This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

SMDP-Based Radio Resource Allocation Scheme in Software-Defined Internet of Things Networks

Xiong Xiong, Lu Hou, Kan Zheng, *Senior Member, IEEE*, Wei Xiang, *Senior Member, IEEE*, M. Shamim Hossain, *Senior Member, IEEE*, and Sk Md Mizanur Rahman

***Abstract*—With rapid development of the Internet of Things (IoT), various machine-to-machine (M2M) communications technologies have emerged in recent years to provide ubiquitous wireless connections for a massive number of IoT devices. This poses significant challenges to network control and management of large-scale IoT networks. Software-defined networking (SDN) is considered a promising technology to streamline network management due to dynamic reconfigurable network elements. Thus, the integration of SDN and IoT provides a potentially feasible solution to strengthening management and control capabilities of the IoT network. Benefit from SDN technology, resource utilization in the IoT network can be further enhanced. In this paper, we first propose a software-defined network architecture for IoT. Then, the resource allocation problem in the proposed SDN-based IoT network is investigated. The optimal problem of maximizing the expected average rewards of the network is formulated as a semi-Markov decision process (SMDP). The optimal solution is obtained through solving the SMDP problem using a relative value iteration algorithm. Simulation results demonstrate that the proposed resource allocation scheme is able to improve the system rewards compared with other comparative resource allocation schemes.**

***Keywords*—IoT; SDN; semi-Markov decision process (SMDP).**

I. Introduction

The advent of the Internet of Things (IoT) has inspired a large variety of new applications that can provide ubiquitous services to make existing industrial systems and peoples life more intelligent, e.g., industrial automation, smart grids, intelligent transportation systems, smart healthcare, smart home, etc [1]. However, with the rapid development of IoT technology, the management and control of IoT networks is met with considerable challenges. It is expected that there will be 100-fold or more machine-to-machine (M2M) connections in the emerging fifth generation (5G) era [2], which requires the network not only to support the massive M2M connections, but also to be optimized with the aid of big data analytics [3]. Meanwhile, the massive number of connections by M2M devices will share the same network with human-to-human (H2H) users. The M2M connections often have different requirements on the network due to the unique characteristics of M2M communications [4] [5].

Furthermore, various M2M communications technologies have emerged in recent years, although none of them can meet all the requirements of M2M connections. At present, M2M communications can be classified into two categories, i.e., cellular M2M solutions, and capillary M2M solutions [6]. To make cellular networks more suited for M2M communications, a great deal of efforts have been made in third generation partnership project (3GPP) recent releases by defining long-term evolution for machine-type communications (LTE-MTC), which aims for low power consumption and moderate costs [7]. Moreover, some clean-slate cellular M2M solutions are also developed to provide long terminal battery life and extended coverage, such as narrowband IoT (NB-IoT) [8]. On the other hand, capillary M2M solutions have been widely used in IoT applications for many years, which provide wireless connections for devices with low power consumption, such as ZigBee, Wi-Fi, etc. In addition, low power wide area (LPWA) technology as another promising solution has been proposed to meet the enhanced coverage requirement of some M2M application scenarios such as rural areas, which provides a long-range connectivity up to several kilometers, such as LoRa [9]. As a consequence, all the techniques and solutions for M2M communications may coexist and work together to support various IoT applications in the future. Thus, the heterogeneity of M2M connectivity also brings about significant challenges to dynamically optimize radio resource allocation and management in the IoT network.

To address these challenges, the software-defined networking (SDN) is considered as a promising candidate technology for simplifying network management due to its programmable and centralized network control [10]. SDN was originally proposed and designed for computer networks to decouple control decisions from forwarding devices such as switches and routers [11]. Many previous efforts have been made to migrate the concept of SDN to wireless networks, especially software-defined mobile networks (SDMN), such as Openradio [12], SoftRAN [13], SoftCell [14], CROWD [15], MobileFlow [16], SoftNet [17], SERVICE [18], etc. Moreover, the integration of SDN and IoT has also been attempted [19]. Sensor OpenFlow

This work is funded in part by National High-Tech R&D Program (863 Program 2015AA01A705), National Science Foundation of China (No.61331009), Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for its funding (PRG-1436-17), National Key Technology R&D Program of China (No.2014ZX03003011-004) and Fundamental Research Funds for the Central Universities (No.2014ZD03-02).

Xiong Xiong, Lu Hou and Kan Zheng are with the Intelligent Computing and Communication (IC^2) Lab, Wireless Signal Processing and Networks Lab (WSPN), Key Lab of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China, 100088. (e-mail: zkan@bupt.edu.cn).

Wei Xiang is with the College of Science and Engineering, James Cook University, Cairns, QLD 4878, Australia.

M. Shamim Hossain and Sk Md Mizanur Rahman are with the Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, KSA.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

is proposed as a software-defined wireless sensor network (WSN) architecture to tackle the inherent problems of WSN, i.e., rigidity to policy changes and difficulty to manage [20]. In [21], the authors propose a SDN controller design for the IoT network, which enables centralized flow scheduling based on network calculus model. In [22], SDN-WISE is designed as a stateful SDN solution for the WSN, and the performance of SDN-WISE is also evaluated based on a prototype implementation. UbiFlow adopts a distributed hashing based overlay structure for flow scheduling and mobility management in a software-defined IoT system [23].

To the best of our knowledge, the research on integrating SDN and IoT networks is still at its infancy. Many problems remain open, e.g., network architecture, protocols, controller design, standardization, etc. In this paper, we first propose a software-defined network architecture for IoT, which is based on the 3GPP proposed architecture for machine type communications (MTC). Then, a novel radio resource allocation scheme is proposed for the software-defined IoT network. It is noted that issues relating to radio resource allocation in SDN-based IoT networks are rarely studied in the literature. We formulate the radio resource allocation problem in the SDN-based IoT network as a semi-Markov decision process (SMDP) [25], which provides a mathematical framework for modeling a wide range of optimization problems [26] [27] [28]. From a global network point of view, the SDN controller can make optimal decisions to maximize the expected average reward of the network. The optimal radio resource allocation policy is obtained through solving the SMDP problem using the relative value iteration algorithm [25]. To verify the performance of our proposed model, a simulation-based study is performed by comparing the performances of our model and other reference algorithms. Numerical results will be presented to demonstrate that our approach is able to enhance the overall resources utilization resulting in improved performance for the SDN-based IoT network.

The remainder of this paper is organized as follows. In Section II, the system model is described, and the resource allocation problem is formulated as an SMDP process. The solution to the SMDP problem is attempted in Section III. Then, Section IV provides numerical results as well as a performance analysis of our proposed scheme. Finally, concluding remarks are drawn in Section V.

II. System Model and Problem Formulation

In this section, we first describe a software-defined heterogeneous network architecture for IoT. Then, our proposed optimal radio resource allocation scheme for this SDN-based IoT network architecture is presented, which is formulated as an SMDP problem. Next, we describe the system states of our proposed model, and the actions that can be taken in each state. The system reward model is also described, which plays a significant role in decision making by the SDN controller.

A. System Model

We consider an SDN-based IoT network as illustrated in Fig. 1, which basically consists of an eNB, N MTC gateways

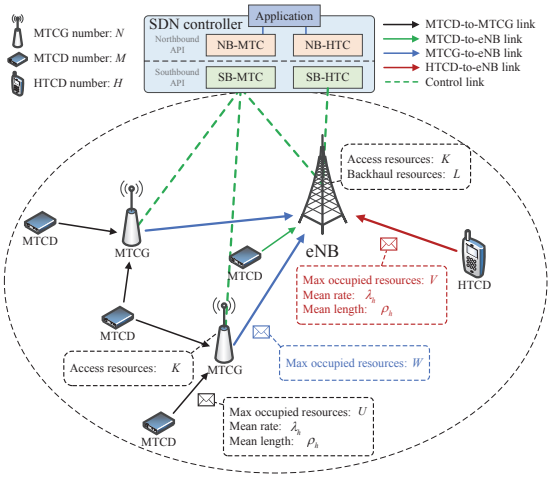

Fig. 1. Illustration of the system model.

(MTCGs), and a local SDN controller. As a new network element, the SDN controller is the control plan of the overall network. Most of the control logics are implemented in the controller, or rest in management applications that interact with the SDN controller via the northbound application program interface (API). The SDN controller has two essential functions. Firstly, it collects the overall status information of the eNB and each MTCG, and offers a global network view to the management applications. Secondly, the SDN controller makes decisions and controls the eNB and MTCGs dynamically based on the obtained network status, which can be implemented by programming through the southbound API, e.g., extensions to the OpenFlow protocol [11].

