Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
August 2017, 24(4): 16–26
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.bupt.edu.cn

# Mobility driven network slicing: an enabler of on demand mobility management for 5G

Wang Hucheng[1,2] (✉), Chen Shanzhi[2], Ai Ming[2], Shi Yan[1]

1. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. State Key Laboratory of Wireless Mobile Communications, China Academy of Telecommunications Technology, Beijing 100191, China

**Abstract**

As different requirements on mobility support will be introduced by diversified communication scenarios in the fifth generation (5G), on demand mobility management is put forward to simplify signaling process, reduce terminal power consumption, improve network efficiency and so on. In order to enable on demand mobility management in 5G networks, a mobility driven network slicing (MDNS) was proposed, which takes individual mobility support requirements into account while customizing networks for different mobile services. Within the MDNS framework, the actual levels of required mobility support are determined by a mobility description system, and network slice templates with the corresponding mobility management schemes are defined by a network slice description function. By instantiating the network slices, each mobile terminal could be directed to the network slice with the most appropriate mobility management scheme. Based on this, a prototype was implemented to validate the feasibility of MDNS framework, i.e. creating multiple network slices with different mobility management schemes. In addition, the performance evaluation on average cost of processing a mobility event is conducted for the proposed MDNS framework and the long term evolution (LTE) system, and operating benefits are analyzed including efficiency and scalability.

**Keywords** mobility driven network slicing, network slicing, on demand mobility management, 5G

## 1 Introduction

The mobility driven network (MDN) [1] foresees that the development of mobile terminals and mobile applications, such as mobile social network, mobile cloud computing, and Internet of things (IOT), will bring many new communication scenarios. According to the International Telecommunications Union-Radiocommunications Sector (ITU-R) Report M.2083-0 [2], the forecasted communication scenarios in 5G include enhanced mobile broadband communication (eMBB), ultra-reliable and low latency communication (URLLC) and massive machine type communication (mMTC). These scenarios involve mobile terminals with different mobility behaviors and the applications with diversified service requirements, and cannot be efficiently served by current LTE system with 'one size fits all' architecture. Therefore, one of the challenges to design an efficient 5G system is how to customize mobile network to serving different communication scenarios and based on what criteria. The traditional way to customize mobile networks is re-building mobile networks, which absolutely is a long and complicated process due to complex and manual management operations of a large number of network nodes as introduced in Ref. [3]. Hence new means to flexibly customize operator's networks at low cost are anticipated. Fortunately, with the development of computer technologies as well as information technologies, e.g. network functions virtualization (NFV), network functions which were provided by dedicated hardware can be implemented as software components running on general purpose computers/hardware. By implementing 5G network functions as software components using NFV

Received date: 20-02-2017
Corresponding author: Wang Hucheng, E-mail: huchengwang@gmail.com
DOI: 10.1016/S1005-8885(17)60219-X

paradigm [4], operators can efficiently customize their mobile networks by orchestrating different sets of network functions to serve different communication scenarios or the 3rd parties. This new way to customize network services is deemed as one of key 5G technologies by next generation mobile networks (NGMN) [5], i.e. network slicing. However, the criteria of creating network slices are not discussed extensively.

Recently, many researches try to introduce network slicing technology into mobile networks. In Ref. [6], network slicing technology is adopted by the 3rd generation partnership project (3GPP) to create core network (CN) slices supporting different network services. Shimojo et al. [7] proposed to categorize the expected future services for slicing mobile networks accordingly. Nguyen et al. [8] introduces a mobile CN architecture which allows the mobile packet CN to be sliced into different control platforms according to either different mobile operators or different radio access technologies. In Ref. [9], Nikaein et al. proposed a new way for dynamic network slicing, where network applications and network functions in each network slice are retrieved from a so called 'network store'. Samdanis et al. [10] standing on the point of reducing capital expenditure and operational expenditure costs, put forward a notion of the 5G network slice broker to support on-demand multi-tenant mobile networks. Zhou et al. [11] proposed a concept of hierarchical network slicing as a service to help operators to offer customized end-to-end cellular networks as a service. However, these researches rarely take mobility support into account while discussing network slicing.

Mobility support is one of the most important capabilities of mobile networks [1]. Traditionally, mobility support is enabled by a statically configured mobility management scheme even in different communication scenarios. However, in 5G networks, due to diversified communication scenarios, enabling appropriate mobility management schemes is expected for better system performance. In order to adaptively provide mobility management in different mobility scenarios, many researches put forward new ideas/concepts, e.g., the Internet Engineering Task Force (IETF) introduces 'on demand mobility management' in Ref. [12], whereas the authors only take the application needs into account while selecting mobility management schemes, thus only Internet protocol (IP) session continuity and/or IP address reachability are differentiated in different mobility management schemes. A concept of 'on demand mobility support' was proposed in Ref. [13] but does not provides detailed solutions.

In this paper, we propose to use network slicing technology to support on demand mobility management. Relying on different mobile network architectures which are generated based on the requirements on mobility support, on demand mobility management can be easily enabled by serving mobile terminals with appropriate networks. With this thought, we put forward a new network slicing framework called 'MDNS framework' to automatically create and manage CN slices with different mobility management schemes. Comparing with the emerged network slicing systems, such as Refs. [8–9], MDNS framework shows following new features:

1) Levels of mobility support are defined to describe the requirements on mobility support.

2) The requirements on mobility support are determined by a mobility description system.

3) Traditional mobility management function is decomposed to 7 elementary functions.

4) Mobility management schemes are defined by orchestrating the required mobility management elementary functions based on the levels of mobility support.

5) Network slice templates (or blueprint) with designated mobility management schemes are defined by a network slice description function.

