High-Confidence Computing 4 (2024) 100193

Contents lists available at ScienceDirect

High-Confidence Computing

journal homepage: www.sciencedirect.com/journal/high-confidence-computing

Research Article

Fog-computing based mobility and resource management for resilient mobile networks

Hang Zhao, Shengling Wang *, Hongwei Shi

School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 25 June 2023
Revised 12 November 2023
Accepted 15 November 2023
Available online 24 November 2023

Keywords:
Resilient mobile networking
Fog computing
Mobility management
Resource management

ABSTRACT

Mobile networks are facing unprecedented challenges due to the traits of large scale, heterogeneity, and high mobility. Fortunately, the emergence of fog computing offers surprisingly perfect solutions considering the features of consumer proximity, wide-spread geographical distribution, and elastic resource sharing. In this paper, we propose a novel mobile networking framework based on fog computing which outperforms others in resilience. Our scheme is constituted of two parts: the personalized customization mobility management (MM) and the market-driven resource management (RM). The former provides a dynamically customized MM framework for any specific mobile node to optimize the handoff performance according to its traffic and mobility traits; the latter makes room for economic tussles to find out the competitive service providers offering a high level of service quality at sound prices. Synergistically, our proposed MM and RM schemes can holistically support a full-fledged resilient mobile network, which has been practically corroborated by numerical experiments.

1. Introduction

Nowadays, mobile networks encounter unprecedented challenges as they thrive since three features are emerging:

- *Large scale*. By 2022, there will be 1.5 mobile devices per capita [1]. The total number of global mobile subscribers will grow from 5.1 billion (66% of the population) in 2018 to 5.7 billion (71% of the population) by 2023 [2]. Solving the contradiction between large-scale users and limited network resources requires mobile networks to have the trait of scalability, handling a growing amount of parallel requests and demands within their acceptable time intervals, where traditional centralized strategy used cannot work [3].
- *Heterogeneity*. Content sharing has become the mainstream of network services instead of host communication, which includes a wide variety of low-bandwidth to high-bandwidth applications (from web browsing to virtual reality/augmented reality). Heterogeneous services ask mobile networks to have the great optimization ability for supporting real-time and non real-time applications to meet their different priority levels of resource sharing.
- *High mobility*. Connected car applications such as fleet management, in-vehicle entertainment systems, vehicle diagnostics, and navigation will be the fastest-growing category, at a 30 percent compound annual growth rate [1]. Such a high-mobility service requires that mobile networks can handle large-scale seamless handoffs, minimizing service disruptions due to the change of wireless access points at mobile nodes (MNs).

To tackle the above challenges, we propose a resilient mobile network framework based on fog computing. Fog computing is defined as an extension of cloud computing from the core to the physical devices at the network edge by Cisco. In fog computing, any devices or infrastructures that can provide computation, storage, and networking services are called *fog nodes* (FNs). Combining with the core cloud, FNs acting as *mini-clouds* along the cloud-to-thing continuum form a cloud-fog computing paradigm.

Introducing fog computing into mobile networks can pave the way towards a hyper-connected world because it has the following traits:

(1) *Consumer proximity*. This distinguishing trait makes all network provisions *localized*, implying that data does not need to move across the network. Such a concentration undoubtedly results in less congestion, latency, and jitter, supporting online and real-time applications in a cost-effective way, and satisfying the stringent QoS requirements induced by the dynamic traits of mobile applications.
(2) *Wide-spread geographical distribution*. On one hand, this trait endows fog computing to be location-aware and device-aware, letting each mobile/wireless device select the most appropriate way to access networks, solving the accessing problem of heterogeneous devices. On the other

---

[*] Corresponding author.
E-mail address: wangshengling@bnu.edu.cn (S. Wang).

https://doi.org/10.1016/j.hcc.2023.100193

hand, this trait makes any FN become a service provider, thus guaranteeing the scalability of networks.

(3) *Elastic resource sharing*. As an extension of the core cloud, it is natural for FNs to inherit the function of network virtualization. Taking advantage of network virtualization, FNs can share extra network resources (such as storage, computation) opportunistically to make up for the insufficiency of local resources [4]. Furthermore, it enables fog computing to be elastic enough to meet heterogeneous resource requirements of various applications.

Mobility management (MM) and resource management (RM) are two fundamental functions of mobile networks. The former answers for location and handoff management, while the latter is in charge of allocating and managing network resources; both count for the effectiveness and efficiency of mobile networks. Through integrating fog computing, MM and RM in our resilient mobile networking framework have the following traits:

- Personalized customization. The proposed MM provides a dynamically customized framework for any specific MN to select the best regional management agent (RMA) and the optimal regional size according to the MN's traffic and mobility traits. The aim of the personalized customization MM is to minimize the handoff delay and optimize the handoff-related transmission path.
- Market-driven resource allocation. The proposed RM makes room for economic tussles to find out that competitive service providers provide a high level of service quality at sound prices. The architecture of the market-driven RM framework can be pure distributed, pure central, or distributed-central mixed, implying it is general.

