributed resource optimization algorithm (ROAGA). In the proposed approach, the probability of migration was significantly reduced even when the mobility of the user equipment was well-handled. Simulation results proved the supremacy of the approach over baseline algorithms in terms of quasi-optimal revenue.

Starting with an argument that presents the migration of user application modules among Fog nodes as one of the solutions to mitigate the mobility issue in the Fog environment, Martin et al. [28] proposed an autonomic framework called MAMF that handles the container migration while adhering to QoS requirements. The proposed framework borrows the concepts of MAPE loop and Genetic algorithm to suitably decide the container migration in Fog within the deadlines for each application. Under this approach, a predetermined value of use location is used for the next time instant to initiate the migration of containers. The re-allocation problem was modeled as an Integer Linear Programming problem within the framework. The experiments were conducted in iFogSim toolkit, which showed

improvement in execution cost, network usage, and request execution delay with the framework when compared to other methods.

Two different analytical models are proposed in [29] to address the issue arisen when the communication with the remote Cloud fails due to mobile devices in the Fog environment and the user does not get the requested service outputs. In the first model, remote Cloud servers were used to execute the task and the results were delivered to the mobile devices through push notification when the mobile was reconnected with the network after the mobile device lost the connection with the Cloud instance in the first place because of the mobility. In the second model, the virtual machine live migration was proposed whenever the mobile device changed the location. The present state of the instance was transferred to a new Cloudlet where the execution was resumed after the offloading. The experimental results showed a decrease in power consumption by 30–78% with the proposed models when compared to existing approaches.

Gosh et al. [30] proposed a collaborative real-time framework named Mobi-IoST that includes Cloud, Fog, Edge, and IoT layers to handle the mobility dynamics of any agent within the framework. In the framework, the spatiotemporal GPS logs along with other situational information are analyzed and fed into a machine learning algorithm that ultimately predicts the location of the moving agents in real-time. A probabilistic graphical model was used to model the mobility of the agent. The model enabled the prediction of the next location of the agent where the processed information was to be delivered. The tasks were delegated among the service nodes based on the mobility model. The framework was proven to provide better QoS in real-time applications and minimize delay and power consumption.

To fulfill the strictest requirements of enormous traffic demands and low latency, Santos et al. [31] proposed Kubernetes-based Fog architecture, which is an open-source orchestration platform on container-level. The authors implemented a network and mobility-aware scheduling approach in a smart city deployment scenario as an extension to existing scheduling mechanisms in Kubernetes. The authors validated the formulation of optimization of IoT services problem using the same approach to prove the applicability of such theoretical approaches in real-life practical deployments. The experimental results highlighted the reduction in network latency by 70% when compared to existing default scheduling mechanisms.

The authors in [32] highlighted the mobility of devices in Fog computing as one of the key factors to influence the application performance. The work then emphasizes the consideration of not just the mobility but the combination of distributed capacity and types of user application also for resource management in Fog computing to deal with the issues created by the mobile devices. The authors compared three different scheduling policies—delay priority, concurrent, and first come first serve (FCFS) and to understand the influence of mobility in resource management and improve the application execution based on their characteristics.

The authors in [27] integrated the Fog architecture with information-centric Internet of Vehicles (IoV) to provide support for the mobility of vehicle nodes through different schemas that consider data characteristics. In the proposed mechanism, the authors also considered the computation, storage, and location information of

Fog nodes for the exchange of information. The feature of IoV was also taken into consideration for communication in a mobile environment.

In a nutshell, various methods such as task offloading, live migration, and task delegation have been well-studied and implemented with demonstrated improvements when it comes to managing mobility in a computing environment regarding resource allocation and provisioning. Fast forward to the future, where all the end devices would offer seamless services to the users without any interruptions and guaranteed quality of service, the novel resource provisioning, and allocation algorithms should combine all those methods to handle the trade-offs within resource management in better ways.

4 Modelling Techniques to Support Mobility to Enhance the QoS of the Applications

Fog computing and its associated Edge-computing paradigms have been introduced to meet the demands of those complex nature applications that need efficiency, performance, reliability, mobility, and scalability at their top priority. The Fog computing approach provides the user with access through local data processing and data output rather than storing and maintaining information in extensively extracted Cloud storage facilities. Compared to Cloud computing that is often based on centralized architectures and provides the computation and storage service at fixed locations, Fog computing, and Edge computing are decentralized, distributed, and hierarchical, and their service locations are close to the end-user [7, 33].

Mobility of Fog computing is a critical challenge for most Fog computing-based real-time IoT applications because mobility requires keeping a network connection alive between sensor nodes and gateways, which are often referred to as Fog nodes. After all, network disruption often leads to unavailability of system services and, in some cases, high latency between network components. Moreover, the rapid advancements in the mobile communication technologies such as 5G wireless, it allows mobile users to off-load their computational processes to nearby deployed servers to reduce the consumption of resource-constrained devices with limited battery, memory, and processing power [34, 35].

To facilitate the application mobility function across Fog nodes, Martin et al. [28] proposed the mobility architecture to address the issue of moving containers corresponding to the user application modules. The transfer of containers is carried out in an automated mobile manner employing an automatic control loop called the Monitor–Analyze–Plan–Execute (MAPE) loop and Genetic Algorithm. The proposed structure incorporates agents who are responsible for gathering information on the environmental context to describe the implementation plan in a well-defined manner. To measure the movement of user devices to the Fog environment to which they are currently attached, the Monitoring mechanism senses a change between the user device and the distributed Fog nodes. The approach uses the pre-defined value

of the user location to launch the migration process to improve service quality for Fog nodes.

QoS management consists of various methods used to delegate essentially distributed resources to Fog-user applications, selecting an acceptable way to allocate virtual resources to the physical resource. Therefore, a Fog-based resource allocation model was introduced in [36], with the main goal of addressing the issues of the mobility of nodes, assigning tasks, and also presenting virtual machine problems in a single Fog computing context. The purpose of the proposed approach is also defined by efficient resource allocation and mobility algorithm, which focuses primarily on optimizing the distribution of resources and reducing the number of users interacting with Fog nodes for different tasks.