For validating the feasibility of MDNS framework enabling on demand mobility management, we implement an MDNS prototype in the SoftNet [14], where different mobility management schemes are implemented in different network slices. In order to demonstrate the efficiency of the proposed MDNS framework, we modeled the user equipment (UE) mobility handling under three types of 5G communication scenarios represented by industrial control scenario, IP multimedia subsystem (IMS) voice in a vehicle scenario and mobile Internet in hotspot area scenario, and then compare the average costs of handling a mobility event by using the MDNS and the LTE system for each communication scenario. Moreover, we analyze the operating benefits to explain that the proposed MDNS framework can create specific mobile networks with designated levels of mobility support rapidly, and enable on demand mobility management efficiently.

The rest of this paper is organized as follows. In Sect. 2, based on the analysis of the challenges for supporting on demand mobility management, the MDNS framework was proposed as well as its work principles. In Sect. 3, a simple prototype of MDNS is implemented for proof of concept. Furthermore, the performance evaluation and the operating benefits analysis are performed in Sect. 4. Finally, the conclusions and future work are provided in Sect. 5.

## 2 MDNS

According to the philosophy of MDN [1], mobility is an inherent feature and a key driving force of future networks. Such new requirements on mobility support will introduce new mobility management schemes as well as corresponding network architectures into 5G. As introduced in NGMN [7], network slicing is an efficient way to create a mobile network with specific functions. We hence propose MDNS to create network slices with different mobility support, i.e., with different mobility management schemes, in 5G. Having such network slices with different capabilities on mobility support, on demand mobility management support can be enabled by directing mobile terminals with different requirements on mobility support to appropriate network slices. Therefore, we conclude that MDNS is an enabler of on demand mobile management.

The implementation of MDNS needs to solve the following key issues. The first one is how to determine the mobility support required by mobile terminals. Unless the requirements on mobility support are clearly described, the network cannot provide the most suitable mobility management scheme. The second key issue is how to decompose traditional mobility management function into a set of independent elementary functionalities, which allows each specific network slice to only include required mobility management elementary functions. The last one is how to apply the selected mobility management schemes to mobile terminals flexibly. Mobile terminals require different mobility support in different scenarios, e.g. a smart phone does not require mobility support at middle night since usually the user is sleeping at home, thus the network shall be able to provide the mobility management adaptively.

The framework of MDNS is shown in Fig. 1, where, a mobility description system is deployed in the application layer to describe the requirements of mobile users on mobility support. A mobility management description function deployed in the operation and management layer is to define mobility management schemes as well as corresponding network functions. A network slice description function is used to define network slices supporting the designated mobility management schemes and a slice manager is to manage the lifecycles of all network slice instances. In the orchestration and control layer and the resource layer, existing NFV-management and orchestration (NFV-MANO) system developed by the European Telecommunications Standards Institute (ETSI) [15] is relied to instantiate network slices.

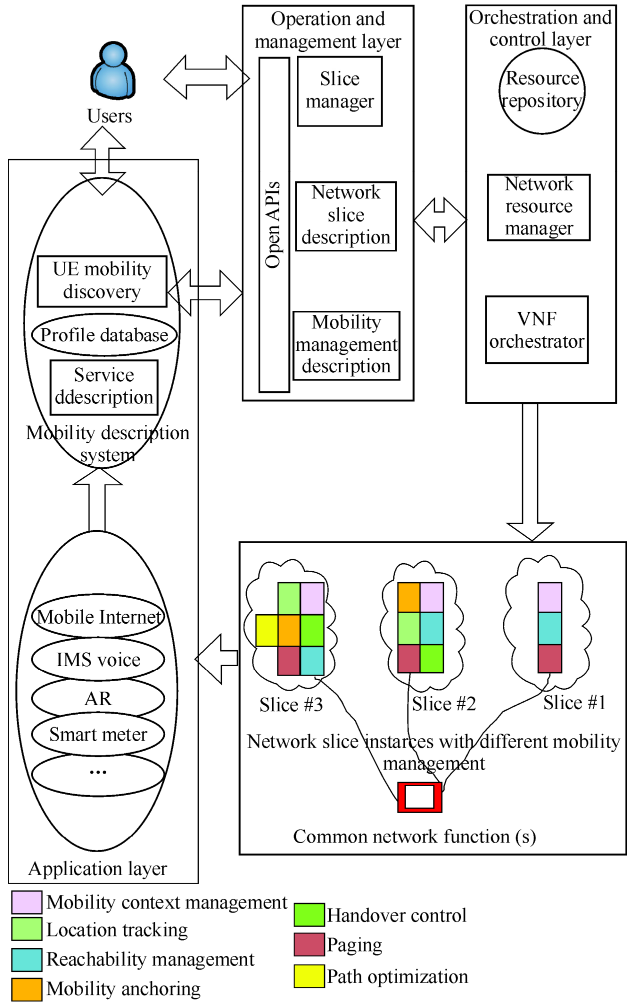



**Fig. 1** Framework of MDNS

As described in the Fig. 2, for enabling on demand mobility management by this framework, the mobility description system has to calculate the levels of mobility support required by groups of target mobile terminals by collecting mobility information of mobile terminals and the characteristics of mobile services. Each level of mobility support shall at least contain following information: the target users' mobility type (e.g. no, low or high mobility), target moving area, service continuity mode (e.g. no session, session continuity required or not), and service provider. Optionally, the mobility description system can suggest a default mobility management scheme for each level of mobility support. With the levels of mobility support, the slice manager within the operator's operating and management system, e.g. business support system/operation support system (BSS/OSS), can request the network slice description function to define network slices with designated levels of mobility supports. Accordingly, the network slice description function can request the mobility management description function to determine the corresponding mobility management schemes, and then to accurately describe the required elementary functionalities. Upon receiving the descriptions of network slices from the network slice description function, the slice manager can request, e.g. NFV-MANO system, to create the expected network slices. Finally, based on the level of mobility support required by each mobile terminal, a network slice with the most suitable mobility management scheme will be selected to serve the mobile terminal.