The eNB and MTCGs are responsible for providing radio access for MTC devices, and the coverage areas of the eNB and MTCGs overlap in the geographical region of the SDN-based IoT network. Thus, M MTC devices (MTCDs) dispersed in the serving area can access to the core network through either the eNB or relaying via an MTCG. The direct connection between an MTCD and the eNB is dubbed the MTCD-to-eNB link, while the MTCD-to-MTCG link describes the wireless link between an MTCD and an MTCG. We assume that the MTCD-to-eNB links and MTCD-to-MTCG links adopt the same NB-IoT M2M communications technique, which is considered as the most potential candidate M2M technology for the future fifth generation (5G) wireless network. Compared to existing cellular techniques, NB-IoT provides limited transmission rates and extended coverage. Each MTCD-to-MTCG and MTCD-to-eNB link can support l_m transmission modes with different transmission rates $\mathrm{C}_m(\cdot)$ in accordance with the signal-to-noise ratio (SNR), which indicates the quality of the channel. The SNR range can be split into l_m non-overlapping intervals with boundary points denoted by $\{\{\theta_{m,i}\}_{i=1}^{l_m+1}\}$, where $\theta_{m,l_m+1} = \infty$. When SNR $\xi \in [\theta_{m,i}, \theta_{m,i+1})$, the transmission mode i is applied and its transmission rate $\mathrm{C}_m(\xi)$ is equal to $\mu_{m,i}$, where $i \in \{1, \ldots, l_m\}$. Suppose each MTCG, as well as the eNB, owns K radio resources (termed *access resources* (ARs)), which are assigned to the MTCD-to-MTCG

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

and MTCD-to-eNB links. Each MTCD-to-MTCG link and MTCD-to-eNB link can consume u ARs of the connected MTCG and eNB, where $u \in \{1,\ldots,U\}$ and $U \leq K$. Moreover, data packets transmitted from an MTCD follow a Poisson distribution with the mean rate of λ_m, and the MTC packet size follows an exponential distribution with a mean size of ρ_m. Therefore, the total arrival rate of the MTC data packets to the network is equal to $M\lambda_m$.

When an MTCD is connected to an MTCG, the MTCG performs as a one-hop relay to forward data packets to the eNB via the MTCG-to-eNB link. It is assumed that the MTCG-to-eNB link employs the 3GPP LTE/LTE-A cellular communications technique, which supports high bandwidth transmission and operates in a different frequency band from NB-IoT. We suppose that the eNB owns L radio resources (termed *backhaul resources* (BRs)) for the MTCG-to-eNB links, and the eNB has many more BRs than ARs, i.e., $K < L$. Each MTCG-to-eNB link occupies w BRs of the eNB, where $w \in \{1,\ldots,W\}$, $W \leq L$. Meanwhile, the eNB is also responsible for supporting H2H communications. H HTC devices (HTCDs) coexist with MTCDs in the serving areas of the network, and interact with the eNB through the HTCD-to-eNB links. The HTCD-to-eNB links and MTCG-to-eNB links share the common BRs of the eNB, which allocates v BRs for each HTCD-to-eNB link, where $v \in \{1,\ldots,V\}$, $V \leq L$. Assume that each HTCD sending data packets also follows an independent Poisson process with a mean rate of λ_h, and the size of the HTC packets also follows an exponential distribution with a mean size of ρ_h. Each MTCG-to-eNB link and HTCD-to-eNB link also can support l_h transmission modes with different transmission rates $\mathrm{C}_h(\cdot)$. Let $\{\{\theta_{h,i}\}_{i=1}^{l_h+1}\}$ be the set of SNR boundary points, where $\theta_{h,l_h+1} = \infty$. When SNR $\xi \in [\theta_{h,i}, \theta_{h,i+1})$, the transmission rate $\mathrm{C}_h(\xi)$ of transmission mode i is equal to $\mu_{h,i}$, where $i \in \{1,\ldots,l_h\}$.

To maximize network revenue, the network must decide the access strategy for each MTCD (i.e., through the eNB or an MTCG), and optimize radio resource allocation for each wireless link. Thus, we formulate the radio resource allocation model in the software-defined IoT network as an SMDP problem. In SDN, all the decision-making procedures are carried out in the SDN controller. When a new MTC or HTC data packet arrives, the SDN controller first evaluates the expected system gain and system expense based on the current status information of the eNB and MTCGs. Then, the SDN controller decides whether to accept or reject the data packets, which MTCGs or eNB should be selected for MTC traffic transmission, and how to allocate radio resources to each wireless link. In the SMDP framework, the decisions adopted by the SDN controller are called *actions*, and the moments when decisions are made are termed *decision epochs*. The action chosen is based on the current *system state* of the network, which includes the current traffic load on each MTCG and the eNB. To make the optimal decision for the reward model, the SDN controller needs to obtain the system reward for each action before making any decision. The long-term expected average reward per unit time of the network is considered as the optimal criterion for the SMDP.

B. Problem Formulation

This subsection aims to formulate the considered optimized problem as an SMDP. The corresponding system states, actions based on each state, and the reward model are described as follows.

1) System States: The system state $\mathcal{S}$ of the software-defined IoT network can be represented by the number of current wireless links in the network with different numbers of radio resources occupied, as well as an event occurred in the system, which could be either the arrival or departure of a data packet. The system state space $\mathcal{S}$ can be denoted as follows, i.e.,

$$\mathcal{S} = \{s \mid s = (\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_N, \boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N, \boldsymbol{s}_m, \boldsymbol{s}_h, e)\}, \tag{1}$$

where $\boldsymbol{g}_i, \boldsymbol{s}_i, \boldsymbol{s}_m, \boldsymbol{s}_h$, $i \in \{1,\ldots,N\}$ are defined as

$$\begin{aligned} \boldsymbol{g}_i &= (g_{i,1}, g_{i,2}, \ldots, g_{i,U}), \\ \boldsymbol{s}_i &= (s_{i,1}, s_{i,2}, \ldots, s_{i,W}), \\ \boldsymbol{s}_m &= (s_{m,1}, s_{m,2}, \ldots, s_{m,U}), \\ \boldsymbol{s}_h &= (s_{h,1}, s_{h,2}, \ldots, s_{h,V}). \end{aligned} \tag{2}$$

The above symbols are explained in detail below

- $\boldsymbol{g}_i$: a vector of $g_{i,u}$, where $u \in \{1,\ldots,U\}$. $g_{i,u}$ represents the number of wireless links between the MTCDs and the ith MTCG that occupy u ARs. The total number of allocated ARs of the ith MTCG should satisfy $\sum_{u=1}^{U}(ug_{i,u}) \leq K$;
- $\boldsymbol{s}_m$: a vector of $s_{m,u}$, where $u \in \{1,\ldots,U\}$. Similarly, $s_{m,u}$ represents the number of MTCD-to-eNB links that occupy u ARs. The total number of allocated ARs of the eNB should also be subjected to $\sum_{u=1}^{U}(us_{m,u}) \leq K$;
- $\boldsymbol{s}_i$: a vector of $s_{i,w}$, where $w \in \{1,\ldots,W\}$. $s_{i,w}$ is the number of the wireless links between the ith MTCG and the eNB that occupy w BRs;
- $\boldsymbol{s}_h$: a vector of $s_{h,v}$, where $v \in \{1,\ldots,V\}$. $s_{h,v}$ indicates the number of HTCD-to-eNB links that occupy v BRs. Thus, the total number of allocated BRs of the eNB should satisfy $\sum_{i=1}^{N}\left(\sum_{w=1}^{W}(ws_{i,w})\right) + \sum_{v=1}^{V}(vs_{h,v}) \leq L$;
- e: An event in the event set $\mathcal{E}$, i.e., $e \in \mathcal{E}$.