To resolve the problem of mobility and increased latency in the vehicle Cloud computing environment, a directional model of vehicle mobility has been proposed [37] to achieve a guaranteed level of road vehicle service. In the proposed model, the entire network is divided into three sub-models based on their movement and rotation around the network. Within each model, vehicles are responsible for communicating with others via roadside units. The purpose of the proposed model is to reduce the latency and response time of the vehicle tasks. Also, various algorithms, such as greedy search algorithms, bipartite matching algorithms are used to solve optimization and cost flow issues.

Chen et al. [38] presented the Edge cognitive computing architecture to enable dynamic service migration based on the behavioral patterns of mobile users. Advanced cognitive services based on various artificial intelligence methods are used at the Edge of the network in the proposed architecture. The benefit of inducing these services to the Edge computing model is that it can achieve higher energy efficiency and service quality relative to current Edge computing paradigms. Besides, the simulation analysis highlights the importance of the proposed architecture in terms of low latency, dynamic user interface, high system resource usage that eventually achieves overall service quality at both user and system level.

The locality of Fog computing has made it a challenge to maintain consistency as mobile users are moving through different access networks. To deal with the locality problem of Fog users, Bi et al. [39] suggested the software-defined networking (SDN) architecture that divides system features into two modules. One module is responsible for managing the mobility of users across the network, while the other module serves as a router and only handles data routing functions. Besides, to demonstrate the flexibility of the proposed model, an effective route optimization algorithm is also proposed that overcomes the problems of network performance overhead as well as delays in data communication. Efficient signaling operations are also advocated to provide mobile users with transparency and usability assistance in Fog computing. The simulation results of the proposed model illustrate the assurance of service continuity for mobile users and increased the efficiency of data transmission when re-registering users on other access networks.

Providing seamless connectivity to mobile users in Fog computing-mobile networks is a major challenge since it requires adequate load balancing mechanisms to offload computing to mobile users and a consistent channel to communicate with

others. To take this concept into account, Ghosh et al. [30] suggested Mobi-IoST—a collaborative system consisting of multiple components offering efficient delivery of various types of services to consumers, regardless of their mobile locations. Also, the proposed system is capable of providing users with an appropriate decision-making process on the knowledge and data they provide. The Cloud portion is responsible for analyzing the location data of IoT devices obtained from Fog nodes based on their mobility patterns. The mobility prediction module predominantly stores the model information and location logs of mobile users in different settings and utilizes the Markov model to make decisions on location logs to detect their locations efficiently.

High latency is also a major challenge in the Fog computing world, where mobile devices are faced with issues of proper access and control of Cloud resources by moving users around the network. To fix this problem, Zhang et al. [40] introduced an effective mobility-based method to transfer virtual machines between Edge Cloud data centers to mitigate network overhead. To meet the migration criteria of virtual machines, two algorithms (M-Edge and M-All) are used to identify machines based on mobile users in the network. Further, weight-based and predictive-based algorithms are used to support user mobility. The strength of the proposed approach is that it greatly decreases the network overhead and latency of the migration of virtual machines in the MEC environment.

Lee et al. [41] presented a Mobility Management System based on the Multi-Access Edge Computing (MEC) model that incorporates the idea of a protected access region called zones where mobile users can access server resources efficiently and monitor contents when traveling. The benefit of having the zone is that mobile users are still able to link and communicate with other mobile users within the specified boundary. In specified zones, mobile users must first register with the nearest zone and transmit their information and then access the network resources identified by the access control list.

Another mobility management issue in a high-density Fog computing environment is discussed by Rejiba et al. [42] where a user-centric mechanism is used to select Fog nodes to perform various tasks. The objective of the proposed work is to allow mobile users to learn and connect to a Fog node using a multi-armed bandit algorithm. In the learning method, an epoch model is used where the total number of Fog nodes is almost equal to the number of specifications (processes) to be completed. As a consequence, the quality of service is achieved by using a limited number of Fog nodes to execute the operation with the best possible capacity. Real-time user data and location patterns are used for simulation purposes.

In MEC environments, users with a mobility orientation frequently change their positions at various periods, so it is very hard to switch between different servers to provide them with the resources they need to maintain the quality of service. To overcome this issue, Peng et al. [43] proposed mobility intelligence and migration enables an online decision-making framework called MobMig to solve the Edge user allocation problem in an efficient and real-time manner. The proposed system completely automates current static location-aware systems that have resulted in system inefficiency and time delay invariant. Proper service selection and allocation

to the right users in the MC system is a major challenge due to resource constraints and the restricted functionality of mobile devices. To overcome these challenges, Wu et al. [44] suggested a heuristic approach by integrating genetic and annealing algorithms to accommodate multiple service requests from mobile users. The benefit of the proposed approach is that it can substantially reduce the response time in the selection and delivery of the service over the MEC network.

A user-centric mobility approach in the Fog computing environment, with the cooperation of user interface and resource allocation scenarios, is introduced by Tong et al. [47] to resolve issues of inappropriate resource allocation and lack of user experience in the context of unpredictable Fog scenarios. In the proposed approach, mix-integer non-linear programming is resolved using a novel algorithm called UCAA—low complexity two-step interactive optimal algorithm. For user experience, two decision-based algorithms, such as semi-definite programming and Kuhn-Munkres are proposed and used in the proposed scenario. Also, to fix the problem of resource allocation, the overall scenario is divided into two phases: transmission power selection and resource allocation, and each phase is individually addressed. The analysis of the proposed approach achieves substantial overall performance in the allocation of server resources and enhanced user experience in Fog computing environments.

With end-user mobility, the migration of services between different mobile users is a challenging task to ensure the quality of service and operating costs of the overall network. To resolve these challenges, Ouyang et al. [48] introduced a dynamic mobility-aware service model to balance the efficiency and cost of the computing infrastructure for end-users. To overcome the unpredictable mobility of users, the Lyapunov optimization technique is used to divide the long-term optimization problem into a set of real-time optimization problems solely based on the NP-hard problem. Also, the proposed method uses the Markov approximation algorithm to find the optimization of such real-time problems. The advantage of the proposed system is that it can substantially reduce the time complexity for large-scale end-user applications.