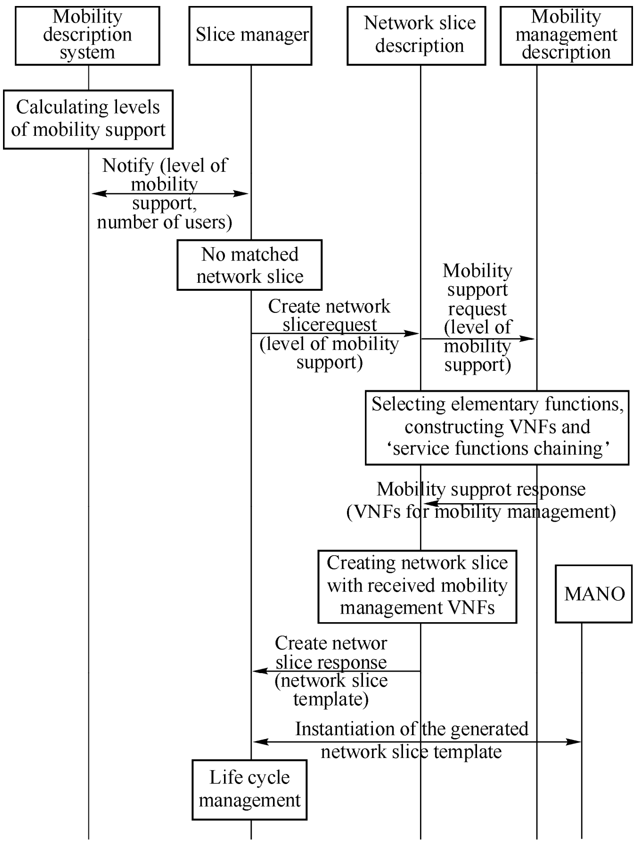



**Fig. 2** Instantiating a network slice with the designated level of mobility support

### 2.1 Mobility description system for determining the levels of mobility support

For clearly describing the mobility support required in different communication scenarios, firstly, we should categorize the communication scenarios forecasted in 5G, then we can define all levels of mobility support which should be provided in all communication scenario. By distinguishing the communication scenarios in multi dimensions such as the mobility types of mobile terminals, session continuity mode, and service provider of mobile services, the levels of mobility support in 5G mobile networks can be defined as shown in Table 1. However, depending on operator's policy, more/less factors can be taken into account, but it is not the key point of this paper.

**Table 1** An example of levels of mobility support

| Level | Mobile terminal | | Service characteristics | | Default mobility management scheme (optional) |
|---|---|---|---|---|---|
| | Mobility type | Moving area | Session continuity | Service type | |
| 1 | Any | Anywhere | No session | No service | CMM without user plane MA |
| 2 | No | One or more cells | Any | Any | DMM without handover control &location tracking |
| 3 | Low | Several tracking areas (TAs) | - | Any | DMM without handover control |
| 4 | Low | Several TAs | No | EMBB | DMM with LMA relocation |
| 5 | Low | Several TAs | Yes | EMBB URLLC | DMM without LMA relocation |
| 6 | Low | Several TAs | Yes | EMBB MMTC | DMM without CMA relocation |
| 7 | High | Whole public land mobile network (PLMN) | - | Any | CMM without handover control |
| 8 | High | Whole PLMN | No | EMBB | CMM with CMA relocation |
| 9 | High | Whole PLMN | Yes | EMBB URLLC | CMM without CMA relocation |
| 10 | High | Whole PLMN | Yes | EMBB MMTC | CMM without CMA relocation |

Based on the defined levels of mobility support, the mobility description system can clearly describe the requirements of mobile terminals on mobility support. By detecting mobility of mobile terminals and the characteristics of applications running on them, the mobility description system can determine the levels of required mobility support, and update their subscriptions to store the determined levels, e.g. in the profile database shown in Fig. 2. Furthermore, the determined levels of mobility support will be provided to the operator's operation and management system via open application programming interfaces (APIs).

The mobility information of mobile termianls is important to determine the levels of mobility support. However, it is not easy to discover mobility information of all mobile terminals by mobile network itself, especially in the network serving huge amount of mobile terminals. Therefore, the mobility description system may employ other mobility discovery platforms, e.g. data mining platform, to analyze and discover their mobility laws. On the other hand, for detecting the characteristics of applications running on mobility terminals, the mobility description system includes service description function as well.

### 2.2 Defining a network slice with designated mobility management scheme

As discussed in Ref. [16], core function decomposition provides a feasible way to obtain fine grained network slices allowing for further optimization in the CN. In the MDNS framework, a specific level of mobility support is enabled by a network slice implementing corresponding mobility management scheme. In order to flexibly define all kinds of mobility management schemes, the MDNS framework also needs to decompose the mobility management function defined in LTE network [17]. As an example, we decompose it into 7 elementary functions, so that any required mobility management scheme can be obtained by orchestrating selected mobility management elementary functions. The decomposed mobility management elementary functions in the MDNS framework are listed below:

1) Mobility context management: maintaining the state machine for each registered mobile terminal, and being responsible for creating, modifying or deleting mobility management context.

2) Location tracking: tracking and recording the location of each mobile terminal in idle mode for discovering it.

3) Paging: paging mobile terminals in idle mode if they have downlink signaling/data arrival.

4) Reachability management: detecting whether a mobile terminal is still reachable, and triggering the deletion of mobility context if the mobile terminal is unreachable.

5) Handover control: processing handover requests and controlling the handover procedures.

6) Mobility anchoring: user plane network function for terminating user plane paths of mobile terminals in mobile network.

7) Path optimization: relocating the mobility anchors based on the current locations of mobile terminals if possible, and/or updating user plane path(s) between mobile terminals and their mobility anchor(s) if necessary.