The event set $\mathcal{E}$ is denoted by

$$\mathcal{E} = \{A_m, A_h\} \cup \mathcal{D}_1 \cup \mathcal{D}_2 \cup \ldots \cup \mathcal{D}_N \cup \mathcal{D}_m \cup \mathcal{D}_h, \tag{3}$$

where $\mathcal{D}_i, \mathcal{D}_m, \mathcal{D}_h$, $i \in \{1,\ldots,N\}$ are defined as

$$\begin{aligned} \mathcal{D}_i &= \{D_{i,u,w} \mid u \in \{1,...,U\},\ w \in \{1,...,W\}\}, \\ \mathcal{D}_m &= \{D_{m,u} \mid u \in \{1,...,U\}\}, \\ \mathcal{D}_h &= \{D_{h,v} \mid v \in \{1,...,V\}\}. \end{aligned} \tag{4}$$

Each event is detailed as follows

- A_m: the network receives a data packet from an MTCD. A_m can be further denoted as $A_m = (\mu_1, \mu_2, \ldots, \mu_N, \mu_{N+1})$, where $\{\mu_i\}_{i=1}^{N}$ represent the transmission rate of the wireless link between the MTCD and the ith MTCG; μ_{N+1} is the transmission rate of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

MTCD-to-eNB link; and $\{\mu_i\}_{i=1}^{N+1} \in \{\{\mu_{m,i}\}_{i=1}^{l_m}\}$. The transmission rate of each wireless link is evaluated by the SDN controller according to the SNR;

- A_h: the network receives a data packet from a HTCD. A_h can be further denoted as $A_h = \mu_h$, where μ_h is the transmission rate of the HTCD-to-eNB link, and $\mu_h \in \{\{\mu_{h,i}\}_{i=1}^{l_h}\}$;
- $\mathcal{D}_i$: an MTC data packet departs from the ith MTCG, where $D_{i,u,w}$ denotes the departure of an MTC data packet that occupies u ARs and w BRs;
- $\mathcal{D}_m$: an MTC data packet departs from the eNB, where $D_{m,u}$ represents the departure of an MTC data packet that occupies u ARs;
- $\mathcal{D}_h$: a HTC data packet departs from the eNB, where $D_{h,v}$ denotes the departure of a HTC data packet that occupires v BRs.

2) Actions: When an event e occurs, the SDN controller excutes an action $a(s)$ from the action set $\mathcal{A}_s$ according to the current state s of the network, i.e.,$a(s) \in \mathcal{A}_s$. The action set can be described by

$$\mathcal{A}_s = \begin{cases} \{(-1,0,0),(i,u,w),(N+1,u,0)\}, & e = A_m \\ \{(-2,0,0),(N+2,0,v)\}, & e = A_h \\ \{(0,0,0)\}, & e \in \mathcal{E} \setminus \{A_m, A_h\}, \end{cases} \quad (5)$$

where $\mathcal{E} \setminus \{A_m, A_h\}$ means the relative complement of $\{A_m, A_h\}$ in $\mathcal{E}$, i.e., $\mathcal{D}_1 \cup \mathcal{D}_2 \cup \ldots \cup \mathcal{D}_N \cup \mathcal{D}_m \cup \mathcal{D}_h$. As can be seen from (5), action $a(s)$ is represented by a triplet, where the first element denotes the transmission strategy of a data packet (reject, transmit via an MTCG or the eNB), and the second and third elements denote the allocation of ARs and BRs, respectively. The details of each action is given as follows

- When $e = A_m$, three types of actions can be chosen from set $\mathcal{A}_s$, i.e., reject, transmitting through an MTCG or the eNB. When 'reject' is chosen, no radio resources would be allocated, and thus $a(s) = (-1,0,0)$; When an MTC packet is transmitted to the ith MTCG with u ARs and w BRs allocated, action $a(s) \in \{(i,u,w)\}$, where $i \in \{1,\ldots,N\}$, $u \in \{1,\ldots,U\}$, and $w \in \{1,\ldots,W\}$; Similarly, when an MTC packet is transmitted to the eNB with u ARs allocated, action $a(s) \in \{(N+1,u,0)\}$;
- When $e = A_h$, the HTC data packet can be either rejected or accepted by the network. The rejection of a HTC packet is denoted by $a(s) = (-2,0,0)$, and the acceptance of a HTC packet with v BRs allocated is described by $a_s \in \{(N+2,0,v)\}$, $v \in \{1,\ldots,V\}$; and
- When $e \in \mathcal{E} \setminus \{A_m, A_h\}$, i.e., the transmission of an MTC or HTC data packet is completed, and no other action is required to be taken by the SDN controller except for updating the status information from the eNB and MTCGs, which is described by $a(s) = (0,0,0)$.

3) Rewards: According to the system state s and the corresponding action a, the system reward can be given by

$$r(s,a) = k(s,a) - \omega(s,a), \quad (6)$$

where $k(s,a)$ is the lump sum income of the system by taking action a in state s, and $\omega(s,a)$ is the expected system cost. $k(s,a)$ can be further defined as follows

$$k(s,a) = \begin{cases} E_2 - \beta t_h, & e = A_h, a(s) = (N+2,0,v) \\ E_1 - \beta t_m, & e = A_m, a(s) = (N+1,u,0) \\ E_0 - \beta t_g, & e = A_m, a(s) = (i,u,w) \\ -P_1, & e = A_m, a(s) = (-1,0,0) \\ -P_2, & e = A_h, a(s) = (-2,0,0) \\ 0, & e \in \mathcal{E} \setminus \{A_m, A_h\}, \\ & a(s) = (0,0,0). \end{cases} \quad (7)$$

When a HTC data packet is accepted by the eNB, the network earns an income of E_2. Meanwhile, the HTC data packet also consumes βt_h resources for occupying v BRs of the eNB during transmission, where t_h denotes the transmission time and β refers to the price per unit time with the same measurement unit as the income. Therefore, $k(s,a) = E_2 - \beta t_h$, when $e = A_h$ and $a(s) = (N+2,0,v)$. Considering that the channel quality and the number of allocated resources have an impact on the transmission time, t_h can be rewritten as

$$t_h = \frac{\delta_h}{v\mu_h}, \quad (8)$$

where δ_h is the length of the HTC packet, which follows an exponential distribution with a mean size of ρ_h.

For MTC data packets, the lump sum income of the system can be derived similarly to HTC data packets. When a data packet is transmitted to the eNB, the network earns E_1 and pays βt_m for the occupation of u ARs of the eNB. On the other hand, when an MTC packet is relayed by an MTCG, the network earns E_0 and expends βt_g for allocating u ARs on the MTCD-to-MTCG link and w BRs on the MTCG-to-eNB link. In these cases, t_m and t_g can be rewritten as

$$\begin{aligned} t_m &= \frac{\delta_m}{u\mu_{N+1}}, \\ t_g &= \frac{\delta_m}{u\mu_m} + \frac{\delta_m}{w\mu_g}, \end{aligned} \quad (9)$$

where δ_m is the length of the MTC packet which follows an exponential distribution with a mean size of ρ_m; μ_m and μ_g denote the transmission rates of the MTCD-to-MTCG link and MTCG-to-eNB link, respectively, which are decided by whichever MTCG is selected by the SDN controller according to the system revenue. When the MTC data packet is forwarded by the ith MTCG, we have $\mu_m = \mu_i$, and μ_g can be derived by $C_h(\xi)$ according to the SNR ξ of the wireless link between the ith MTCG and the eNB.

When an MTC or HTC data packet is rejected, the network receives a penalty of $(-P_1)$ or $(-P_2)$, respectively. Moreover, the network receives no income when a data packet transmission is completed, i.e., $k(s,a) = 0$ for $a(s) = (0,0,0)$ and $e \in \mathcal{E} \setminus \{A_m, A_h\}$.

On the other hand, the expected system cost $\omega(s,a)$ given in (6) can be defined as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

$$\omega(s,a) = o(s,a)\tau(s,a), \quad a(s) \in \mathcal{A}_s \tag{10}$$

where $\tau(s,a)$ is the expected time duration between two continuous decision epochs; $o(s,a)$ is the cost rate of the system, which is determined by the total number of occupied ARs and BRs

$$\begin{aligned} o(s,a) = &\sum_{i=1}^{N}\Big(\sum_{u=1}^{U} u g_{i,u} + \sum_{w=1}^{W} w s_{i,w}\Big) \\ &+ \sum_{u=1}^{U} u s_{m,u} + \sum_{v=1}^{V} v s_{h,v}. \end{aligned} \tag{11}$$

III. Solution to the SMDP problem

In this section, we first present the state transition probability that has a significant effect on deriving the optimal policy. Then, an average reward criterion is used as the performance criterion, since we focus primarily on the long term performance of the network. Our main objective is to maximize the expected average reward of the system by making optimal decisions at decision epochs. Finally, a relative value iteration algorithm is utilized to obtain the optimal policy.