Optimal task scheduling and off-loading tasks in MEC are both critical challenges due to vehicle mobility, moving patterns, and varying traffic loads. To resolve these problems, the energy-efficient dynamic decision-based approach is introduced by Huang et al. [49] where a transmission system called uplink is implemented to allow traffic from vehicles to roadside units. A dynamic process (or task) offloading mechanism is used to reduce latency, energy consumption, and packet rate. Moreover, a resource allocation method is also proposed to tackle the different complexity of each vehicle and its waiting pattern for resource accessibility. For dynamic task offloading and efficient allocation of resources, a Lyapunov mechanism is proposed to ensure the efficiency of the system.

Table 2 provides a detailed comparison of existing techniques for modeling the mobility used to improve the quality of service (QoS) of the applications. Following parameters such as problem addressed, techniques used, performance challenges, use cases, experimental setup, evaluation criteria, and data set, are used for comparison purposes.

Table 2 Comparison of existing techniques for modelling the mobility used to improve the Quality of Service (QoS) of the applications

Ref	Problem addressed	Techniques used	Performance challenges	Use cases	Experimental setup	Evaluation criteria	Dataset used
[28]	Migrations of containers or applications	MAPE loop and genetic algorithm	Quality of service	Not defined	iFogSim Toolkit	Network usage, delay (Execution and Request) Monetary cost	Vehicular dataset, (General Departmental Council of Val de Marne)
[36]	Migration of applications, Resource allocation issues	Not defined	Quality of service	Not defined	CloudSim tool kit	Processing time	Not defined
[38]	Service migration issue	Edge cognitive computing	Energy efficiency, Quality of service	Not defined	Android application program	Latency, Service resolution, User Experience	Not defined
[39]	Mobility support to mobile users	An efficient route optimization algorithm	Quality of experience	Not defined	Mininet-WiFi package	System cost, Handover latency	Not defined
[30]	Mobility dynamics of the moving agent	Machine learning approaches, probabilistic graphical model	Energy efficiency Quality of service	Not defined	Matlab	Delay, Power consumption, Accuracy, Recall F-measure	Mobility dataset generated from 100 smart phones
[45]	Communication inefficiency and security of data	Not defined	Security Reliability Efficiency	Smart grid	IoT platform ThingSpeak, Matlab, IEEE 13-bus test grid	Different Simulation Times (Initialization, execution, termination), energy computation	Not defined

(continued)

Table 2 (continued)

Ref	Problem addressed	Techniques used	Performance challenges	Use cases	Experimental setup	Evaluation criteria	Dataset used
[40]	Network overhead issue because of migration of virtual machines	Naïve migration algorithms (M-All and M-Edge)	Quality of service	Not defined	Not defined	Trajectory, Overhead analysis, Performance analysis	Microsoft Geolife Trajectories 1.3
[41]	Access to MEC Architecture, Unnecessary delays in request processing	Zone based access	Quality of service	Not defined	CloudSim and Edge CloudSim	User mobility model, Number of mobile users, average task size, WLAN bandwidth, VM processor speed,	Not defined
[42]	Mobility issue in user patterns	Bandit learning model	Quality of service	Not defined	Not defined	Medium and high Fog nodes density	Android mobile data collected at downtown street in Vilanova I La Geltru
[43]	Edge user resource allocation issue	Random way point (RWP) mobility model OPT, FFP, and Greedy Approach	Quality of service	Central business district in Melbourne, Australia	Not defined	User coverage rate and capacity	Google map
[46]	Reliability of smart mobility applications	Not defined	Quality of service	Vehicular network	Python based modules	High density vs low density	Real VANET called BusNet

			Service of Quality		Build on simulation tool		
[44]	Service composition in mobile users and communities	Krill-Herd algorithm	Service of Quality	Mobile web service	Build on simulation tool using .Net platform	Population size, maximum iterations, searching constant	Not defined
[47]	Lack of mobility support and high delay	semi-definite programming based algorithm, Kuhn-Munkres algorithm	user experience, system performance	IIoT Healthcare smart grid smart traffic	Matlab	User-centric utility	Not defined
[48]	Service migration across mobile users	Lyapunov optimization, Markov approximation	Performance Time complexity	Not defined	ONE simulator	Average perceived latency, long term cost	Video streaming data
[49]	optimal task offloading decisions	Lyapunov optimization	Energy consumption, Quality of Experience	Vehicular Networks	Not defined	Transmission power, Noise power, Channel bandwidth, CPU frequency, Packet drop rate, Energy consumption	Not defined
[50]	Mobility-Aware Task Offloading issue in vehicular networks	Not defined	Performance Service Quality	Not defined	Vehicular network	System cost, Latency threshold,	Not defined
[51]	Delay and energy consumption during sending and receiving request from server	Weighted majority game theory	Latency Energy efficiency	Health care	Test bed is created to gather heart rate, Raspberry Pi	Data transmission, delay, energy	Real-time patient data (heart rate)

5 Mathematical Models for Mobility Based Resource Allocation

This section discusses various existing mathematical models that have been used to solve the mobility challenges in dynamic resource allocations.

Aazam et al. [52] proposed a mathematical model for the dynamic allocation of resources in fog computing and, in particular, in the industrial internet, to achieve both qualities of experience and quality of services. The net promoter score is used in the proposed model to calculate the user's input from a given scale with total scores of 0–10, divided into three parts, i.e. negative (0–6), neutral (7–8), and positive (9–10). By calculating these scores from the user's historic net promoter score, the user is assigned dynamic resources with different cases ranging from the default, smaller, and higher scores. Lu et al. [53] proposed an efficient mechanism to achieve a dynamic allocation of resources in the Fog-based high-speed train setup and to improve communication performance in the mobility environment. Also, to maximize the energy efficiency of the proposed mechanism, a mathematical model is proposed to solve the problem of rapid convergence using iterative algorithms. In the mathematical model, the following scenarios related to resource allocation problems are considered, such as subcarrier allocation, transmitting power, and antenna allocation. Also, each scenario is then resolved by decoupling the cases from the others in iterative ways.