Table 1 gives an example of suitable mobility management schemes for diverse communication scenarios. In Table 1, centralized mobility management (CMM) implies control plane mobility management functions are centrally deployed in CN, e.g. the mobility management scheme in LTE network, distributed or decentralized mobility management (DMM) implies some or all control plane mobility management functions are deployed in radio access network (RAN), e.g. the distributed mobility management scheme introduced in Ref. [18], local mobility anchor (LMA) and central mobility anchor (CMA) represent user plane anchor closing to RAN and user plane anchor in CN respectively.

With the determined mobility management scheme, the network slice description function can define the corresponding network slice template. Finally, this network slice template will be sent to the NFV-MANO system for instantiation. Just as the introduced in Ref. [15], a virtual network function (VNF) may contain several VNF components (VNFCs), thus a VNF for mobility management can be an mobility management elementary function or a combination of several mobility management elementary functions. In later case, each mobility management elementary function acts as a VNFC.

### 2.3 Flexibly applying mobility management schemes to mobile terminals

In the MDNS framework, different mobility management schemes are provided by different network slices, so that suitable network slices need to be selected while mobile terminals with different requirements on mobility support accessing the network. For example, in Fig. 1, the network slice #1 supporting distributed mobility management scheme without handover control and area registration should be selected to serve a smart meter device. Whereas, the network slice #2 supporting CMN scheme without path optimization should be selected for a smart phone using IMS voice in wide area. In order to select a suitable mobility management scheme, an mobility profile detection and network control (MPDC) function is introduced into the MDNS framwork as a part of network slice selection function, which is a shared network function shown in Fig. 1. When a mobile terminal is accessing the network, the MPDC function will first determine the level of mobility support that the mobility terminal requires, then select a suitable network slice based on the determined level of mobility support.

Although the MDNS framework proposes to create network slices based on the requirements on mobility support, it does not imply that any level of mobility support will be enabled by a dedicated network slice, if a particular level of mobility support is required by only few mobile terminals, the MDNS may not create a dedicated network slice but update an existing network slice to support an additional mobility management scheme. In this case, as shown in Fig. 3, the MPDC function is further responsible for selecting the most suitable mobility management scheme for each mobile terminal, where the MPDC function acts as a service function chaining (SFC) control plane defined in Ref. [19].

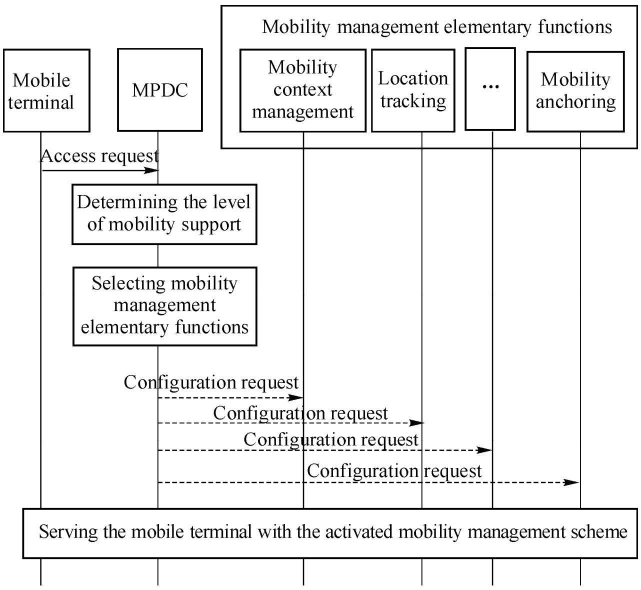
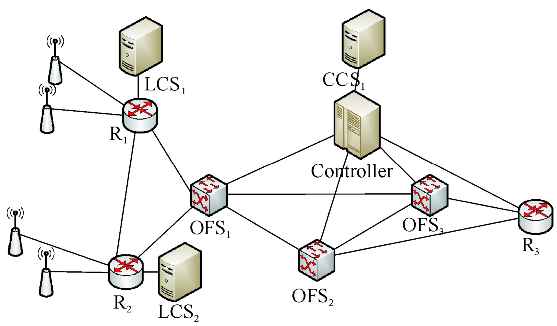



**Fig. 3** Enabling different mobility management schemes in same network slice

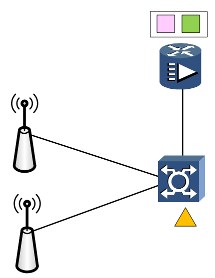
By chaining the selected mobility management elementary functions, the network can dynamically activate the required mobility management scheme to serve different mobile terminals.

## 3 Prototype implementation

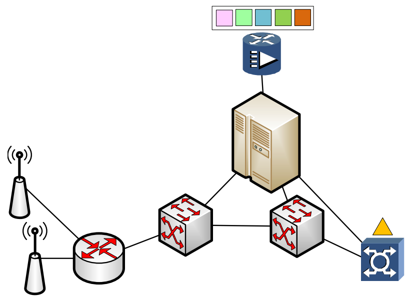
The main idea of MDNS is slicing mobile networks to support different mobility management schemes separately. For validating this concept, an infrastructure network as introduced in Ref. [14] is used, where control plane network function can be flexibly deployed in RAN or CN. In the prototype system, control plane elementary functionalities for mobility management are deployed in three servers, i.e. two local control servers (LCSs) in RAN and a central control server (CCS) in CN. These servers are general purpose computers, each of which runs multiple kernel-based virtual machine (KVM) with designated applications, e.g., application for handover control. The LCS and CCS in CN are interconnected via local area network (LAN). The user plane consists of IP routers and OpenFlow switches (OFSs). Based on the network configuration, the IP routers can act as user plane mobility anchors or intermediate data forwarding nodes, e.g. in the Fig. 4, $R_2$ acts as intermediate router in the logical network #2, but acts as LMA in logical network #3. Since each server can run multiple virtual machines, we can build multiple networks concurrently in the prototype system.