The proposed MM and RM do not work individually. They form a mutually-supporting resilient system to improve the performance of mobile networks in terms of handoff and transmission. In detail, the market-driven RM can help MM select the best RMA to satisfy the QoS requirements of MNs, while the personalized customization MM can assist RM in improving transmission performance through optimizing handoff-related transmission paths.

The rest of the article is organized as follows. The next section introduces the state-of-the-art work on MM and RM. The personalized customization MM and the market-driven RM are proposed in Sections 3 and 4 respectively. Section 5 shows the performance analysis by numerical results. We conclude our paper in Section 6.

2. Related work

In this section, the state-of-the-art work on MM and RM is introduced.

2.1. Mobility management

MM aims to offer seamless services for an MN with minimum network disconnection time caused by handoffs. According to whether there is a need to modify the protocol stacks of MNs, MM protocols can be categorized into two kinds: host-based and network-based.

2.1.1. Host-based mobility management schemes

In host-based MM schemes, to support continuous communication during mobility, the protocol stacks of MNs are required to be modified to support MM signaling processes consisting of movement detection, route solicitation, duplicate address detection, and location updating. Pioneering work on host-based MM schemes includes Internet Engineering Task Force (IETF) standards, mobile IPv4 (MIPv4) (RFC 3344), MIPv6 (RFC 3775), MIPv4 regional registration (RFC 4857), and hierarchical MIPv6 (HMIPv6) (RFC 4140).

MIPv4 and MIPv6 serve for IPv4 and IPv6 networks respectively. They have common idea in their solutions: a home agent (HA) is deployed in the network to bind an MN's identifier with locator. The former refers to an IPv4/IPv6 address indicating the MN's identification while the latter is an IPv4/IPv6 address denoting its location. Once the MN moves from one sub-network to another, it is required to register the new locator to the HA. Hence, when the MN moves far from the HA, the delay of registering the HA prolongs, increasing the handover latency. This situation will worsen if the MN performs handovers frequently.

To make up for the deficiencies of MIPv4 and MIPv6, the method of hierarchizing networks was proposed. Hence, the whole network consists of multiple regions, resulting in two kinds of handoffs: macro-mobility (handoffs across regions) and micro-mobility (handoffs within a region). Here, MIPv4 and MIPv6 are employed to deal with macro-mobility whereas some specific micro-mobility schemes, such as MIPv4 regional registration and HMIPv6, are employed to cope with micro-mobility. When an MN performs handoffs within a region, it needs to register the regional mobility agent rather than the HA, thus reducing the handoff latency of micro-mobility.

Although micro-mobility schemes (e.g., HIMPv6) were designed to enhance the performance of macro-mobility schemes (e.g., MIPv6) through hierarchizing networks, Wang et al. [5] do not think HMIPv6 can always outperform MIPv6. To figure out whether to hierarchize a network and how to hierarchize it if necessary, they developed a model to analyze the application scopes of MIPv6 and HMIPv6, based on which an optimal choice of mobility management (OCMM) scheme was proposed. To minimize handoff latency and packet delivery latency, OCMM makes customized MM schemes for each MN. In detail, OCMM selects the best MM schemes between MIPv6 and HMIPv6 in light of the mobility and service characteristics of an MN, addressing whether to hierarchize the network. Further, OCMM chooses the best mobility anchor point and the optimal regional size when HMIPv6 is adopted, addressing how to hierarchize the network.

Another approach proposed to reduce the handover latency is Fast handover for Mobile IPv6 (FMIPv6) [6]. Using FMIPv6, an MN may send packets as soon as it recognizes a new subnet link, and a new access router can deliver packets to a mobile node as soon as it recognizes the node's attachment.

In addition, centralized MM has obvious drawbacks such as performance bottlenecks and single points of failure, therefore distributed MM has been considered. Lee et al. [7] proposed a host-based distributed MM protocol and experimentally confirmed that the approach can further improve the switching latency and throughput of the network when facing the challenges of flattening and high traffic in mobile networks. In [8], an efficient security protocol for binding updates was proposed, expanding host-based mobility management solutions, in which, without the MN's assistance, the security of the mobility management is designed to gain possible benefits such as mobility localization and easy extension.

2.1.2. Network-based mobility management schemes

Host-based MM schemes require MNs to install mobility functionality, greatly limiting their application scope, especially in the era of IoT, where most devices have weak capacities (e.g., storage and computation). To solve this problem, network-based MM schemes aim to provide transparent mobile connectivity to MNs, meaning that it is unnecessary to implement additional software stacks for MNs.

Proxy mobile IPv6 (PMIPv6) is a typical network-based MM scheme that is standardized by the IETF in RFC 5213. PMIPv6 deploys a local mobility anchor and a mobile access gateway to track the movements of MNs and perform the required mobility signaling on behalf of MNs. Since there is no mobility-related signaling, PMIPv6 can minimize the signaling overhead of MNs and is more appropriate for IoT devices. Based on PMIPv6, Berguiga et al. proposed an improved Fast Handover Proxy Mobile IPv6 for Sensor Networks, which effectively reduces packet loss and improves network throughput [9].