A. Transition probability

An action taken by the SDN controller causes the state transition, which is characterized by the state transition probability. To obtain the transition probability, one should first derive the mean rate of events. When an action a is selected at the current state s, the system will transit to the next state j before the next decision epoch. The interval between two continuous decision epochs is denoted by $\tau(s,a)$, which is given in (10). Therefore, the mean rate $\gamma(s,a)$ of events for a given s and a is the sum rate of all events in the system, which is the reciprocal of $\tau(s,a)$. To ease of exposition, the following two mathematical operators $\mathbf{1}_{x,y}$ and $\bar{\mathbf{1}}_{x,y}$ are introduced

$$\mathbf{1}_{x,y} = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{if } x \neq y, \end{cases} \quad \bar{\mathbf{1}}_{x,y} = \begin{cases} 0, & \text{if } x = y \\ 1, & \text{if } x \neq y. \end{cases} \tag{12}$$

Accordingly, the mean rate $\gamma(s,a)$ of events can be expressed as

$$\begin{aligned} &\gamma(s,a) = \tau^{-1}(s,a) \\ &= \begin{cases} \gamma_0(s,a), & e \in \mathcal{E} \setminus \{A_m, A_h\} \\ & e = A_m, a(s) = (-1,0,0) \\ & e = A_h, a(s) = (-2,0,0) \\ \gamma_0(s,a) + \frac{1}{\rho_h} v \bar{\mathrm{C}}_h, & e = A_h, a(s) = (N+2,0,v) \\ \gamma_0(s,a) + \frac{1}{\rho_m} u \bar{\mathrm{C}}_m, & e = A_m, a(s) = (N+1,u,0) \\ \gamma_0(s,a) & e = A_m, a(s) = (i,u,w), \\ + \frac{1}{\rho_m} \bar{\mathbf{1}}_{s'_{i,w},0} u \bar{\mathrm{C}}_m & \\ + \frac{1}{\rho_m} \bar{\mathbf{1}}_{g'_{i,u},0} w \bar{\mathrm{C}}_g, & \end{cases} \end{aligned} \tag{13}$$

where $\gamma_0(s,a)$ can be further denoted by

$$\begin{aligned} \gamma_0(s,a) &= M\lambda_m + H\lambda_h \\ &+ \frac{1}{\rho_m} \sum_{i=1}^{N}\sum_{u=1}^{U}\sum_{w=1}^{W} \Big(\bar{\mathbf{1}}_{s'_{i,w},0} g_{i,u} u \bar{\mathrm{C}}_m + \bar{\mathbf{1}}_{g'_{i,u},0} s_{i,w} w \bar{\mathrm{C}}_g\Big) \\ &+ \frac{1}{\rho_m} \sum_{u=1}^{U} s_{m,u} u \bar{\mathrm{C}}_m + \frac{1}{\rho_h} \sum_{v=1}^{V} s_{h,v} v \bar{\mathrm{C}}_h, \end{aligned} \tag{14}$$

where $g_{i,u}$ and $s_{i,w}$ are the elements in the current state s, which are given in (2); $g'_{i,u}$ and $s'_{i,w}$ are the corresponding values of $g_{i,u}$ and $s_{i,w}$ in the next state j, respectively, which are affected by the action of the current state $a(s)$.

When a data packet transmission is completed, or a data packet is rejected by the network, the mean rate of events $\gamma(s,a)$ is equal to $\gamma_0(s,a)$ as shown in (14), where $M\lambda_m$ and $H\lambda_h$ are the total arrival rates of MTC and HTC data packets, respectively. Moreover, the reset part of $\gamma_0(s,a)$ is the departure rate of the system. Since the long term performance of the network is mainly concerned, we take the average transmission rate as the capacity for each wireless link. Denote by $\bar{\mathrm{C}}_m$ the average transmission rate of the MTCD-to-MTCG and MTCD-to-eNB links; $\bar{\mathrm{C}}_g$ and $\bar{\mathrm{C}}_h$ the average transmission rates of the MTCG-to-eNB links and HTCD-to-eNB links, respectively. $\bar{\mathrm{C}}_m$ and $\bar{\mathrm{C}}_h$ are defined as follows [24]

$$\begin{aligned} \bar{\mathrm{C}}_m &= \sum_{i=1}^{l_m} \mathrm{P}_m(i)\mu_{m,i}, \\ \bar{\mathrm{C}}_h &= \sum_{i=1}^{l_h} \mathrm{P}_h(i)\mu_{h,i}, \end{aligned} \tag{15}$$

where $\mathrm{P}_m(i)$ and $\mathrm{P}_h(i)$ are the probabilities of applying the transmission mode i of $\mathrm{C}_m(\cdot)$ and $\mathrm{C}_h(\cdot)$, respectively. To simplify the analysis, it is assumed that each transmission mode is chosen with an equal probability, i.e., $\mathrm{P}_m(i) = 1/l_m$ and $\mathrm{P}_h(i) = 1/l_h$. $\bar{\mathrm{C}}_g$ is compute through averaging the transmission rates of the wireless links between N MTCGs and the eNB.

When a HTC data packet is accepted by the network, the eNB allocates v more radio resources for its transmission, and thus $\gamma(s,a)$ is larger than $\gamma_0(s,a)$ by $\frac{1}{\rho_h} v \bar{\mathrm{C}}_h$. Similarly, the mean rate of events is equal to $\gamma_0(s,a) + \frac{1}{\rho_m} u \bar{\mathrm{C}}_m$, when an MTC data packet is transmitted to the eNB with u resources occupied. For the MTC data packet forwarded by the MTCG, the mean rate of events is equal to $\gamma_0(s,a) + \frac{1}{\rho_m} \bar{\mathbf{1}}_{s'_{i,w},0} u \bar{\mathrm{C}}_m + \frac{1}{\rho_m} \bar{\mathbf{1}}_{g'_{i,u},0} w \bar{\mathrm{C}}_g$.

The state transition probability is represented by $q(j|s,a)$, which means the state is transited from state s to state j under action a. To simplify the expression of $q(j|s,a)$, the following function is defined

$$\begin{aligned} &\mathrm{F}_s(x_1 + m_1, x_2 + m_2, ..., x_n + m_n, e') \\ &\equiv (..., x_1 + m_1, ..., x_2 + m_2, ..., x_n + m_n, ..., e'), \\ &\text{if } s = (..., x_1, ..., x_2, ..., x_n, ..., e), \end{aligned} \tag{16}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

which means that only element x_i in state s is updated to x_i+m_i and other elements remain the same, when the system enters the next state j with new event e'. Therefore, the state transitions for different actions can be obtained as follows

- For $e = A_m$ in the current state s, $q(j|s,a)$ can be obtained via (18)∼(19), which are categorized by the available action a of the current state s. When $a(s) = (-1,0,0)$, $q(j|s,a)$ can be calculated by (19), while (17) and (18) are used for the cases of $a = (N+1,u,0)$ and $a = (i,u,w)$, respectively. Each transition probability in each formula is calculated in the same way. That is, the mean rate of all events $\gamma(s,a)$ divides the rate of the event e with action a. When the event e' in the next state j is A_m, coefficient $1/l_m^{(N+1)}$ is introduced due to the assumption that $\{\mu_i\}_{i=1}^{N+1}$ of A_m adopt each transmission mode with the same probability. For the same reason, the coefficient $1/l_h$ is also used for cases where $e' = A_h$. In addition, the symbols in (18)∼(19) are explained as: $i, i_1 \in \{1,\dots,N\}$ are the indexes of the MTCG; $u, u_1 \in \{1,\dots,U\}$ denote the number of ARs; $w, w_1 \in \{1,\dots,W\}$ represent the number of BRs for the MTCG-to-eNB link, and $v, v_2 \in \{1,\dots,V\}$ are the resource allocated for the HTCD-to-eNB link;

$$q(j|s,a)\mid_{a=(N+1,u,0)}=\begin{cases}\frac{1}{l_m^{(N+1)}}\cdot\frac{M\lambda_m}{\gamma(s,a)}, & j=\mathrm{F}_s(s_{m,u}+1,A_m)\\ \frac{1}{l_h}\cdot\frac{H\lambda_h}{\gamma(s,a)}, & j=\mathrm{F}_s(s_{m,u}+1,A_h)\\ \frac{g_{i,u_1}u_1\bar{\mathrm{C}}_m+s_{i,w}w\bar{\mathrm{C}}_g}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(s_{m,u}+1,g_{i,u_1}-1,\\ & \qquad s_{i,w}-1,D_{i,u_1,w})\\ \frac{(s_{m,u_1}+\mathbf{1}_{u,u_1})u_1\bar{\mathrm{C}}_m}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(s_{m,u}+1,s_{m,u_1}-1,D_{m,u_1})\\ \frac{s_{h,v}v\bar{\mathrm{C}}_h}{\rho_h\gamma(s,a)}, & j=\mathrm{F}_s(s_{m,u}+1,s_{h,v}-1,D_{h,v})\end{cases}\tag{17}$$