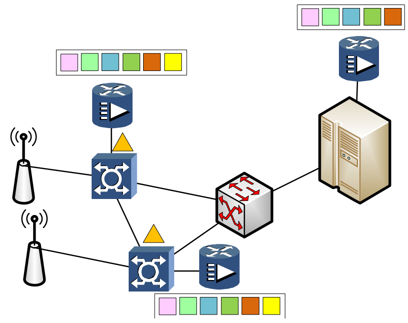

(a) Physical network topology



(b) Logical network #1 for industrial control



(c) Logical network #2 for IMS voice



(d) Logical network #3 for mobile Internet service in hotspot area

**Fig. 4** Prototype of dynamically enabling mobility management schemes in different network slices

By appropriately distributing mobility management elementary functions to RAN or CN based on the designated level of mobility support, the proposed prototype system implements three different mobility management schemes to serve different services shown in the Fig. 4. The first mobile network that we built aims to serve industrial control service. It is assumed that the mobile terminals used by industrial control service are connected to local application servers for obtaining services with low latency, and they are with no/low mobility in the factory, so we suppose the serving network should quickly response the requests from the mobile terminals, provide low latency data transmission. Based on this assumption, the mobility management function in traditional mobile network is tailored as shown in Fig. 4, where only locally deployed network functions for mobility context management and handover control need to be supported in the control plane, and a LMA is deployed closing to RAN. In contrast, in Fig. 4(c) designed for supporting IMS voice/video services, there is no limitation on the moving range of mobile terminals, thus CMM scheme with centralized network control and centralized mobility anchor is applied. Furthermore, as IMS service requires session continuity, i.e. IP address allocated to each mobile terminal cannot be changed, it is not applicable to apply path optimization. The last experimental network aims to provide mobile broad band services in hotspot area. It is supposed that mobile terminals have limited moving range while visiting mobile broad band services in the hotspot area.

In this case, DMM introduced in Ref. [13] can save lots of signaling cost and improve data forwarding efficiency, because of localized mobility event handling and traffic offloading.

## 4 Performance evaluation and operating benefits analysis

MDNS provides an efficient way to enable on demand mobility management in 5G networks, which not only improves the efficiency of mobility management, but also brings the benefits to network management and operation.

### 4.1 Performance evaluation

In this work, the average cost of processing a mobility event is considered as the performance metric. We define the cost of processing a mobility event as the control plane network resources consumed to handle this mobility event, e.g. location update, handover, and then derive the average cost of processing a mobility event:

$$\overline{C} = \sum_{n=1}^{N} r_n C_n \tag{1}$$

where $r_n$ is the ratio of event $n$ happens, $\sum_{n=1}^{N} r_n = 1$, and $C_n$ represent the cost of handling the mobility event $n$.

As there are different mobility events in different communication scenarios, we need to consider specific communication scenario for comparison. Take the scenarios of industrial control (Scenario A), IMS voice in a vehicle (Scenario B) and mobile Internet in hotspot area (Scenario C) as examples, and then calculate the average cost of processing a mobility event under each scenario.

For Scenario A, it is assumed that the mobile terminals are static, and the connections with the network are always kept for low latency communication. Therefore, the MDNS framework only activates mobility context management and handover control elementary functions to serve them. Based on this assumption, the average cost of processing a mobility event equals the cost of handling a handover event, i.e.

$$\overline{C_{\mathrm{A}}^{\mathrm{MDNS}}} = C_{\mathrm{H}} \tag{2}$$

where $C_{\mathrm{H}}$ implies the cost of handling a handover event by the MDNS.

For Scenario B, the MDNS framework needs to activate most of control plane mobility management elementary functions, including mobility context management, location tracking, reachability management, handover control and paging. If mobility events defined in LTE system are assumed to be applied to Scenario B, we need to calculate the cost of handling service request, $S_1$ release, handover and tracking area update (TAU) procedure. In order to calculate the cost of handing such mobility events, we first model the mobility of a mobile terminal by a discrete-time Markov chain model as shown in Fig. 5.

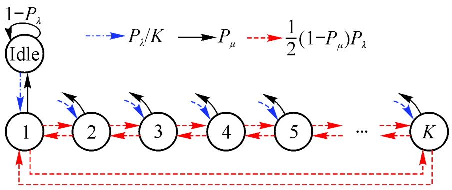



**Fig. 5** Mobility model of a mobile terminal under Scenario B

In the discrete-time Markov chain model, it is assumed that, each cell represents the area served by a macro base station. Each tracking area covers $K$ cells serving the mobile terminal. A mobile terminal can only move one cell to another cell with the probability $P$=1/2. Therefore, the state space can be defined to $S=\{S_{\text{idle}}, S_i(1\leqslant i\leqslant K)\}$, the state $S_{\text{idle}}$ represents that a mobile terminal has no active session, the state $S_i$, $1\leqslant i\leqslant K$ represents the mobile terminal with active session(s) locating at the cell $i$. We further assume that the rate of session arrival in a cell follows Poisson distribution with a mean $\lambda$. The session duration follows an exponential distribution with a mean $1/\mu$. And, the state of mobile terminal is only changed at the end of each time slot $\tau$ and only one change is allowed at a time. Then, the stationary probability can be expressed as followed:

$$\left.\begin{aligned}
&\pi_{\text{idle}}+\sum_{i=1}^{K}\pi_i=1\\
&\pi_{\text{idle}}=(1-P_\lambda)\pi_{\text{idle}}+P_\mu\sum_{i=1}^{K}\pi_i\\
&\pi_1=\frac{1}{K}P_\lambda\pi_{\text{idle}}+\frac{1}{2}(1-P_\mu)P_\gamma(\pi_2+\pi_K)\\
&\pi_i=\frac{1}{K}P_\lambda\pi_{\text{idle}}+\frac{1}{2}(1-P_\mu)P_\gamma(\pi_{i-1}+\pi_{i+1})\\
&\pi_K=\frac{1}{K}P_\lambda\pi_{\text{idle}}+\frac{1}{2}(1-P_\mu)P_\gamma(\pi_{K-1}+\pi_1)
\end{aligned}\right\}\tag{3}$$