Considering that the software-defined networking (SDN) framework is an efficient solution to the growing challenges and requirements of future 4G and 5G networks, Jia proposed a route optimized PMIPv6 virtualized function over the SDN framework [10], where PMIPv6 mechanism is transformed into a suit of virtualized network functions and implemented for SDN-based evolved packet core networks. [11] only focused on the handoff management of MM, and proposed a simple handoff scheme according to the speed of MNs to lower handoff latency for heterogeneous cloud small cell networks. In [12], SDN-MM avoids packet loss and tunneling costs, giving mobile consumers better QoS through separating the duties of packet forwarding and mobility management. HSD-DMM [13] as a distributed scheme that uses dynamic anchor selection to reduce the overhead of signaling and packet delivery, address the shortcomings of centralized protocols, and enhance the flexibility of larger, geographically dispersed networks.

2.2. Resource management

As a distributed computing paradigm, fog computing is widely used in IoT applications to ensure the real-time QoS requirements of mobile users with limited data processing speed and bandwidth. Different from cloud computing, because an intermediate layer composed of fog nodes is added between cloud services and mobile devices, it faces more challenging resource management problems that can be solved in the two dimensions of space and time.

From the perspective of space, the most critical problem is a matching problem: it is necessary to find the optimal mapping between cloud services and fog nodes, fog nodes and mobile nodes, as well as IoT applications and fog nodes, to minimize latency and improve QoS. For example, Zhang et al. proposed a three-layer hierarchical game framework based on game theory to solve the resource allocation among data service operators (SAO) FNs and data service subscribers (ADSSs), including the virtualized network and the asymmetric information problem, as well as the resource matching from the FNs to the ADSSs in the physical network [14]. Nguyen et al. proposed a novel market-based resource allocation framework to solve how to allocate multiple resource types of capacity-limited heterogeneous fog nodes to competing services in which fairness, efficiency, and privacy-preserving are considered [15]. A decentralized multi-SP resource allocation scheme is proposed in [16], where an improved matching algorithm is designed to solve the total profit maximization problem in resource allocation.

In the fog computing system, the trait of consumer proximity allows users to offload part of their data or computational tasks to the fog nodes. The following problem is to balance energy consumption and delay performance. For example, Liu et al. utilized queueing theory to design cost models for the MDs to find the optimal offloading probability and transmit power for each MD [17]. To make full use of nearby distributed FN nodes, load balancing is also an important goal in resource management. In [18], Huang et al. proposed a candidate FN-based resource allocation algorithm which is an energy-saving and load-balancing allocation scheme. For fog computing environments, a reconfigurable resource management framework [19] is proposed that allows users to offload tasks and perform communications. By integrating integer linear programming formulas and heuristics, the scheme reduces the energy consumption of the system and improves the fairness of bandwidth usage. FN nodes are always equipped with computing and storage capabilities, which enable them to perceive information in the network and optimize resource allocation to ensure an effective and efficient resource management framework. For example, In [20], Lu et al. provided an efficient resource provisioning scheme in fog computing to minimize the total cost for multiple mobile users.

In the time dimension, the main problem is how to schedule the tasks of a set of IoT services to ensure the quality of service in the fog node network. In [21], the authors constructed a task-scheduling algorithm for the fog node to optimize the number of concurrent tasks and ensure that the tasks can be completed on time. In [22] proposed a resource allocation and management technique TRAM that provided a scheduling algorithm for the resource hierarchy process in fog computing environments, effectively reducing the execution time, network consumption, energy consumption, and average cycle latency of tasks. Wadhwa et al. [23] proposed the optimized task scheduling and preemption (OSCAR) model to overcome resource management and throughput transactions caused by large data transfers and experimentally showed that the proposed scheme effectively improves the QoS of the network.

This paper diverges from the conventional researches in network optimization, where MM and RM have traditionally been explored in isolation. While MM and RM have individually been extensively investigated for their roles in improving network performance, our work introduces a collaborative scheme that integrates both approaches. Unlike previous studies, our novel perspective emphasizes a resilient system framework for optimizing networks. By intertwining MM and RM functionalities, our schema not only enhances traditional performance metrics but also fortifies networks against unforeseen crisis. This distinctive approach sets our work apart from existing solutions, opening new avenues for research in the pursuit of more robust and adaptable networks.

3. Personalized customization MM

In this section, we will introduce the framework of the personalized customization MM with the solution of key factor and its usage in the handoff process.

3.1. Framework of the proposed MM

From the viewpoint of network structure, MM mechanisms can be categorized into two types: central and distributed. In the central MM, there is only one management agent, i.e., the home agent (HA), deployed to bind the identifiers and locators of all MNs. Once an MN moves from one access network to another, a binding updating (BU) process should be executed through HA. In the distributed MM, the whole network is divided into several regions, where each region deploys one RMA to manage the mobility of all MNs in that region. Therefore, handoffs are also classified into two kinds: intra-region handoffs and inter-region ones. When an intra-region handoff happens, only the local binding updating (LBU) process to RMA is needed, while an inter-region handoff should incur LBU and BU processes to both the new RMA and HA, respectively. Thus, by decreasing the frequency of BUs to HA, the distributed MM improves MNs' handoff experiences.