Lee et al. [54] presented a mathematical model to solve resource problems in the Fog computing-based industrial internet of things. The purpose of presenting the mathematical model is to explain the relationship between computing cost and service popularity to solve the problem of constructing utilities for shared dynamic resources used in Fog computing. Babu and Biswash [55] presented a mathematical model for a mobility-based management technique designed to achieve node-to-node communication in Fog computing-based 5G networks. The following scenarios, including energy consumption, communication latency, robustness, signaling costs, delay, and latency, are considered in this model. Hui et al. [56] proposed an idea to build a resource allocation mechanism for the Edge-computing environment based on mobility to achieve better stability and secure system data operations. To formulate their proposed idea, a differential mathematical model is presented to define the relationship between the resources allocated to the intrusion detection system (IDS) and the users. To enhance the applicability of their model, the quantitative analysis is also combined with the proposed differential model. By having this integrated model, the system is better able to maintain scalability and security.

Xiang et al. [57] presented a mathematical model called JSNC for efficient slicing of mobile and Edge computing resources in the Fog-based Cloud computing environment. This mathematical modeling aims to analyze the problems for minimizing the latency or delay between the transmission of resources, performing operations on user traffic gather from multiple traffic classes, and applying the constraints on different capacities network. The proposed mathematical model is integrated

with a mixed-integer nonlinear spatial modulation to effectively evaluate the two heuristic approaches, including sequential fixing and greedy approaches. Oueida et al. [58] presented mathematical modeling using the Petri net framework for the validation of non-consumable resources in Cloud computing and Edge computing. This mathematical framework provides for the validation of the known framework used in the medical field called the Emergency Department (ED). For validation of the proposed model, the following parameters, such as patient length of stay, patient waiting time, and resource utilization rate, are used as basic performance measurement criteria. Zhang et al. [59] presented a mathematical model for the Joint Optimization Framework to optimize the content cache and resource allocation issues in MEC. The presented model is integrated with the policy gradient and value integration methods for determining the performance of communication links with two scenarios such as vehicle to infrastructure and vehicle to vehicle. At the top of that, the mathematical model is also given for the content of the cache scenario in MEC.

Huang et al. [60] proposed an energy-efficient, enhanced learning algorithm for task offloading and resource allocation in a Fog-based vehicular network. To validate the proposed algorithm, a mathematical model is then presented covering the different delay aspects, for instance, delay of task execution. Lin et al. [61] proposed a task offloading and resource allocation model based on a new multi-objective resource constraint mechanism for smart devices in a Fog-based Cloud environment. The proposed model is enhanced by a regression algorithm to make user requests without repeating the sequence. To validate their proof, a mathematical model is presented for the resolution of tasks related to offloading and resource allocation issues, taking into account different evaluation parameters, including processing time of virtual machines, tasks, completion time, and energy consumption.

The comparison of these models are presented in Table 3.

6 Application Use Cases

Fog computing is fully accessible to a wide variety of potential IoT applications, covering a significant proportion of many industries and businesses. It offers various advantages to those applications where real-time connectivity, streaming, fast processing, delay less communication, and low latency are the most and the top priority requirements. Figure 1 illustrates the taxonomy of mobility-based use cases and applications in the Fog/Edge environments.

6.1 Vehicular Networks

As part of the innovative IoT technologies, Vehicular Ad-hoc Networks (VANETs) offers an opportunity for the modern world by introducing new ways of linking

Table 3 Comparison of mobility-based resource allocation mathematical models

Ref	Application domain	Problems formulated	Model/technique used	Evaluation parameters considered
[52]	Industrial Internet of Things	Dynamic resource allocation	Net promoter score	Quality of service, Quality of experience
[53]	High-speed trains	Dynamic resource allocation, Communication performance	Mixed integer programming	Subcarrier allocation Transmit power Antenna allocation
[54]	Industrial Internet of Things	Resource portioning	Not defined	Quality of service, Quality of experience
[55]	5G networks	Mobility management	Net promoter score	Energy consumption, Communication delay, Robustness, Signalling cost, Delay, Latency
[56]	Intrusion detection system	Resource allocation	Differential Equation Model	Scalability
[57]	5G networks	Efficient slicing of resources	Not defined	Scalability
[58]	Smart Healthcare	Resource allocation	Petri net framework	Patient length stay, Patient waiting time, Resource utilization rate
[59]	Internet of Vehicles	Content cache and resource allocation	Markov decision process model	Scalability
[60]	Vehicular Networks	Task offloading and resource allocation	Deep Reinforcement Learning	Task execution delay, Energy consumption, Offloading execution time, Local execution time
[61]	Smart Healthcare	Task offloading and resource allocation	Fruit Fly Optimization	Processing time of virtual machines, Completion time, Energy consumption, Data center cost

on-road vehicles and passengers and offering new driving paradigms such as intelligent auxiliary driving and automatic driving [46]. Further, it enhance the existing transportation system with the new intelligent features including navigation, city traffic anomaly detection, bus stop arrival time estimation, and path finding along with the planning as well [62]. It also can detect and respond in real-time
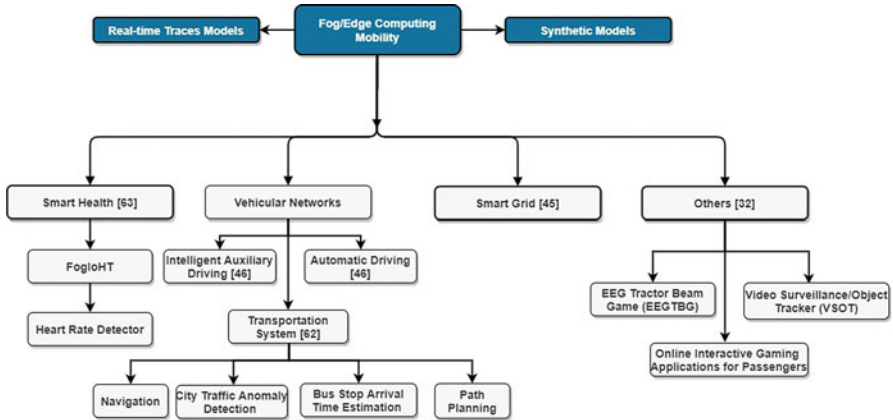
Fig. 1 Taxonomy of fog/edge mobility supported applications

to unforeseen incidents where traditional concepts such as Cloud computing are unable to perform immediate actions. Fog Computing considers VANETs to be a further prerequisite for the technological challenge, bringing computer tools, applications, and services closer to their users and enabling mobile applications and services to be delivered with a minimum of delay. To address the mobility challenges of smart applications in the VANET context, Pereira et al. [46] proposed a simplified architecture that uses proof-of-concept technologies to provide Fog computing mobility services across VANET safely and efficiently. The architecture tested on a smaller traffic dataset where the achieved results show the reliability and quality of the information broadcast over a short period.