Based on this model, we can derive the probability of each mobility event: the probability of service request event is $P_{\text{SR}}=\pi_{\text{idle}}P_\lambda$. The probability of $S_1$ release event is $P_{S_1}=(1-\pi_{\text{idle}})P_\mu$. The probability of TAU event is $P_{\text{TAU}}=(\pi_1+\pi_K)(1/2)(1-P_\mu)P_\gamma+(2/K)\pi_{\text{idle}}(1/2)(1-P_\lambda)\cdot P_\gamma$. And the probability of handover event is $P_{\text{H}}=\sum_{i=1}^{K}\pi_i(1-P_\mu)P_\gamma$. Then, the average cost of processing a mobility event can be expressed as:

$$\overline{C_{\text{B}}^{\text{MDNS}}}=\frac{P_{\text{SR}}}{P_{\text{SR}}+P_{S_1}+P_{\text{TAU}}+P_{\text{H}}}C_{\text{SR}}+\frac{P_{S_1}}{P_{\text{SR}}+P_{S_1}+P_{\text{TAU}}+P_{\text{H}}}\cdot C_{S_1}+\frac{P_{\text{TAU}}}{P_{\text{SR}}+P_{S_1}+P_{\text{TAU}}+P_{\text{H}}}C_{\text{TAU}}+\frac{P_{\text{H}}}{P_{\text{SR}}+P_{S_1}+P_{\text{TAU}}+P_{\text{H}}}C_{\text{H}}\tag{4}$$

where $C_{\text{SR}}$ implies the cost of handling a service request event by the MDNS, $C_{S_1}$ implies the cost of handling a $S_1$ release event by the MDNS, $C_{\text{TAU}}$ implies the cost of handling a TAU event by the MDNS.

For Scenario C, the mobile terminals are assumed to move within a designated area, e.g. a tracking area. The MDNS therefore provides localized mobility management control, i.e. activates mobility context management, location tracking, reachability management, handover control, paging and path optimization on the RAN level. In order to calculate the cost of handing such mobility events, we also need to model the mobility of a mobile terminal. As introduced in Refs. [20–21], the network deployment in a hopspot area can be model to a grid topology as shown in Fig. 6, where each cell $(i, j)$ represents the area served by a small cell base station, and all cells are within a tracking area. With this assumption, we can also model the mobility by a discrete-time Markov chain model. The state space can be expressed as $S=\{S_{\text{idle}}, S_i^j(0\leqslant i\leqslant R, 1\leqslant j\leqslant 4i)\}$, the state $S_{\text{idle}}$ represents that a mobile terminal has no active session, the state $S_i^j$ $(0\leqslant i\leqslant R, 1\leqslant j\leqslant 4i)$ represents the mobile terminal locating at the cell $(i, j)$. We still assume that the rate of session arrival in a cell follows Poisson distribution with a mean $\lambda$, the session duration follows an exponential distribution with a mean $1/\mu$, the residence time of the mobile terminal in each cell follows an exponential distribution with a mean $1/\gamma$, and the state of the mobile terminal is only changed at the end of each time slot $\tau$ and only one change is allowed at a time. Finally, we derive the stationary probability of the mobile terminal in each cell similar as those presented in Ref. [21].

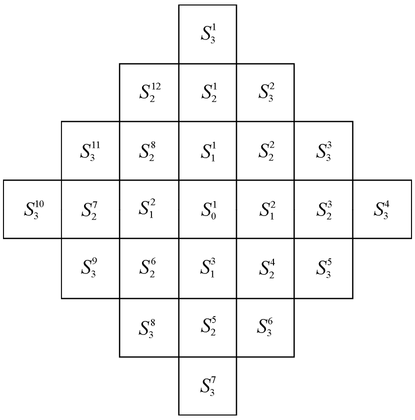



**Fig. 6** Grid network deployment in a hotspot area

At last, we derive the probability of each mobility event under this scenario: the probability of service request event is $P_{\mathrm{SR}} = \pi_{\mathrm{idle}} P_{\lambda}$. The probability of $S_1$ release event is $P_{S_1} = (1-\pi_{\mathrm{idle}})P_{\mu}$. And the probability of handover event is $P_{\mathrm{H}} = (1-\pi_{\mathrm{idle}})(1-P_{\mu})P_{\gamma}$ . Then, the average cost of processing a mobility event is:

$$\overline{C_{\mathrm{C}}^{\mathrm{MDNS}}} = \frac{P_{\mathrm{SR}}}{P_{\mathrm{SR}} + P_{S_1} + P_{\mathrm{H}}} C_{\mathrm{SR}} + \frac{P_{S_1}}{P_{\mathrm{SR}} + P_{S_1} + P_{\mathrm{H}}} C_{S_1} + \frac{P_{\mathrm{H}}}{P_{\mathrm{SR}} + P_{S_1} + P_{\mathrm{H}}} C_{\mathrm{H}} \tag{5}$$

Let us choose the existing LTE system as the baseline for comparison. As the LTE system requires that all mobility events of a mobile terminal are handled by a dedicated MME, for any types of mobility events, the MME has to prepare whole mobility management resources. With this consideration, we assume the average cost of processing a mobility event equals the cost of processing any mobility event, i.e.

$$\overline{C^{\mathrm{LTE}}} = C_{\mathrm{att}} + C_{\mathrm{SR}} + C_{S_1} + C_{\mathrm{TAU}} + C_{\mathrm{H}} \tag{6}$$

where $C_{\mathrm{att}}$ implies the cost of handling an attach event by the MDNS.