- For $e = A_h$ in the current state s, $q(j|s,a)$ can also be obtained by (19) with $a(s) = (-2,0,0)$. When $a(s) = (N+2,0,v)$, $q(j|s,a)$ can be calculated by (20), which is derived similarly to (18)∼(19). Moreover, each symbol in (20) is the same as in (18)∼(19);

$$q(j|s,a)\mid_{a(s)=(N+2,0,v)}=\begin{cases}\frac{1}{l_m^{(N+1)}}\cdot\frac{M\lambda_m}{\gamma(s,a)}, & j=\mathrm{F}_s(s_{h,v}+1,A_m)\\ \frac{1}{l_h}\cdot\frac{H\lambda_h}{\gamma(s,a)}, & j=\mathrm{F}_s(s_{h,v}+1,A_h)\\ \frac{g_{i,u}u\bar{\mathrm{C}}_m+s_{i,w}w\bar{\mathrm{C}}_g}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(s_{h,v}+1,g_{i,u}-1,\\ & \qquad s_{i,w}-1,D_{i,u,w})\\ \frac{s_{m,u}u\bar{\mathrm{C}}_m}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(s_{h,v}+1,s_{m,u}-1,D_{m,u})\\ \frac{(s_{h,v_1}+\mathbf{1}_{v,v_1})v_1\bar{\mathrm{C}}_h}{\rho_h\gamma(s,a)}, & j=\mathrm{F}_s(s_{h,v}+1,s_{h,v_1}-1,D_{h,v_1})\end{cases}\tag{20}$$

$$q(j|s,a)\mid_{a=(i,u,w)}=\begin{cases}\frac{1}{l_m^{(N+1)}}\cdot\frac{M\lambda_m}{\gamma(s,a)}, & j=\mathrm{F}_s(g_{i,u}+1,s_{i,w}+1,A_m)\\ \frac{1}{l_h}\cdot\frac{H\lambda_h}{\gamma(s,a)}, & j=\mathrm{F}_s(g_{i,u}+1,s_{i,w}+1,A_h)\\ \frac{g_{i_1,u_1}u_1\bar{\mathrm{C}}_m+s_{i_1,w_1}w_1\bar{\mathrm{C}}_g}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(g_{i,u}+1,g_{i_1,u_1}-1,\\ & \qquad s_{i,w}+1,s_{i_1,w_1}-1,D_{i_1,u_1,w_1}),\\ & \qquad i\neq i_1\\ \frac{(g_{i,u_1}+\mathbf{1}_{u,u_1})u_1\bar{\mathrm{C}}_m}{\rho_m\gamma(s,a)} & j=\mathrm{F}_s(g_{i,u}+1,g_{i,u_1}-1,\\ +\frac{(s_{i,w_1}+\mathbf{1}_{w,w_1})w_1\bar{\mathrm{C}}_g}{\rho_m\gamma(s,a)}, & \qquad s_{i,w}+1,s_{i,w_1}-1,D_{i,u_1,w_1})\\ \frac{s_{m,u_1}u_1\bar{\mathrm{C}}_m}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(g_{i,u}+1,s_{i,w}+1,\\ & \qquad s_{m,u_1}-1,D_{m,u_1})\\ \frac{s_{h,v}v\bar{\mathrm{C}}_h}{\rho_h\gamma(s,a)}, & j=\mathrm{F}_s(g_{i,u}+1,s_{i,w}+1,\\ & \qquad s_{h,v}-1,D_{h,v})\end{cases}\tag{18}$$

$$q(j|s,a)\mid_{a(s)\in\{(-1,0,0),(-2,0,0),(0,0,0)\}}=\begin{cases}\frac{1}{l_m^{(N+1)}}\cdot\frac{M\lambda_m}{\gamma(s,a)}, & j=\mathrm{F}_s(A_m)\\ \frac{1}{l_h}\cdot\frac{H\lambda_h}{\gamma(s,a)}, & j=\mathrm{F}_s(A_h)\\ \frac{g_{i,u}\cdot u\bar{\mathrm{C}}_m+s_{i,w}\cdot w\bar{\mathrm{C}}_g}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(g_{i,u}-1,s_{i,w}-1,D_{i,u,w})\\ \frac{s_{m,u}u\bar{\mathrm{C}}_m}{\rho_m\gamma(s,a)}, & j=\mathrm{F}_s(s_{m,u}-1,D_{m,u})\\ \frac{s_{h,v}v\bar{\mathrm{C}}_h}{\rho_h\gamma(s,a)}, & j=\mathrm{F}_s(s_{h,v}-1,D_{h,v})\end{cases}\tag{19}$$

- For $e \in \mathcal{E}\setminus\{A_m,A_h\}$ in the current state s, the action $a(s) = (0,0,0)$ and $q(j|s,a)$ can be also obtained by (19).

B. Average reward model

The time duration between two continuous decision epochs $\tau(s,a)$ follows an exponential distribution, whose cumulative distribution function (CDF) is shown below

$$F(t|s,a) = 1 - e^{-\gamma(s,a)t}, \quad t > 0. \tag{21}$$

Therefore, the expected average reward $r(s,a)$ during time $\tau(s,a)$ can be calculated based on the average reward model defined in [25], i.e.,

$$\begin{aligned} r(s,a) &= k(s,a) - o(s,a)E_s^a\{\tau\}\\ &= k(s,a) - \frac{o(s,a)}{\gamma(s,a)}. \end{aligned}\tag{22}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

C. Solution

The average reward of policy π is defined as

$$\psi^{\pi} = \lim_{N\to\infty} \frac{E^{\pi}\left\{\sum_{n=1}^{N} r(s_n, a_n)\right\}}{E^{\pi}\left\{\sum_{n=1}^{N} \tau(s_n, a_n)\right\}}, \tag{23}$$

where s_n and a_n refer to the state and action at the decision epoch n, respectively. The optimal reward ψ^* is denoted as

$$\psi^* = \psi^{\pi^*} = \sup_{\pi\in\Pi} \psi^{\pi}, \tag{24}$$

where Π is the set of any feasible policy π. Moreover, π^* is the optimal policy that can be derived by solving the Bellman equation, i.e.,

$$\nu(s) = \max_{a\in\mathcal{A}_s}\{r(s,a) - \psi\tau(s,a) + \sum_{j\in\mathcal{S}} q(j|s,a)\nu(j)\},\ s\in\mathcal{S} \tag{25}$$

where $\nu(s)$ is the potential function of state s. Then, the uniformization transformation is applied to transform the average reward SMDP to a discrete-time model so as to simplify the analysis [25]. To realize uniformization, parameter η is introduced and defined as

$$\begin{aligned}\eta \equiv M\lambda_m + H\lambda_h + K(N+1)\frac{\mu_{m,l_m}}{\rho_m} \\ + L\max\left\{\frac{\mu_g}{\rho_m}, \frac{\mu_{h,l_h}}{\rho_h}\right\},\quad \eta < \infty.\end{aligned} \tag{26}$$

Thus, the uniformed reward function $\widetilde{r}(s,a)$ and uniformed transition probability $\widetilde{q}(j|s,a)$ are obtained as

$$\widetilde{r}(s,a) = r(s,a)\frac{\gamma(s,a)}{\eta}, \tag{27}$$

$$\widetilde{\psi} = \frac{\psi}{\eta}, \tag{28}$$

$$\widetilde{q}(j|s,a) = \begin{cases} 1 - \frac{[1-q(s|s,a)]\gamma(s,a)}{\eta}, & j = s \\ \frac{q(j|s,a)\gamma(s,a)}{\eta}, & j \neq s \end{cases}. \tag{29}$$

Therefore, the Bellman equiation can be rewritten as

$$\widetilde{\nu}(s) = \max_{a\in\mathcal{A}_s}\{\widetilde{r}(s,a) - \widetilde{\psi} + \sum_{j\in\mathcal{S}} \widetilde{q}(j|s,a)\widetilde{\nu}(j)\},\ s\in\mathcal{S}. \tag{30}$$

Since the proposed SMDP model has finite state and action spaces, a relative value iteration algorithm can be applied to solve the Bellman equation as shown in **Algorithm 1**, which often offers a much faster rate of convergence with respect to the span seminorm [25]. Let $\Phi(\boldsymbol{\nu})$ be the span of vector $\boldsymbol{\nu}$, which is defined as follows

$$\Phi(\boldsymbol{\nu}) = \max_{s\in\mathcal{S}} \nu(s) - \min_{s\in\mathcal{S}} \nu(s). \tag{31}$$

For $\boldsymbol{\nu}\in\mathcal{V}$, $\Phi(\boldsymbol{\nu})$ is a seminorm on $\mathcal{V}$. In the result, the relative value iteration algorithm can obtain a vector of decision rules $d_\varepsilon(s)$ that constitute the optimal policy π^*.