In another study, Yang et al. [50] suggested a mobility-aware task offloading scheme to tackle the issues of computing, time selection, communication, high latency, and efficient resource allocation in VANETs. The proposed scheme uses the concept of MEC where each MEC server works independently or in conjunction with access points to efficiently perform mobility aware off-loading tasks. Besides, MEC servers also use the location-based off-loading system to unload tasks to adjacent mobility access points by moving vehicles. The benefit of the proposed scheme is that mobility vehicles based on initial locations may either select a local computing access point or reload their processing tasks to the next access points to achieve the quality of service and balance latency and network computation costs at an optimal stage.

6.2 Smart Healthcare

Smart healthcare is one of the most ambitious applications that combines current computing paradigms such as IoT, Fog, and Cloud to provide patients with an

enhanced and futuristic range of services. In smart healthcare, patient-related data is collected by different sensors, such as smartwatches, wrist bands, thermometers, and processed by different intermediate nodes, such as Fog, and further shared with Cloud servers, to take specific health-related actions. The network used for this form of application is called the body area network, where users (patients, doctors) can access their health data using various mobile devices and have separate control over data stored in the Cloud. To reduce the impact of energy consumption and processing delays of various Cloud servers in the smart healthcare environment, Fog computing has been proposed where various intermediate nodes called Fog nodes serve the processing and storage of patient data. To explain this scenario, Mukherjee et al. [63] suggested a Fog-based smart healthcare infrastructure where different indoor and outdoor sensors in the body area network gather patient data and send it to Fog nodes. Each Fog node used the concept of game theory called weighted majority to minimize the average latency, jitters, and energy consumption of the overall Fog computing-based health-care system.

6.3   Smart Grid

Smart grid is also a potential application of Fog-based IoT that combines the benefits of various ICT technologies to provide reliable, secure, and high-quality power services to consumers in an effective and specified manner. Smart grids are designed to monitor the power consumption of each household and to build a secure communication channel between users and different energy providers. Current solutions used by smart grid networks are solely focused on single or consolidated Cloud paradigms where Cloud providers are responsible for collecting and storing energy usage data and maintaining the profile of each household user. However, with an increasing number of smart devices connected to smart grid architectures, Cloud-based approaches often fail to provide real-time services on user data measurement and sometimes cause more delays in the network. In order to overcome these issues, Fog-based smart grid frameworks are introduced where Fog nodes are able to perform different computing services on user data rather than sending and storing Cloud parties [45].

6.4   Others

The scope of Fog and Edge computing paradigms is not limited to a few applications but provides a wide range of applications used in human daily life. For example, vision and hearing for mobility-impaired users, video surveillance, augmented reality, and mostly for gaming frameworks. One advantage of these Fog-based applications is that they require very low latency for communication between stations and Clouds [32].

7 Future Direction of Mobility-Based Resource Allocation and Provisioning in Fog and Edge related Computing Paradigms

This section concludes several future directions from the potential challenges of existing mobility-based resource allocation and provisioning frameworks for Fog and Edge-related computing paradigms. These challenges provide guidelines for researchers to develop efficient solutions for mobility-based resource allocation in Fog/Edge computing.

7.1 Mobility-Based Resource Allocation and Provisioning

Achieving the efficiency of Fog resources using different security enhancement policies is a difficult challenge, as each policy has its limitations, and therefore a multi-dimensional enhancement platform to support a variety of policies should be proposed to maximize profit and efficiency. In a mobility-based Fog computing environment, the energy consumption of mobile IoT devices is a challenging problem as most devices spend their energy and execution time connecting to different Fog devices in different regions, and therefore energy-efficient solutions for the placement and location of devices need to be considered as part of future work. The distribution of computational tasks on different Fog nodes in a resource-restricted environment is a difficult task, as Fog nodes must be aware of the computational capabilities of the IoT nodes deployed and their remaining use capacity. Also, the migration of services or applications from the Cloud to Fog nodes is a challenging issue that needs to be considered for efficient delivery and proper use to achieve QoS. Moreover, the available resources of Fog/Edge nodes should be properly managed through virtualization and efficiently allocate the resources because of limited resource nodes. To this end, efficient allocation, provisioning of resources, and scheduling algorithms should be developed to improve QoS in the Fog and Edge-based environment.

7.2 Security and Privacy

Despite offering a large number of benefits to users in terms of distributed processing, minimizes latency, mobility support, and position awareness, etc., the security and privacy of users and their data stored and exchanged by various Fog computing nodes is becoming a critical challenge. While a range of security solutions have been put forward to ensure the authentication, authorization, access control, and availability of Fog services, as well as the confidentiality, integrity, and reliability of data stored on Cloud servers. Nevertheless, due to a significant increase

in data volume and processing nodes, more robust and secure solutions based on quantum cryptography and blockchain could also be used to build a trustworthy relationship between Fog computing nodes.

7.3   Power Utilization and Management

Fog nodes have to manage a large number of concurrent requests from computers, users, and other Fog nodes simultaneously. Researchers have suggested various approaches to solve this situation, such as adding more Fog nodes and increase the resource vector that will ultimately accommodate multiple requests to some extent. However, on the other hand, these approaches contribute to the extra power usage of the overall network, which greatly reduces the energy and efficiency of restricted mobile devices. For the power utilization and management of both Fog and Edge nodes, this problem must be tackled to achieve the QoS of the underlying network and to maximize the efficiency of Fog nodes during the migration of services and tasks among other nodes.

7.4   Fault Tolerance

Fault tolerance is an essential challenge in the Fog computing environment that ensures the continuous delivery of services and operations to mobile nodes regardless of location and network processing efforts. This also ensures that every node completes their task in an event of breakdown with little to no human interference. To achieve fault tolerance of MC nodes, a range of failover and redundancy solutions such as RAID models, backup of user data, upgrade of security patches, constant power supply, etc. However, these solutions require extra hardware but instead provide Fog nodes services at fixed locations, so there is a need to improve capabilities and design low power fault tolerance techniques for the mobile nature of Fog nodes.