Based on the developed mathematical model and cost functions, we perform numerical simulations by using Matlab. The values of parameters used in the simulation are given in the Table 2, where *R* indicates the radius of the hotspot area in term of the number of cells, and the cost of handling a mobility event is represented by a scalarization value. The scalarization values for the MDNS based solution are derived based on the number of messages that the network has to process.

**Table 2** Parameters used in the evaluation

| Parameter | Value |
|---|---|
| $\lambda$ / s$^{-1}$ | 0.001 |
| $\mu$ / s$^{-1}$ | 0.0025 , 0.005 , 0.01, 0.02, 0.04,0.08 |
| $\gamma$ / s$^{-1}$ | 0.025 , 0.05 , 0.1, 0.2, 0.4,0.8 |
| $\tau$ / s$^{-1}$ | 0.01 |
| $K$ | 5 |
| $R$ | 5 |
| $C_{\mathrm{SR}}$ | 1 |
| $C_{S_1}$ | 1 |
| $C_{\mathrm{TAU}}$ | 3 |
| $C_{\mathrm{H}}$ | 5 |
| $C_{\mathrm{att}}$ | 4 |

Fig. 7 shows the average costs of processing a mobility event while using mobility management mechanism in the LTE system and the proposed MDNS framework for the industrial control scenario. As only handover event needs to be handled in the industrial control scenario, the average cost of using MDNS based solution is a constant value. The average cost in the LTE system is also a constant value, because we assumed that, for any mobility event, the MME has to prepare whole mobility management resources to handle it.

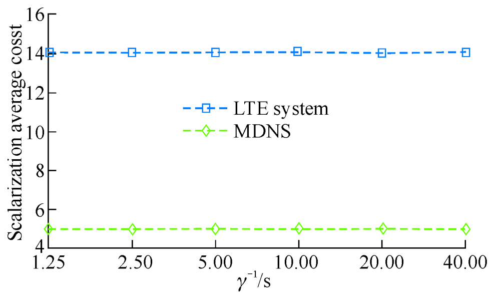
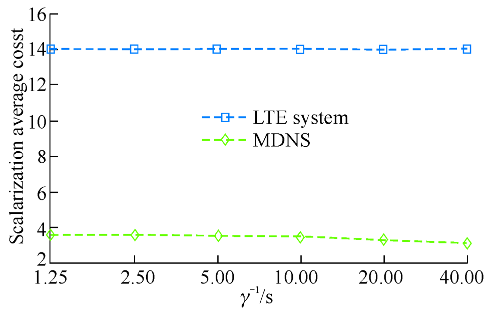
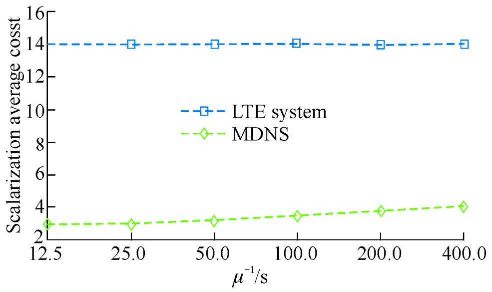



**Fig. 7** Average costs of processing a mobility event against cell residence time under the Scenario A

Fig. 8 shows the average costs of processing a mobility event while using mobility management mechanism in the LTE system and the proposed MDNS framework under the IMS voice in a vehicle scenario. According to Fig. 8(a), with the increase of average residence time of the mobile terminal in each cell, the average cost of using MDNS based solution decreases slowly, that is because the increased average residence time reduces the frequency of handover which contributes the most cost. Whereas in Fig. 8(b), with the increase of average session duration, the average cost of using MDNS based solution increase slowly, that is because the increased average session duration increases the probability of handover.



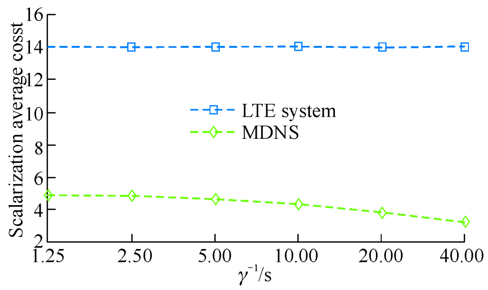

(a) Average costs of processing a mobility event against cell residence time



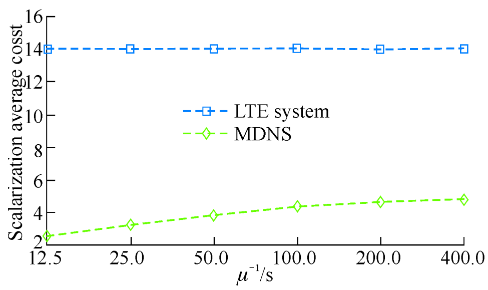

(b) Average costs of processing a mobility event against session duration

**Fig. 8** Average costs of processing a mobility event under the Scenario B

Fig. 9 shows the average costs of processing a mobility event while using mobility management mechanism in the LTE system and the proposed MDNS framework under the mobile internet in a hot spot area scenario.




(a) Average costs of processing a mobility event against cell residence time




(b) Average costs of processing a mobility event against session duration

**Fig. 9** Average costs of processing a mobility event under the Scenario C

Similarly, due to increased average residence time of the mobile terminal in each cell, less handover events need to be handled, and then the system cost of handling mobility events is reduced as shown Fig. 9(a). With the increase of average session duration, the probability of handover increase, which leads to more cost on handling mobility events.