Fog computing is naturally suitable for developing distributed MM due to its trait of geographical distribution. Hence, we place

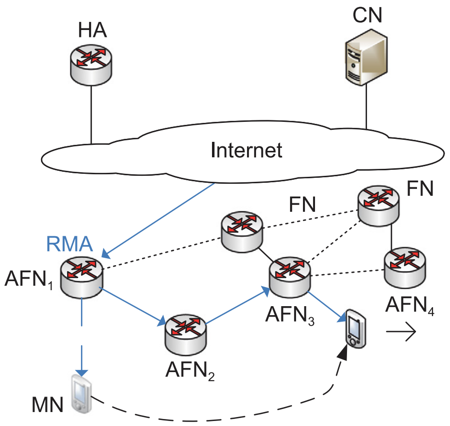

Fig. 1. Framework of the proposed MM.

the distributed MM on the fabric of fog computing, whose framework is shown in Fig. 1, where any FN that can provide wireless access is called an access FN (AFN). Since FNs usually have storage and computation capacities, each AFN is required to act as an RMA when it has enough network resources, implying it should record the identifiers and locators of MNs if necessary.

3.2. Optimal regional size

Obviously, the proposed MM framework based on fog computing can inherit the traits of distributed MM. As we described above, the distributed MM outperforms the central one in terms of MNs' handoff experiences. However, this advantage is obtained at the cost of data delivery performance. In detail, the data sent by a corresponding node (CN) has to arrive at the RMA first and then tunnel to the MN, incurring the *triangle routing* issue. For example, in Fig. 1, when an MN roams from AFN_2 to AFN_3, it only notifies its current location (i.e., AFN_3) to the RMA (i.e., AFN_1), which shortens the process of handoff. Nevertheless, CN has to deliver data through the RMA to the MN rather than directly sending data to AFN_3 since it does not know the new location of the MN. This undoubtedly lengthens the data delivery delay. Since the data delivery cost is directly related to the tunnel length from the MN's current AFN to the RMA, we use the hops from the RMA to the farthest AFN in the region to represent the size of the region (s). Thus, to describe the total cost (C) of the distributed MM based on the fabric of fog computing, we use the following metric:

$$C(\upsilon, l, s) = \alpha C_H(\upsilon, s) + \beta C_T(l, s) \tag{1}$$

In (1), C_H and C_T are respectively the expected handoff latency cost and data delivery cost. $\alpha > 0$ and $\beta > 0$ are the coefficients, indicating the degree of influence of the two costs on the total cost C. The larger α means handoff has greater impact on C and otherwise, data delivery matters. The regional size s influences both C_T and C_H. Specifically, the larger the size is, the higher the data delivery cost is, and vice versa. υ and l are the MN's mobility rate and data arrival rate, separately affecting the expected handoff latency cost and data delivery cost.

The concept of *region* is introduced by the distributed MM to shorten the handoff delay of the centralized MM. Hence, the expected handoff latency cost $C_H(\upsilon, s)$ is determined by comparing the expected handoff latencies of these two kinds of MM. That is,

$$C_H(\upsilon, s) = \frac{(M(s) - 1)D_{intra} + D_{inter} - M(s)D_{inter}}{T(s, \upsilon)} \tag{2}$$

where $M(s)$ is the handoff times of the MN within an RMA region, which is obviously related to the regional size s. In the distributed MM, there are $M(s)-1$ intra-region handoffs plus one inter-region handoff, while in the central MM, all handoffs belong to the inter-region ones. D_{intra} and D_{inter} are respectively the intra-region and inter-region handoff latencies. $T(s, \upsilon)$ is the average time the MN resides in the RMA region, which is correlated to the regional size s and the MN's mobility rate υ. Specifically, $T(s, \upsilon) = (M(s)+1)/\upsilon$. Thus, $((M(s) - 1)D_{intra} + D_{inter})/T(s, \upsilon)$ and $M(s)D_{inter}/T(s, \upsilon)$ are respectively the expected handoff delay latencies. Due to $D_{intra} \leq D_{inter}$, $C_H(\upsilon, s)$ is non-positive, implying it is actually the handoff profit of the distributed MM compared with the central one.

In the distributed MM, all data sent to the MN should be tunneled from the RMA to the current AFN, which is the extra cost of the distributed MM compared with the central one. Hence,

$$C_T(l, s) = l \cdot \tilde{s} \tag{3}$$

where $\tilde{s}$ is the average tunnel length when the regional size is s.

Among the three factors affecting the total cost in (1), υ and l pertain to the MN side, indicating its mobility and service traits; and s belongs to the network side, reflecting how the network divides MM regions. Because the division of regions is completely logical rather than physical, we can provide the personalized optimal region (s^*) for each MN in light of its mobility and service traits. That is

$$s^* = \arg_s \min C(\upsilon, l, s) \tag{4}$$

Because the increase of regional size will reduce C_H and increase C_T, there always exists an optimal regional size (s^*) to minimize C, that is, the optimal length of tunnel between the RMA and the farthest AFN in the region.

3.3. Handoff procedure of the proposed MM

The handoff procedure of the personalized customization MM is depicted in Fig. 2:

(1) After the MN receives the announcement from a new AFN and determines to perform a handoff, it will first check whether it moves across the region boundary by judging whether the tunnel length between the RMA and the new AFN n is longer than s^*. If not, the intra-region handoff occurs, where the MN only needs to send the LBU to the old RMA informing of the new AFN. After binding the MN's current location, RMA acknowledges the MN with the tunnel length n to the MN, which is updated according to the routing information.