Algorithm 1 Relative Value Iteration Algorithm

Step 1. Initialization:
- Select $\widetilde{\boldsymbol{\nu}}^0 \in \mathcal{V}$.
- Choose a base state $s^* \in \mathcal{S}$, and specify $\varepsilon > 0$.
- Set $\boldsymbol{w}^0 = \widetilde{\boldsymbol{\nu}}^0 - \widetilde{\nu}^0(s^*)\boldsymbol{e}$, where $\boldsymbol{e}$ is a vector of ones.
- Set $n = 0$.

Step 2. Set

$$\widetilde{\boldsymbol{\nu}}^{n+1} = \max_{a\in\mathcal{A}_s}\{\boldsymbol{r}_a + \mathbf{P}_a\boldsymbol{w}^n\},$$

$$\boldsymbol{w}^{n+1} = \widetilde{\boldsymbol{\nu}}^{n+1} - \widetilde{\nu}^n(s^*)\boldsymbol{e},$$

where $\boldsymbol{r}_a$ is the vector of $\widetilde{r}(s,a)$; $\mathbf{P}_a$ is the transition probability matrix under action a, which consists of $\widetilde{q}(j|s,a)$.

Step 3. If $\Phi(\widetilde{\boldsymbol{\nu}}^{n+1} - \widetilde{\boldsymbol{\nu}}^n) < \varepsilon$, go to **Step 4**. Otherwise, $n = n+1$ and return to **Step 2**.

Step 4. Choose $d_\varepsilon \in \arg\max_{a\in\mathcal{A}_s}\{\boldsymbol{r}_a + \mathbf{P}_a\boldsymbol{\nu}^n\}$.

IV. Performance Evaluation

To evaluate the performance of our proposed radio resource allocation scheme for the SDN-based IoT network, we develop a Matlab simulator. In this section, we first introduce the simulation environment and then analyze the numerical results.

In the simulation, we consider an SDN-based IoT network that consists of an eNB, an MTCG, and an SDN controller. There are also 200 MTCDs and 20 HTCDs evenly scattered in the serving area of the network. Each MTCG and the eNB contains up to $K = 3$ ARs, and the eNB also owns $L = 5$ BRs. The maximum number of ARs allocated to each MTCD-to-eNB and MTCD-to-MTCG wireless link is $U = 2$. That is, a wireless link can be assigned 1 or 2 units of ARs based on the decision taken by the SDN controller. Moreover, the eNB can allocate $W = 2$ and $V = 2$ BRs at most to each MTCG-to-eNB and HTCD-to-eNB link, respectively. We assume that the MTCD-to-MTCG and MTCD-to-eNB links support $l_m = 2$ transmission modes with rates $\mu_{m,1} = 10$ and $\mu_{m,2} = 20$. The MTCG-to-eNB and HTCD-to-eNB links also support $l_h = 2$ transmission modes, which provide higher transmission rates, i.e., $\mu_{h,1} = 100$ and $\mu_{h,2} = 200$. In order to expedite the execution time and to reduce memory usage, the aforementioned system parameters are configured in a small scale, which helps reduce the number of states in the state space as well as the number of feasible actions corresponding to the states. This will not affect the generality and accuracy of our simulations.

Moreover, the average sizes of the MTC and HTC packets are $\rho_m = 5$ and $\rho_h = 50$, respectively. Each HTCD transmits data packets to the network with an arrival rate of $\lambda_h = 0.1$. To further study the relationship among the parameters, the arrival rate of the MTC data packets λ_m varies from 0.001 to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

0.1 for performance comparison, which is specified in small values due to infrequent M2M communications traffic. The other parameters used in the simulation are summarized in Table I.

TABLE I. SIMULATION PARAMETERS.

Parameter	Value	Parameter	Value
N	1	λ_m	$0.001 - 0.1$
M	200	λ_h	0.1
H	20	ρ_m	5
K	3	ρ_h	50
L	5	E_0	10
U	2	E_1	10
V	2	E_2	10
W	2	P_1	10
β	10	P_2	5
$\mu_{h,1}$	100	$\mu_{m,1}$	10
$\mu_{h,2}$	200	$\mu_{m,2}$	20

For comparison purposes, we will also benchmark the proposed scheme against the following comparative schemes:

- **Greedy scheme:** The algorithm that aims at maximizing the system current reward at the decision epoch;
- **Channel precedence (CP) scheme:** a heuristic algorithm that is based on the channel quality of the wireless links between the MTCD and the candidate eNB or MTCGs. In general, the wireless link with a better channel quality can provides a higher transmission rate. The SDN controller chooses the eNB or MTCG that can provide the highest transmission rate with enough spare resources for MTC packet transmission, which is always allocated as much resources as possible; and
- **eNB precedence (EP) scheme:** When there are spare ARs available in the eNB, the SDN controller gives priority to selecting the eNB for MTC data packet transmission, which is always allocated with the maximal number of ARs that the eNB can support. Otherwise, the SDN controller makes the decision that maximizes the system current reward in the decision epoch.

Furthermore, several metrics are specified to evaluate the performance of our proposed scheme. The first metric is the blocking rate of data packets, which represents the probability of data packets rejected by the network. The second metric is the expected average reward of the network defined in (22).

Before compare the performance of our proposed scheme with those of the other reference schemes, we first present the action probabilities with different arrival rates of MTC data packets. Fig. 2 shows that the SDN controller is more likely to select the eNB for MTC data packet transmission when λ_m is small. This is because the transmission via the eNB is through only one wireless link, which may consume less transmission time and receive a good system reward. With the increase of λ_m, the eNB allocates more ARs for MTC data packet transmission, which results in insufficient ARs in the eNB. Thus, the action probability of selecting the eNB decreases, while the action probability of selecting the MTCG increases. When an MTC data packet is transmitted by the eNB, the probabilities of resource allocation under each possible action (i.e., $a \in \{(2,1,0),(2,2,0)\}$) are depicted in Fig. 3. When λ_m

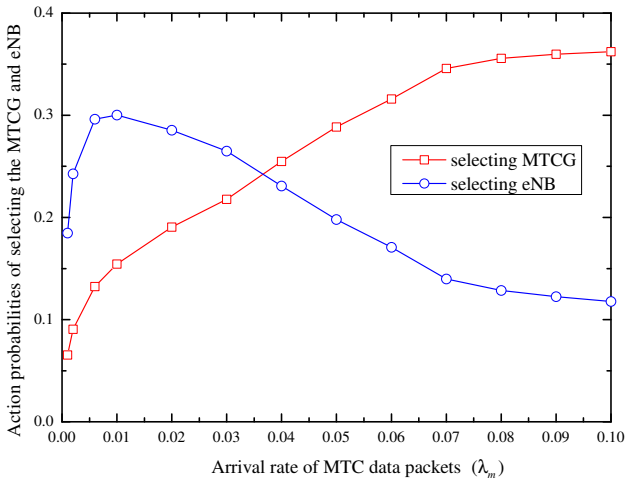

Fig. 2. Action probabilities of selecting the MTCG and eNB with various arrival rates of MTC data packets per MTCD.