7.5   Support For Application Placements Strategies

Application placement strategies in Fog and Edge computing offer a way to meet the challenging needs of complex resources for an increasing number of IoT devices in time-critical scenarios. Since IoT devices operate constantly in the deployed environments for the sensing and computation of data, many Cloud/Fog resources are required to perform tasks in time. Current application placement approaches

often fail to satisfy the resource needs of a growing number of devices and only applicable to IoT applications that do not always change their locations, such as parking sensors. It is, therefore, a challenging task to develop an application placement strategy that needs support for the mobility and heterogeneous design of IoT devices and takes less time to perform parallel tasks.

7.6   Support Interoperability

The design model of Fog and Edge-based computing environments consists of heterogeneous nature of devices distributed at remote locations, connected via various data centers, and using a wide variety of protocols such as wireless, Bluetooth, 4G, and 5G. With the abundance of various types of technical components with specific design models and capacities, it is also difficult to work seamlessly and share resources among others. A proper interoperability framework is thus required that can manage the dynamic nature of devices and protocols, as well as fulfill the need for a common remote resource sharing platform to improve efficiency and transparency at a significant level.

7.7   Unified and Dynamic Resource Management and Provisioning

A crucial and difficult issue is to develop a single resource management framework, which is similar to personal computer Meta OS, to provide on-demand and cross-domain services for heterogeneous IoT devices. The platform should be a centralized IoT-based management framework that distinguishes hardware and software logically from devices and ensures that all software and hardware resources, including commodity OSes and their applications, are managed uniformly and that versatile services are provided via heterogeneous IoT devices. Required services can be configured on the IoT devices and performed dynamically according to the request. If the instance OS on the IoT device cannot support the requested service, the server-side will load a compatible instance OS for the service. So it is a crucial challenge in this regard to get dynamic OS boots and service loading on lightweight IoT devices from the Edge servers. While using the computing resources of Edge servers, IoT-based computing infrastructure will be capable of using a more powerful computing paradigm. The main challenge is how machine tasks are distributed between terminals and Edge servers. The development of various techniques for partitioning a task has been accelerated by evolving distributed computing environments to allow for simultaneous partitioned tasks at several geographic locations [17, 64].

8 Conclusion

Near-to-the-Edge services have now started delivering exciting services in every field with the help of nodes and devices in paradigms such as Fog and Edge. These nodes are highly dynamic and mobile, which pose several challenges in resource management including provisioning and allocation while serving the users for different applications. As such, in this chapter, we presented a comprehensive list of challenges for a MEC paradigm that can deliver seamless services, irrespective of its highly mobile nodes and devices. We also presented a true reflection of the current-state-of-the-art of the works done, centered around the mobility of the end devices, to address the challenges while managing the resource. Based on our analyses, we also highlighted the need for such a mobile environment to be integrated with emerging technologies like 5G and SDN along with other future research directions for MEC.

References

1. Sneha Tammishetty, T Ragunathan, Sudheer Kumar Battula, B Varsha Rani, P RaviBabu, RaghuRamReddy Nagireddy, Vedika Jorika, and V Maheshwar Reddy. Iot-based traffic signal control technique for helping emergency vehicles. In *Proceedings of the First International Conference on Computational Intelligence and Informatics*, pages 433–440. Springer, 2017.
2. KC Ujjwal, Saurabh Garg, James Hilton, Jagannath Aryal, and Nicholas Forbes-Smith. Cloud computing in natural hazard modeling systems: Current research trends and future directions. *International Journal of Disaster Risk Reduction*, page 101188, 2019.
3. Hamidreza Arasteh, Vahid Hosseinnezhad, Vincenzo Loia, Aurelio Tommasetti, Orlando Troisi, Miadreza Shafie-khah, and Pierluigi Siano. Iot-based smart cities: a survey. In *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, pages 1–6. IEEE, 2016.
4. Flavio Bonomi, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. Fog computing and its role in the internet of things. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pages 13–16, 2012.
5. Sudheer Kumar Battula, Saurabh Garg, James Montgomery, and Byeong Ho Kang. An efficient resource monitoring service for fog computing environments. *IEEE Transactions on Services Computing*, 2019.
6. Jürgo S Preden, Kalle Tammemäe, Axel Jantsch, Mairo Leier, Andri Riid, and Emine Calis. The benefits of self-awareness and attention in fog and mist computing. *Computer*, 48(7):37–45, 2015.
7. Ranesh Kumar Naha, Saurabh Garg, Dimitrios Georgakopoulos, Prem Prakash Jayaraman, Longxiang Gao, Yong Xiang, and Rajiv Ranjan. Fog computing: Survey of trends, architectures, requirements, and research directions. *IEEE access*, 6:47980–48009, 2018.
8. Sonia Shahzadi, Muddesar Iqbal, Tasos Dagiuklas, and Zia Ul Qayyum. Multi-access edge computing: open issues, challenges and future perspectives. *Journal of Cloud Computing*, 6(1):30, 2017.
9. Minh-Quang Tran, Duy Tai Nguyen, Van An Le, Duc Hai Nguyen, and Tran Vu Pham. Task placement on fog computing made efficient for iot application provision. *Wireless Communications and Mobile Computing*, 2019, 2019.
10. Maurizio Capra, Riccardo Peloso, Guido Masera, Massimo Ruo Roch, and Maurizio Martina. Edge computing: A survey on the hardware requirements in the internet of things world. *Future Internet*, 11(4):100, 2019.

11. Hasan Ali Khattak, Hafsa Arshad, Saif ul Islam, Ghufran Ahmed, Sohail Jabbar, Abdullahi Mohamud Sharif, and Shehzad Khalid. Utilization and load balancing in fog servers for health applications. *EURASIP Journal on Wireless Communications and Networking*, 2019(1):91, 2019.

12. Pavel Mach and Zdenek Becvar. Mobile edge computing: A survey on architecture and computation offloading. *IEEE Communications Surveys & Tutorials*, 19(3):1628–1656, 2017.