(2) if $n > s^*$, the MN needs to further determine whether the new AFN belongs to its RMA candidate set. If not, the intra-region handoff described in the previous step incurs.

(3) When $n > s^*$ and the new AFN belongs to its own RMA candidate set, MN sets the current AFN as the new RMA and initiates location registration with the new RMA; After binding the identifier and locator of MN, the new RMA will update the optimal region size (s^*) according to Eq.(4) and reset $n = 0$, which will be sent to the MN with an ACK. Once receiving the ACK, MN initiates a deleted binding request to the old RMA and a location update to the HA.

The RMA candidate set of the MN is a set of RMA which can satisfy the QoS requirements of the MN. The RMA is responsible for tunneling data for all MNs in the region, so whether the RMA's resources are sufficient is very critical. Therefore, in the above process, when $n > s^*$, if the newly accessed AFN is not in the RMA candidate set, the MN will still maintain the original RMA instead of changing it.

After the MN completes the registration with the HA, the HA will find a list of RMA candidates for it according to the current

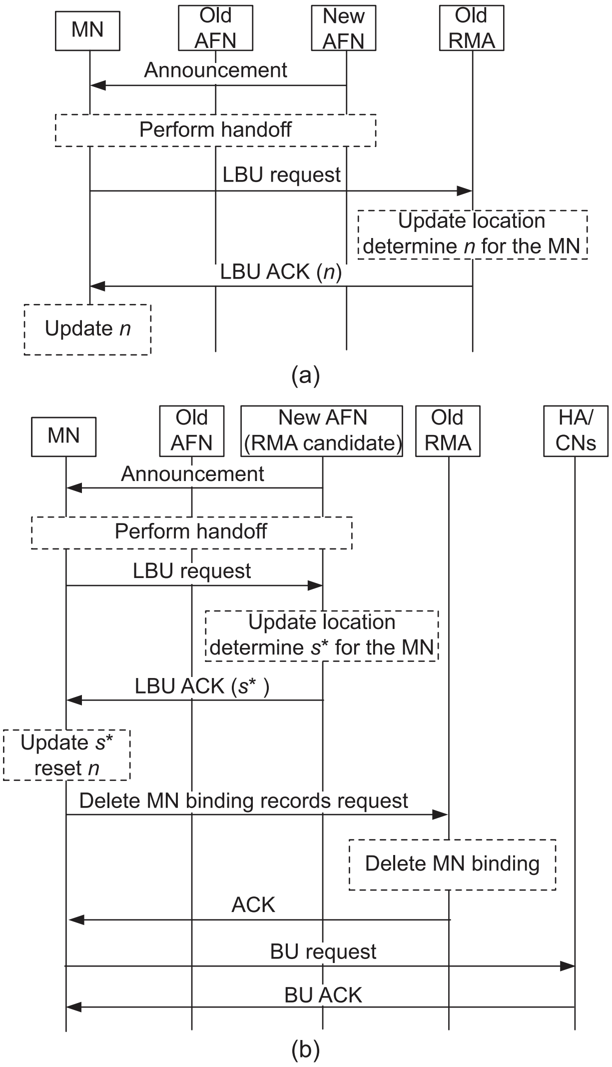

Fig. 2. Handoff procedure of the proposed MM. (a) Flowchart of intra-region handoff. (b) Flowchart of inter-region handoff.

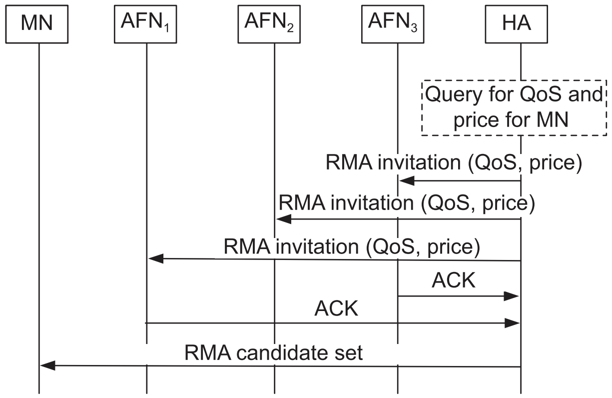

Fig. 3. Selection of RMA candidates for the MN.

RMA location to meet the demand for replacing RMA when the MN roams out of the current region. The specific steps are shown in Fig. 3:

(1) Once the HA returns BU ACK to the MN, it will be triggered to select the potential RMA in the next region for the MN. To that aim, the HA will query for the QoS requirements and corresponding price negotiated with the MN before[1];
(2) The HA broadcasts *RMA invitation req* containing the MN's QoS requirements and price to each AFN nearby MN's current region for asking whether they are capable of serving as an RMA candidate for the MN;
(3) Any AFN receiving the invitation will response whether it is resource-qualified as an RMA;
(4) If the HA does not receive the respond from an AFN within a given time duration, it will consider that the AFN cannot have the capability;
(5) In light of the responses given by the AFNs, HA sends the list of AFN candidates to the MN. Hence, if the MN roams among these AFNs, one of them will act as its RMA if inter-region handoff happens.