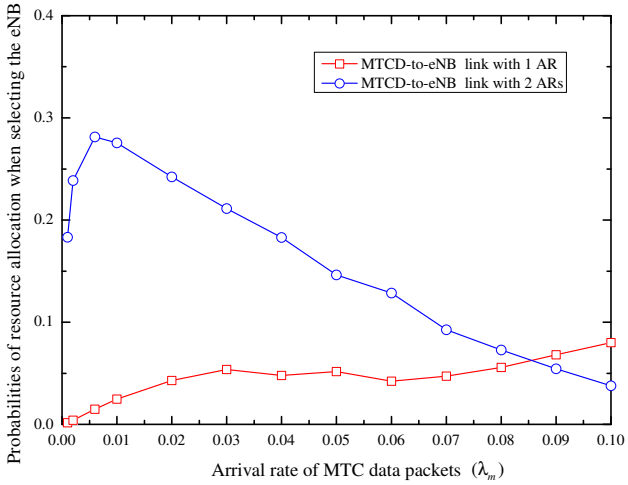

Fig. 3. Probabilities of resource allocation when selecting the eNB for MTC data packet transmission with various arrival rates of MTC data packets per MTCD.

is small, the eNB tends to allocate two ARs for the MTCD-to-eNB links, while one AR is more likely to be allocated to the MTCD-to-eNB links with the increase of λ_m. This is because there are sufficient ARs in the eNB, which allocates the most number of ARs to maximize the system reward with a high probability. When an MTC data packet is forwarded via the MTCG, the probabilities of resource allocation to the MTCD-to-MTCG and MTCG-to-eNB links with various actions are shown in Fig. 4. With the increase of λ_m, almost all the curves increase except for the one that allocates two ARs to the MTCD-to-MTCG links, which starts decreasing slightly around the point where $\lambda_m = 0.07$. Moreover, it can be seen that the MTCD-to-MTCG and MTCG-to-eNB links are allocated two BRs and two ARs with high probabilities, respectively. This is because the network load is not high enough within the range of $\lambda_m \in [0.001, 0.1]$, and the MTCG

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

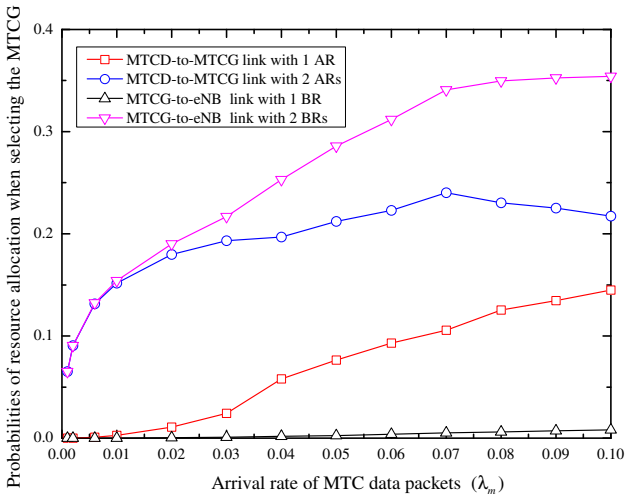

Fig. 4. Probabilities of resource allocation when selecting the MTCG for MTC data packet transmission with various arrival rates of MTC data packets per MTCD.

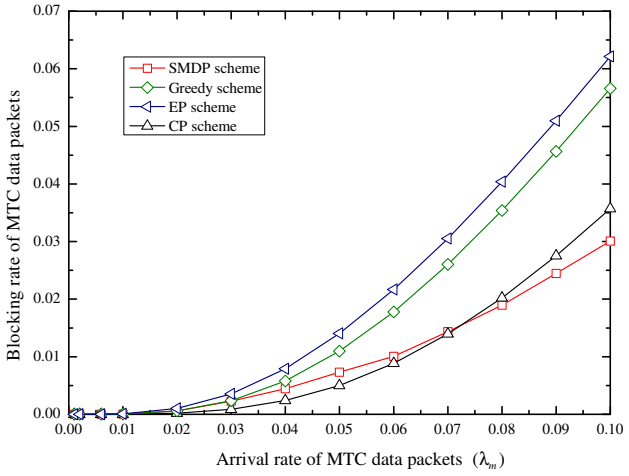

Fig. 5. Blocking rate of MTC data packets with various arrival rates of MTC data packets per MTCD of each scheme.

still has resources available to accommodate new connections.

Then, we compare the blocking rates of the proposed SMDP-based model and other reference schemes. Fig. 5 plots the blocking rate of MTC data packets with various arrival rates of MTC data packets. When λ_m is very small, the blocking rates of MTC data packets of all the comparative schemes are close to zero. This is because there are sufficient ARs and BRs in the MTCG and eNB to be allocated for the wireless links, and the network seldom rejects MTC data packets. With the increase of λ_m, the MTCG and eNB have to allocate ARs and BRs to more wireless links so as to support more MTC data packet transmissions, which consumes available network resources resulting in a gradually increased blocking rate of MTC data packets. Among all the comparative schemes, the EP scheme has the highest blocking rate, since

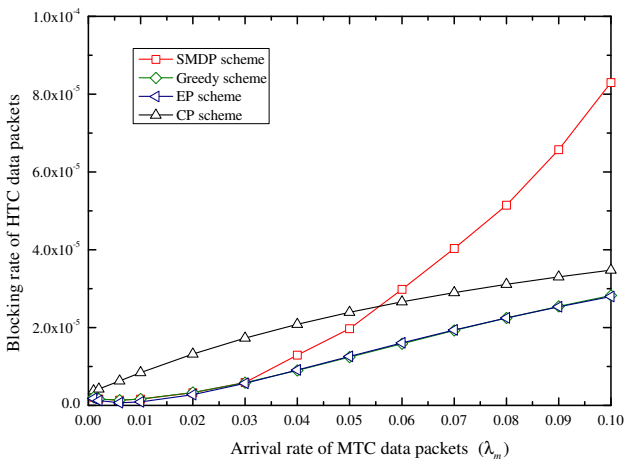

Fig. 6. Blocking rate of HTC data packets with various arrival rates of MTC data packets per MTCD of each scheme.

the EP scheme always first chooses the eNB for MTC data packet transmission, and then allocates ARs to the transmission as much as possible. This may exhaust the network resources rapidly resulting in insufficient spare resources for new arriving packets. For the same reason, the EP scheme also offers an inferior performance in terms of the MTC blocking rate. Compared to these two schemes, the proposed SMDP scheme and the CP scheme have better performances. The MTC blocking rate curves of the SMDP scheme and CP scheme cross at the point around $\lambda_m = 0.07$. When λ_m is smaller than the crossing point, the CP scheme has a lower blocking rate than the SMDP scheme under the low network load. This is attributable to the assumption that each wireless link adopts different transmission rates with the same probability, which makes the CP scheme select the MTCG or eNB equally likely. Therefore, the CP scheme has an effect on balancing the network load, and can reduce the possibility of network congestion. However, when λ_m is larger than the crossing point, our proposed SMDP scheme outperforms not only the CP scheme, but also all the other reference schemes. Moreover, the advantage of the SMDP scheme becomes more evident under a heavier traffic load (i.e, with the continuous increase of λ_m). The reason is that the SMDP scheme aims to optimize resource allocation based on the long-term average reward, which tries to avoid the rejection of MTC data packets that may penalize the system reward.

On the other hand, the blocking rates of HTC data packets with various arrival rates of MTC data packets are plotted in Fig. 6. The blocking rate of HTC data packets of each scheme varies within a small range, which has little impact on the service quality of HTC users. Therefore, the total blocking rate of data packets has the same trend as that of MTC data packets as can be observed from Fig. 7. Our proposed SMDP scheme has an acceptable performance in terms of the blocking rate, and outperforms nearly all the other reference schemes except the CP scheme when λ_m is small.

We further compare the expected average rewards of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2016.2580677, IEEE Sensors Journal

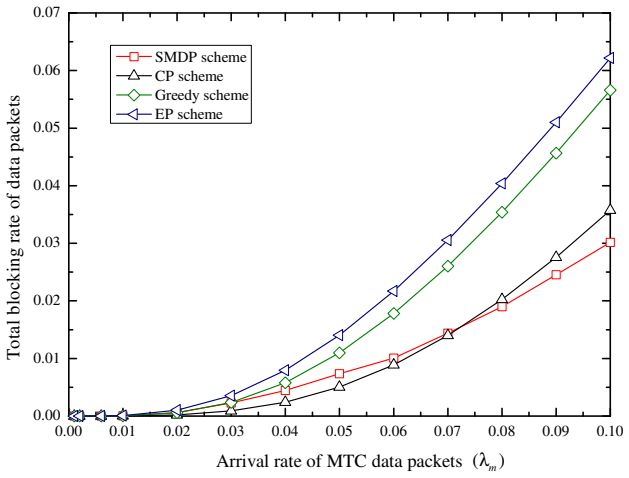

Fig. 7. Total blocking rate of data packets with various arrival rates of MTC data packets per MTCD of each scheme.