13. Yonal Kirsal, Glenford Mapp, and Fragkiskos Sardis. Using advanced handover and localization techniques for maintaining quality-of-service of mobile users in heterogeneous cloud-based environment. *Journal of Network and Systems Management*, 27(4):972–997, 2019.

14. Ranesh Kumar Naha, Saurabh Garg, Andrew Chan, and Sudheer Kumar Battula. Deadline-based dynamic resource allocation and provisioning algorithms in fog-cloud environment. *Future Generation Computer Systems*, 104:131–141, 2020.

15. Cheol-Ho Hong and Blesson Varghese. Resource management in fog/edge computing: a survey on architectures, infrastructure, and algorithms. *ACM Computing Surveys (CSUR)*, 52(5):1–37, 2019.

16. Mostafa Ghobaei-Arani, Alireza Souri, and Ali A Rahmanian. Resource management approaches in fog computing: A comprehensive review. *Journal of Grid Computing*, pages 1–42, 2019.

17. Ju Ren, Hui Guo, Chugui Xu, and Yaoxue Zhang. Serving at the edge: A scalable iot architecture based on transparent computing. *IEEE Network*, 31(5):96–105, 2017.

18. Haijun Zhang, Na Liu, Xiaoli Chu, Keping Long, Abdol-Hamid Aghvami, and Victor CM Leung. Network slicing based 5g and future mobile networks: mobility, resource management, and challenges. *IEEE communications magazine*, 55(8):138–145, 2017.

19. Argyrios G Tasiopoulos, Onur Ascigil, Ioannis Psaras, and George Pavlou. Edge-map: Auction markets for edge resource provisioning. In *2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 14–22. IEEE, 2018.

20. Mengting Liu, F Richard Yu, Yinglei Teng, Victor CM Leung, and Mei Song. Distributed resource allocation in blockchain-based video streaming systems with mobile edge computing. *IEEE Transactions on Wireless Communications*, 18(1):695–708, 2018.

21. Yangzhe Liao, Liqing Shou, Quan Yu, Qingsong Ai, and Quan Liu. Joint offloading decision and resource allocation for mobile edge computing enabled networks. *Computer Communications*, 2020.

22. Muhammad Waqas, Yong Niu, Manzoor Ahmed, Yong Li, Depeng Jin, and Zhu Han. Mobility-aware fog computing in dynamic environments: Understandings and implementation. *IEEE Access*, 7:38867–38879, 2018.

23. Shreya Ghosh, Jaydeep Das, and Soumya K Ghosh. Locator: A cloud-fog-enabled framework for facilitating efficient location based services. In *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 87–92. IEEE, 2020.

24. S Babu and Sanjay Kumar Biswash. Fog computing–based node-to-node communication and mobility management technique for 5g networks. *Transactions on Emerging Telecommunications Technologies*, 30(10):e3738, 2019.

25. Jindou Xie, Yunjian Jia, Zhengchuan Chen, and Liang Liang. Mobility-aware task parallel offloading for vehicle fog computing. In *International Conference on Artificial Intelligence for Communications and Networks*, pages 367–379. Springer, 2019.

26. Shashank Shekhar, Ajay Chhokra, Hongyang Sun, Aniruddha Gokhale, Abhishek Dubey, Xenofon Koutsoukos, and Gabor Karsai. Urmila: Dynamically trading-off fog and edge resources for performance and mobility-aware iot services. *Journal of Systems Architecture*, page 101710, 2020.

27. Dongyu Wang, Zhaolin Liu, Xiaoxiang Wang, and Yanwen Lan. Mobility-aware task offloading and migration schemes in fog computing networks. *IEEE Access*, 7:43356–43368, 2019.

28. John Paul Martin, A Kandasamy, and K Chandrasekaran. Mobility aware autonomic approach for the migration of application modules in fog computing environment. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–20, 2020.

29. Anwesha Mukherjee, Deepsubhra Guha Roy, and Debashis De. Mobility-aware task delegation model in mobile cloud computing. *The Journal of Supercomputing*, 75(1):314–339, 2019.
30. Shreya Ghosh, Anwesha Mukherjee, Soumya K Ghosh, and Rajkumar Buyya. Mobi-iost: mobility-aware cloud-fog-edge-iot collaborative framework for time-critical applications. *IEEE Transactions on Network Science and Engineering*, 2019.
31. José Santos, Tim Wauters, Bruno Volckaert, and Filip De Turck. Resource provisioning in fog computing: From theory to practice. *Sensors*, 19(10):2238, 2019.
32. Luiz F Bittencourt, Javier Diaz-Montes, Rajkumar Buyya, Omer F Rana, and Manish Parashar. Mobility-aware application scheduling in fog computing. *IEEE Cloud Computing*, 4(2):26–35, 2017.
33. Tarik Taleb, Konstantinos Samdanis, Badr Mada, Hannu Flinck, Sunny Dutta, and Dario Sabella. On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration. *IEEE Communications Surveys & Tutorials*, 19(3):1657–1681, 2017.
34. Jianbing Ni, Kuan Zhang, Xiaodong Lin, and Xuemin Sherman Shen. Securing fog computing for internet of things applications: Challenges and solutions. *IEEE Communications Surveys & Tutorials*, 20(1):601–628, 2017.
35. Rodrigo Roman, Javier Lopez, and Masahiro Mambo. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78:680–698, 2018.
36. Sathish Kumar Mani and Iyapparaja Meenakshisundaram. Improving quality-of-service in fog computing through efficient resource allocation. *Computational Intelligence*, 2020.
37. Yalan Wu, Jigang Wu, Long Chen, Gangqiang Zhou, and Jiaquan Yan. Fog computing model and efficient algorithms for directional vehicle mobility in vehicular network. *IEEE Transactions on Intelligent Transportation Systems*, 2020.
38. Min Chen, Wei Li, Giancarlo Fortino, Yixue Hao, Long Hu, and Iztok Humar. A dynamic service migration mechanism in edge cognitive computing. *ACM Transactions on Internet Technology (TOIT)*, 19(2):1–15, 2019.
39. Yuanguo Bi, Guangjie Han, Chuan Lin, Qingxu Deng, Lei Guo, and Fuliang Li. Mobility support for fog computing: An sdn approach. *IEEE Communications Magazine*, 56(5):53–59, 2018.
40. Fei Zhang, Guangming Liu, Bo Zhao, Xiaoming Fu, and Ramin Yahyapour. Reducing the network overhead of user mobility–induced virtual machine migration in mobile edge computing. *Software: Practice and Experience*, 49(4):673–693, 2019.
41. Juyong Lee, Daeyoub Kim, and Jihoon Lee. Zone-based multi-access edge computing scheme for user device mobility management. *Applied Sciences*, 9(11):2308, 2019.
42. Zeineb Rejiba, Xavier Masip-Bruin, and Eva Marin-Tordera. A user-centric mobility management scheme for high-density fog computing deployments. In *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–8. IEEE, 2019.
43. Qinglan Peng, Yunni Xia, Zeng Feng, Jia Lee, Chunrong Wu, Xin Luo, Wanbo Zheng, Hui Liu, Yidan Qin, and Peng Chen. Mobility-aware and migration-enabled online edge user allocation in mobile edge computing. In *2019 IEEE International Conference on Web Services (ICWS)*, pages 91–98. IEEE, 2019.
44. Hongyue Wu, Shuiguang Deng, Wei Li, Jianwei Yin, Xiaohong Li, Zhiyong Feng, and Albert Y Zomaya. Mobility-aware service selection in mobile edge computing systems. In *2019 IEEE International Conference on Web Services (ICWS)*, pages 201–208. IEEE, 2019.
45. Miodrag Forcan and Mirjana Maksimović. Cloud-fog-based approach for smart grid monitoring. *Simulation Modelling Practice and Theory*, 101:101988, 2020.
46. Jorge Pereira, Leandro Ricardo, Miguel Luís, Carlos Senna, and Susana Sargento. Assessing the reliability of fog computing for smart mobility applications in vanets. *Future Generation Computer Systems*, 94:317–332, 2019.
47. Shiyuan Tong, Yun Liu, Mohamed Cheriet, Michel Kadoch, and Bo Shen. Ucaa: User-centric user association and resource allocation in fog computing networks. *IEEE Access*, 8:10671–10685, 2020.