### 4.2 Operating benefits analysis

According to the MDNS, the instantiation of any specialized network slice is automatically triggered by the slice manager function, which simplifies the network configurations, and then improves the efficiency of network management. For each specialized network slice supporting designated mobility management schemes, the MDNS requires only necessary mobility management elementary functions to be enabled, thus the specialized networks can be implemented with reduced network resources including computation, storage and transport. In addition, compared with uniform mobility management scheme applied to all user equipments in LTE network, customized mobility management schemes are provided in specialized networks, so that mobility events handling for mobile terminals attached to this network can be refined as much as possible. With this approach, on demand mobility management can be realized with simplified network control logic but efficient network resource utilization.

Furthermore, in MDNS, mobility management function is decomposed into a set of independent elementary functions. Therefore, any specific mobility management scheme can be flexibly implemented by selecting and instantiating required set of elementary function. In this way, the scalability of 5G mobile networks is provided to network operators to support new mobile services.

## 5 Conclusions and future work

On demand mobility management is a critical component of 5G mobile networks to serve different communication scenarios, but there are still challenges for efficiently realizing it. For tackling the challenges, based on the rethink of the capability of mobile network and the nature of network slicing, we propose a new network slicing mechanism, i.e. MDNS, as well as a framework implementing MDNS. The MDNS framework can create network slices with different mobility management schemes, so that on demand mobility management can be efficiently enabled by providing different mobility support in different network slices. The feasibility of this framework is validated by the implemented MDNS prototype. The cost efficiency of the MDNS framework handling mobility events is demonstrated by the developed mathematical model. Moreover, the analysis on it shows that, the MDNS framework can bring benefits on network management and operation because it can automatically create and manage network slices with different mobility management schemes. Future work includes implementing the whole MDNS system, and subsequently, quantitatively evaluating the management complexity of creating or revoking a network slice with a designated mobility management scheme, i.e. the cost on time, computation, storage and transport aspects.

**Acknowledgements**

This work was supported by the National Science and Technology Major Project of China (2017ZX03001014), the National Natural Science Foundation of China for Distinguished Young Scholar (61425012).

## References

1. Chen S Z, Shi Y, Hu B, et al. Mobility-driven networks (MDN): from evolutions to visions of mobility management. IEEE Network, 2014, 28(4): 66−73
2. ITU-R M.2083-0. IMT vision─framework and overall objectives of the future development of IMT for 2020 and beyond. 2015
3. Ahmed M F, Talhi C, Cheriet M. Towards flexible, scalable and autonomic virtual tenant slices. Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM'15), May 11−15, 2015, Ottawa, Canada. Piscataway, NJ, USA: IEEE, 2015: 720−726
4. Abdelwahab S, Hamdaoui B, Guizani M, et al. Network function virtualization in 5G. IEEE Communications Magazine, 2016, 54(4): 84−91
5. Iwamura M. NGMN view on 5G architecture. Proceedings of the IEEE 81st Vehicular Technology Conference (VTC-Spring'15), May 11−14, 2015, Glasgow, UK. Piscataway, NJ, USA: IEEE, 2015: 5p
6. 3GPP TR 23.799 v14.0.0. Study on architecture for next generation system (release 14). 2016
7. Shimojo T, Takano Y, Khan A, et al. Future mobile core network for efficient service operation. Proceedings of the IEEE 1st Conference on Network Softwarization (NetSoft'15), Apr 13−17, 2015, London, UK. Piscataway, NJ, USA: IEEE, 2015: 6p
8. Nguyen V G, Kim Y H. Slicing the next mobile packet core network. Proceedings of the 11th International Symposium on Wireless Communications Systems (ISWCS'14), Aug 26−29, 2014, Barcelona, Spain. Piscataway, NJ, USA: IEEE, 2014: 901−904
9. Nikaein N, Schiller E, Favraud R, et al. Network store: exploring slicing in future 5G networks. Proceedings of the 10th International Workshop on Mobility in the Evolving Internet Architecture (MobiArch'15), Sept 7, 2015, Paris, France. New York, NY, USA: ACM, 2015: 8−13
10. Samdanis K, Costa-Perez X, Sciancalepore V. From network sharing to multi-tenancy: the 5G network slice broker. IEEE Communications Magazine, 2016, 54(7): 32−39
11. Zhou X, Li R P, Chen T, et al. Network slicing as a service: enabling enterprises' own software-defined cellular networks. IEEE Communications Magazine, 2016, 54(7): 146−153
12. Yegin A, Kweon K, Lee J, et al. On demand mobility management. draft-ietf-DMM-on-demand-mobility-02. IETF, 2016
13. 3GPP TR 22.864 v14.0.0. FS_SMARTER−network operation (release 14). 2016
14. Wang H C, Chen S Z, Xu H, et al. SoftNet: a software defined decentralized mobile network architecture toward 5G. IEEE Network, 2015, 29(2): 16−22
15. ETSI GS NFV-MAN 001 v1.1.1. Network functions virtualisation (NFV); management and orchestration. 2014
16. Sama M R, An X L, Wei Q, et al. Reshaping the mobile core network via function decomposition and network slicing for the 5G era. Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, Apr 3−6, 2016, Doha, Qatar. Piscataway, NJ, USA: IEEE, 2016: 7p
17. 3GPP TS 23.401 v13.6.1. General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E-UTRAN) access (release 13). 2016
18. Chan H, Liu D, Seite P, et al. Requirements for distributed mobility management. IETF RFC 7333. 2014
19. Halpern J, Pignataro C. Service function chaining (SFC) architecture. IETF RFC 7665. 2015
20. Wang H C, Chen S Z, Ai M, et al. Localized mobility management for 5G ultra dense network. IEEE Transactions on Vehicular Technology (In publish)
21. Balakrishnan R, Akyildiz I F. Local mobility anchoring for seamless handover in coordinated small cells. Proceedings of the 2013 IEEE Global Communications Conference (GLOBECOM'13), Dec 9−13, 2013, Atlanta, GA, USA. Piscataway, NJ, USA: IEEE, 2014: 4489−4494

(Editor: Wang Xuying)