When $n > s^*$, if the MN finds that the current AFN does not belong to the RMA candidate set, it will actively initiate an RMA invitation request to the HA containing its current AFN address. The HA will select qualified RMA candidates nearby for the MN according to Steps (2)-(5) of the above process.

4. Market-driven RM

As shown in Fig. 3, each AFN needs to judge whether it is resource-qualified to act as an RMA candidate for an MN. This process is in charge of RM. The traits of *large scale* and *heterogeneity* of mobile networks require RM to optimize resource allocation to solve the contradiction between limited network resources and large-scale users with widely varied QoS requirements. To tackle this challenge, a market-driven RM based on fog-computing framework is proposed.

Taking advantage of network virtualization, FNs can share network resources opportunistically to make up for the insufficiency of local resources. Thus, a network ecosystem is formed with a variety of resource providers and consumers, where the law of natural selection and survival of the fittest are also applicable to yield optimal RM. That needs a mechanism for beneficial tussles among all FNs in the above network ecosystem. Since the behaviors of any rational and intelligent individuals are profit-driven, the above tussles originate from their economic relationship.

Our RM makes room for economic tussles to find out the competitive resource providers for offering a high level of service quality at sound prices, whose framework is illustrated by Fig. 4, where each FN or cloud server has three functional planes, namely *market plane*, *orchestration plane* and *provision plane*. To exchange market information, FNs or cloud servers can interact with each other or the central market server, implying our framework is general to support three structures: pure distribution, pure centrality, and mixed distribution and centrality. The interfaces of resource pools among FNs or cloud servers enable network resources to be borrowed and lent among FNs and cloud servers. The function of each plane is detailed as follows.

4.1. Market plane

As described in Section 3, an MN can negotiate with the HA in terms of the QoS provided by the network. When the resources of the involved network node are insufficient, it can purchase from other nodes. To achieve that aim, the market plane needs to find an optimal provider or the optimal combination of providers for the clients from a great spectrum of choices. Suitable economic models should be built to deduce a

---

[1] The MN can negotiate with the HA on QoS and price at any time, which will be executed when it roams to the next new RMA.

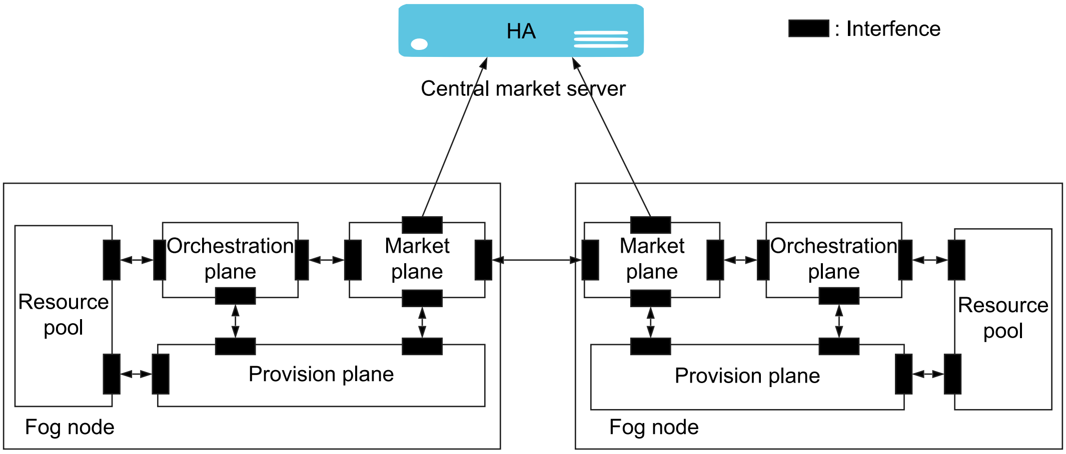

Fig. 4. Market-driven RM framework.

set of Pareto-optimal provider combinations. The models based on game theory are good choices because game theory can efficiently deal with the economic relationship to solve the optimal strategies for marketing behavior.

4.2. Orchestration plane

The orchestration plane has two functions: managing the resource pool and determining the contract with its clients, who can be other peers or MNs.

The total amount of resources in the resource pool, no matter whether they are real (the local) or virtual (the external) ones borrowed from other providers employing virtualization technologies, should be divided into three parts: the one which is being used, the unused part, and the reservation part. The reservation part is employed to deal with the resource consumption as a candidate RMA as well as the bursty and unexpected resource demand for guaranteeing the quality of service. The orchestration plane needs to develop some algorithm to optimize the amount of reservation resource to avoid *over-reservation* and *under-reservation*, the former leading to a low resource utilization ratio while the latter resulting in bad customers' satisfactions.

Although the market plane outputs a set of Pareto-optimal resource transaction schemes taking into account all possible economic factors, the final contracts are determined in the orchestration plane, which selects the best resource transaction scheme from the output of the market place based on some non-economic factors. For example, if there is a need to borrow resources from other providers, the orchestration plane will select the one(s) with a good reputation or high degree of trust based on previous cooperations.