48. Tao Ouyang, Zhi Zhou, and Xu Chen. Follow me at the edge: Mobility-aware dynamic service placement for mobile edge computing. *IEEE Journal on Selected Areas in Communications*, 36(10):2333–2345, 2018.

49. Xiaoge Huang, Ke Xu, Chenbin Lai, Qianbin Chen, and Jie Zhang. Energy-efficient offloading decision-making for mobile edge computing in vehicular networks. *EURASIP Journal on Wireless Communications and Networking*, 2020(1):35, 2020.

50. Chao Yang, Yi Liu, Xin Chen, Weifeng Zhong, and Shengli Xie. Efficient mobility-aware task offloading for vehicular edge computing networks. *IEEE Access*, 7:26652–26664, 2019.

51. Anwesha Mukherjee, Debashis De, and Soumya K Ghosh. Fogioht: A weighted majority game theory based energy-efficient delay-sensitive fog network for internet of health things. *Internet of Things*, page 100181, 2020.

52. Mohammad Aazam, Khaled A Harras, and Sherali Zeadally. Fog computing for 5g tactile industrial internet of things: Qoe-aware resource allocation model. *IEEE Transactions on Industrial Informatics*, 15(5):3085–3092, 2019.

53. Lingyun Lu, Tian Wang, Wei Ni, Kai Li, and Bo Gao. Fog computing-assisted energy-efficient resource allocation for high-mobility mimo-ofdma networks. *Wireless Communications and Mobile Computing*, 2018, 2018.

54. Gaolei Li, Jun Wu, Jianhua Li, Kuan Wang, and Tianpeng Ye. Service popularity-based smart resources partitioning for fog computing-enabled industrial internet of things. *IEEE Transactions on Industrial Informatics*, 14(10):4702–4711, 2018.

55. S Babu and Sanjay Kumar Biswash. Fog computing–based node-to-node communication and mobility management technique for 5g networks. *Transactions on Emerging Telecommunications Technologies*, 30(10):e3738, 2019.

56. Hongwen Hui, Chengcheng Zhou, Xingshuo An, and Fuhong Lin. A new resource allocation mechanism for security of mobile edge computing system. *IEEE Access*, 7:116886–116899, 2019.

57. Bin Xiang, Jocelyne Elias, Fabio Martignon, and Elisabetta Di Nitto. Joint network slicing and mobile edge computing in 5g networks. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2019.

58. Soraia Oueida, Yehia Kotb, Moayad Aloqaily, Yaser Jararweh, and Thar Baker. An edge computing based smart healthcare framework for resource management. *Sensors*, 18(12):4307, 2018.

59. Mu Zhang, Song Wang, and Qing Gao. A joint optimization scheme of content caching and resource allocation for internet of vehicles in mobile edge computing. *Journal of Cloud Computing*, 9(1):1–12, 2020.

60. Xinyu Huang, Lijun He, and Wanyue Zhang. Vehicle speed aware computing task offloading and resource allocation based on multi-agent reinforcement learning in a vehicular edge computing network. *arXiv preprint arXiv:2008.06641*, 2020.

61. Kai Lin, Sameer Pankaj, and Di Wang. Task offloading and resource allocation for edge-of-things computing on smart healthcare systems. *Computers & Electrical Engineering*, 72:348–360, 2018.

62. Quan Yuan, Haibo Zhou, Jinglin Li, Zhihan Liu, Fangchun Yang, and Xuemin Sherman Shen. Toward efficient content delivery for automated driving services: An edge computing solution. *IEEE Network*, 32(1):80–86, 2018.

63. Anwesha Mukherjee, Debashis De, and Soumya K Ghosh. Fogioht: A weighted majority game theory based energy-efficient delay-sensitive fog network for internet of health things. *Internet of Things*, page 100181, 2020.

64. Yaoxue Zhang, Ju Ren, Jiagang Liu, Chugui Xu, Hui Guo, and Yaping Liu. A survey on emerging computing paradigms for big data. *Chinese Journal of Electronics*, 26(1):1–12, 2017.