4.3. Provision plane

The provision plane is an intermediate layer to connect the upper planes and customers. On the one hand, it reports resource demands and contract violation information to the market and orchestration planes. On the other hand, it produces resource instances according to the contracts determined by the orchestration plane, thus serving for customers. Due to the nature of user proximity, the provision plane should deal with privacy and security issues. In detail, the provision plane has sensitive information concerning customers' accounts, contracts, the changing rules of resource usage over time, and so on. Once such sensitive information is released, the customers' economic status and habits can be analyzed and then revealed. In addition, some attacks, such as denial of service and spoof address of mobile or wireless devices, can also easily happen in this plane. All these privacy and security issues call for smart solutions and countermeasures in the provision plane.

5. Numerical results

In this section, we analyze how the optimal regional size varies with the mobility rate and data arrival rate.

In our numerical analysis, to prevent the negligence of one type of cost due to the magnitude differences between $C_H(v, s)$ and $C_T(l, s)$, we have set α and β to 1 and 0.01, respectively. These values can also be finely designed according to the attributes of the task in practice. According to [5], the range of l is set to $[0.1, 2.1]$ and the range of v is $[0.01, 0.05]$. For simplicity, we employ the one-dimensional random walk model to describe the mobility of the MN. Thus, the expected number of handoffs (i.e., $M(s)$ in (3)) needed by the MN to move out of a region is proportional to the square of the region size s [24]. The average tunnel length is set to $\tilde{s} = s/2$. D_{intra} and D_{inter} are set to 4 and 16 respectively [25].

Fig. 5 shows that the optimal regional size increases with the mobility rate but decreases with the data arrival rate. The reason behind this fact is that the increase of mobility rate makes the number of handoffs rise. At this time, expanding the optimal regional size will reduce the number of inter-region handoffs, obtaining more profit than the central MM in terms of handoff performance; with the rise of the data arrival rate, the extra delay of tunneling data prolongs with the tunnel length. In this case, shortening the tunnel length, implying that reducing the optimal regional size, can decrease the data transmission cost.

6. Conclusion

Fog computing is an emerging paradigm which is suitable to solve the challenges of mobile networks as they thrive. In this article, we propose a novel mobile networking framework to achieve resilience, including the personalized customization MM and the market-driven RM. They form a mutually-supporting system to improve the resilience of mobile networks in terms of handoff and transmission. The market-driven RM can help MM select the best regional management agent to satisfy the QoS requirements of a mobile node when its data is tunneled, while the personalized customization MM can assist RM in improving transmission performance through optimizing handoff-related transmission paths by solving the optimal regional size. Numerical experiments show that our proposed method is practical and can improve the resilience of the network according to the optimal regional size for different the mobility rate and data arrival rate.

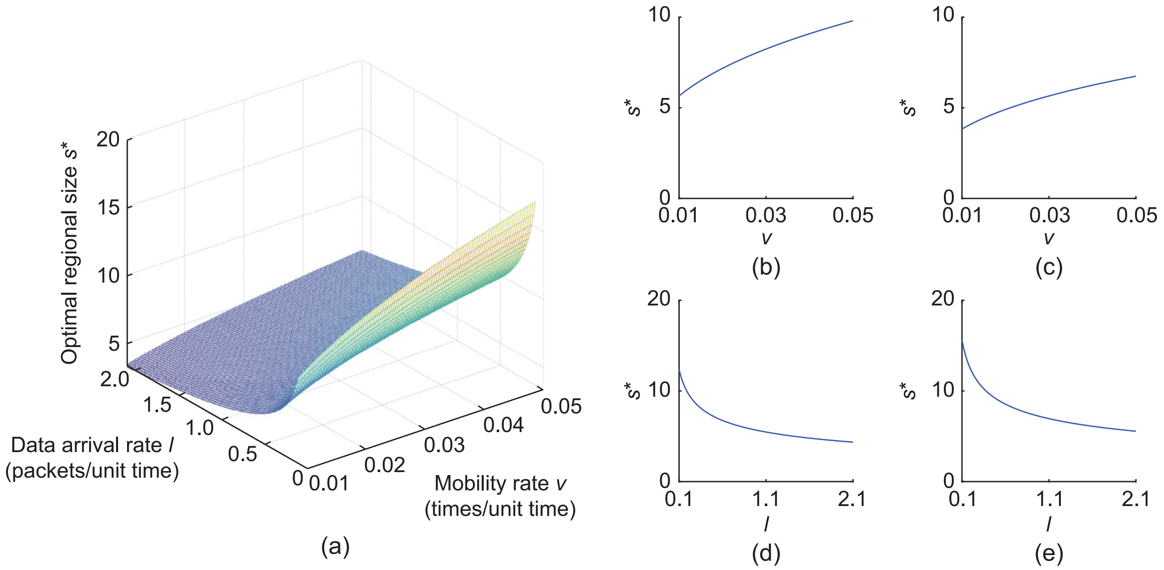

Fig. 5. Influence of the mobility rate and the data arrival rate on the optimal regional size. (b) $l = 0.5$. (c) $l = 1.5$. (d) $l = 0.02$. (e) $l = 0.04$.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (61772044, 62077044, and 62293555), the Major Program of Science and Technology Innovation 2030 of China (2022ZD0117105), and the Major Program of Natural Science Research Foundation of Anhui Provincial Education Department, China (2022AH040148).

References

[1] V. Cisco, Cisco visual networking index: Forecast and trends, 2017–2022, in: White Paper 1, 2018.

[2] U. Cisco, Cisco Annual Internet Report (2018–2023) White Paper, 2020, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/whitepaper-c11-741490.html. [Online](Accessed 26 March 2021).

[3] Z. Cai, T. Shi, Distributed query processing in the edge-assisted iot data monitoring system, IEEE Internet Things J. 8 (2020) 12679–12693.

[4] T. Shi, Z. Cai, Y. Li, Query recombination: To process a large number of concurrent top-k queries towards iot data on an edge server, in: 2022 IEEE 42nd International Conference on Distributed Computing Systems, ICDCS, IEEE, 2022, pp. 559–569.

[5] S. Wang, Y. Cui, S.K. Das, W. Li, J. Wu, Mobility in ipv6: Whether and how to hierarchize the network? IEEE Trans. Parallel Distrib. Syst. 22 (2011) 1722–1729.

[6] R. Koodli, Fast Handovers for Mobile IPv6, Technical Report, 2005.

[7] J.-H. Lee, J.-M. Bonnin, X. Lagrange, Host-based distributed mobility management support protocol for ipv6 mobile networks, in: 2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob, IEEE, 2012, pp. 61–68.

[8] S. Lee, H.-K. Choi, E. Kim, J.-H. Lee, Secure and efficient binding updates in host-based distributed mobility management, Wirel. Netw. 25 (2019) 1443–1459.

[9] A. Berguiga, A. Harchay, A. Massaoudi, H. Youssef, Fpmipv6-s: A new network-based mobility management scheme for 6lowpan, Internet Things 13 (2021) 100045.

[10] W.-K. Jia, Architectural design of an optimal routed network-based mobility management function for SDN-based epc networks, in: Proceedings of the 11th ACM Symposium on QoS and Security for Wireless and Mobile Networks, ACM, 2015, pp. 67–74.

[11] H. Zhang, C. Jiang, J. Cheng, V.C. Leung, Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks, IEEE Wirel. Commun. 22 (2015) 92–99.

[12] Y. Bi, G. Han, C. Lin, M. Guizani, X. Wang, Mobility management for intro/inter domain handover in software-defined networks, IEEE J. Sel. Areas Commun. 37 (2019) 1739–1754.

[13] E. Amiri, N. Wang, S. Vural, R. Tafazolli, HSD-DMM: Hierarchical software defined distributed mobility management, in: 2021 IEEE 20th International Symposium on Network Computing and Applications, NCA, IEEE, 2021, pp. 1–7.

[14] H. Zhang, Y. Zhang, Y. Gu, D. Niyato, Z. Han, A hierarchical game framework for resource management in fog computing, IEEE Commun. Mag. 55 (2017) 52–57.

[15] D.T. Nguyen, L.B. Le, V.K. Bhargava, A market-based framework for multi-resource allocation in fog computing, IEEE/ACM Trans. Netw. 27 (2019) 1151–1164.

[16] C. Zhang, H. Du, Q. Ye, C. Liu, H. Yuan, Dmra: A decentralized resource allocation scheme for multi-sp mobile edge computing, in: 2019 IEEE 39th International Conference on Distributed Computing Systems, ICDCS, IEEE, 2019, pp. 390–398.

[17] L. Liu, Z. Chang, X. Guo, S. Mao, T. Ristaniemi, Multiobjective optimization for computation offloading in fog computing, IEEE Internet Things J. 5 (2017) 283–294.

[18] X. Huang, W. Fan, Q. Chen, J. Zhang, Energy-efficient resource allocation in fog computing networks with the candidate mechanism, IEEE Internet Things J. 7 (2020) 8502–8512.

[19] N. Godinho, H. Silva, M. Curado, L. Paquete, A reconfigurable resource management framework for fog environments, Future Gener. Comput. Syst. 133 (2022) 124–140.

[20] S. Lu, J. Wu, Y. Duan, N. Wang, J. Fang, Towards cost-efficient resource provisioning with multiple mobile users in fog computing, J. Parallel Distrib. Comput. 146 (2020) 96–106.

[21] L. Yin, J. Luo, H. Luo, Tasks scheduling and resource allocation in fog computing based on containers for smart manufacturing, IEEE Trans. Ind. Inform. 14 (2018) 4712–4721.

[22] H. Wadhwa, R. Aron, Tram: Technique for resource allocation and management in fog computing environment, J. Supercomput. 78 (2022) 667–690.

[23] H. Wadhwa, R. Aron, Optimized task scheduling and preemption for distributed resource management in fog-assisted iot environment, J. Supercomput. 79 (2023) 2212–2250.

[24] T. Antal, S. Redner, The excited random walk in one dimension, J. Phys. A: Math. Gen. 38 (2005) 2555.

[25] S. Pack, X. Shen, J.W. Mark, J. Pan, Adaptive route optimization in hierarchical mobile ipv6 networks, IEEE Trans. Mobile Comput. 6 (2007) 903–914.