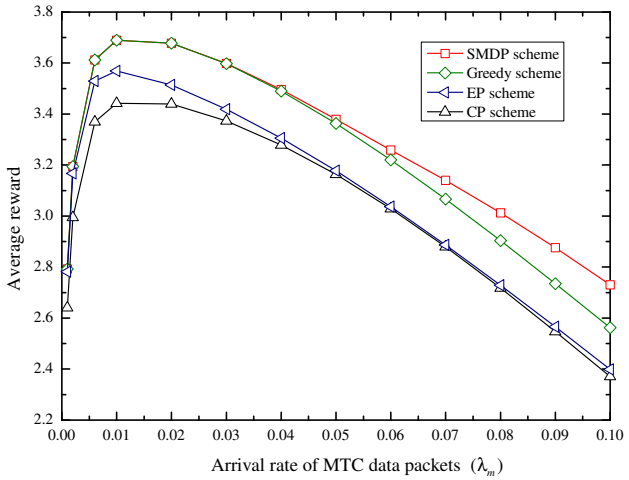

Fig. 8. Average system reward with various arrival rates of MTC data packets per MTCD of each scheme ($\lambda_h = 0.1$).

SMDP-based model and the other reference schemes. The expected average rewards of the network with various arrival rates of MTC data packets are shown in Fig. 8. When λ_m is very low, the MTCG and eNB have a large umber of spare ARs and BRs available for MTCD access, and almost all MTC data packet transmissions are admitted by the network. Therefore, the expected average reward of each scheme increases rapidly with the increase of λ_m. However, when the resources of the MTCG and eNB are nearly depleted (i.e., aroud $\lambda_m = 0.01$), the expected average rewards of the comparative schemes reach their peak value. However, when λ_m continues to increase, the expected average rewards of all the schemes decrease gradually. This is because the network suffers from a heavier traffic load and becomes more likely to reject MTC data packet transmission, which tends to penalize the system reward. Thus, the average reward of the network is a concave function of the arrival rate of MTC data packets. As can be seen from Fig. 8, our proposed scheme significantly outperforms the other reference schemes. When λ_m is less than 0.04, there is no much difference between the SMDP scheme and the greedy scheme, because both schemes tend to adopt the same action. That is, when there are enough resources in the MTCG and eNB, the network tends to allocate resources to the wireless links as much as possible with the objective of achieving a good revenue. However, with the increase of λ_m, the difference between the SMDP scheme and the greedy scheme becomes larger, and the advantage of the SMDP scheme becomes more evident. This is because that the greedy scheme always allocates the maximum number of radio resources to MTC data packet transmission so as to achieve the highest system reward in the decision epoch, without consideration of whether the residual resources are enough for the next request. Thus, the greedy scheme may rapidly exhaust network resources, and the network may risk rejecting new data packets without sufficient radio resources. By contrast, the proposed SMDP-based radio resource allocation model makes decisions by considering both the instant lump sum income and the system expenses. From the long-term perspective, the SMDP scheme can achieve the optimized average reward. The CP and EP schemes do not aim to optimize the system reward. Therefore, they obtain worse expected average rewards.

V. Conclusions and Future Work

In this paper, we investigated the resource allocation problem in the SDN-based IoT network. The problem was formulated as an SMDP process with the objective of maximizing the expected average reward of the network. Then, we obtained the optimal solution to the SMDP problem via a relative value iteration algorithm, which determines which MTCGs or eNB should be selected for MTC packet transmission and how many resources to be allocated for each wireless link. Simulation and numerical results were presented to demonstrate the superiority of our proposed scheme in comparison to the reference schemes.

In our future work, we will analyze the optimal resource allocation policy in the SDN-based IoT network that employs multiple M2M communications technologies to further validate the scalability of our proposed scheme.

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347-2376, 4th Quarter 2015.

[2] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36-43, May 2014.

[3] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Netw.*, vol. 30, no. 1, pp. 44-51, Jan. 2016.

[4] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 184-192, Jul. 2012.

[5] K. Zheng, F. Liu, L. Lei, C. Lin, and Y. Jiang, "Stochastic performance analysis of a wireless finite-state Markov channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 782-793, Feb. 2013.

[6] M. Chen, J. Wan, S. Gonzalez, X. Liao, and V. C. M. Leung, "A survey of recent developments in home M2M networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 98-114, 1st Quarter 2014.

[7] S. Andreev, O. Galinina, A. Pyattaev et al., "Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 32-40, Sep. 2015.

[8] 3GPP RAN Tdoc RP-151621, "New WI proposal: narrow band IOT," approved in RAN plenary #69 meeting, Sep. 2015.

[9] LoRa, https://www.lora-alliance.org/

[10] B. A. A. Nunes, M. Mendonca, Xuan-Nam Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Commun. Surveys Tuts*, vol. 16, no. 3, pp. 1617-1634, 3rd Quarter 2014.

[11] D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14-76, Jan. 2015.

[12] M. Bansal, J. Mehlman, S. Katti, and P. Levis, "Openradio: a programmable wireless dataplane," in *Proc. 1st Workshop Hot Topics Softw. Defined Netw.*, Helsinki, Finland, Aug. 2012, pp. 109-114.

[13] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: Software defined radio access network," in *Proc. 2nd Workshop Hot Topics Softw. Defined Netw.*, Hong Kong, China, Aug. 2013, pp. 25-30.

[14] X. Jin, L. Erran Li, L. Vanbever, and J. Rexford, "SoftCell: Scalable and flexible cellular core network architecture," in *Proc. 9th Int. Conf. Emerging Netw. Exp. Technol.*, Santa Barbara, USA, Dec. 2013, pp. 163-174.

[15] H. Ali-Ahmad et al., "CROWD: An SDN approach for densenets," in *Proc. 2nd Eur. Workshop Softw. Defined Netw.*, Berlin, Germany, Oct. 2013, pp. 25-31.

[16] K. Pentikousis, Y. Wang, and W. Hu, "MobileFlow: Toward software-defined mobile networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 44-53, Jul. 2013.

[17] H. Wang, S. Chen, H. Xu, M. Ai, and Y. Shi, "SoftNet: A software defined decentralized mobile network architecture toward 5G," *IEEE Network*, vol. 29, no. 2, pp. 16-22, Mar. 2015.

[18] K. Zheng, L. Hou, H. Meng, Q. Zheng, N. Lu, and L. Lei, "Soft-defined heterogeneous vehicular network: Architecture and challenges," *IEEE Netw.*, accepted, arXiv preprint arXiv:1510.06579.

[19] K. Sood, S. Yu, and Y. Xiang, "Software defined wireless networking opportunities and challenges for Internet of Things: A review," *IEEE Internet Things J.*, vol., no. 99, pp. 1-1, Sep. 2015.

[20] T. Luo; H.-P. Tan, and T. Q. S. Quek, "Sensor OpenFlow: Enabling software-defined wireless sensor networks," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1896-1899, Nov. 2012.

[21] Z. Qin, G. Denker, C. Giannelli, P. Bellavista, and N. Venkatasubramanian, "A software defined networking architecture for the Internet-of-Things," *IEEE Network Operations and Management Symposium (NOMS)*, pp. 1-9, May 2014.

[22] L. Galluccio, S. Milardo, G. Morabito, and S. Palazzo, "SDN-WISE: Design, prototyping and experimentation of a stateful SDN solution for WIreless SEnsor networks," *IEEE INFOCOM*, Kowloon, Apr. 2015, pp. 513-521.

[23] D. Wu, D. I. Arkhipov, E. Asmare, Z. Qin, and J. A. McCann, "UbiFlow: Mobility management in urban-scale software defined IoT," *IEEE INFOCOM*, Kowloon, Apr. 2015, pp.208-216.

[24] S. Zhang, F. R. Yu, and V. C. M. Leung, "Joint connection admission control and routing in IEEE 802.16-based mesh networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1370-1379, Apr. 2010.

[25] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley, 2005.

[26] K. Zheng, H. Meng, P. Chatzimisios, L. Lei, and X. Shen, "An SMDP-based resource allocation in vehicular cloud computing systems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7920-7928, Dec. 2015.

[27] L. Lei, Z. Zhong, K. Zheng, J. Chen, and H. Meng, "Challenges on wireless heterogeneous networks for mobile cloud computing," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 34-44, Jun. 2013.

[28] L. Lei, Y. Kuang, N. Cheng, X. Shen, Z. Zhong and C. Lin, "Delay-Optimal Dynamic Mode Selection and Resource Allocation in Device-to-Device Communications‖Part II: Practical Algorithm," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3491-3505, May 2016.

BIOGRAPHY

Xiong Xiong received his B.S. degree from Beijing University of Posts and Telecommunications (BUPT), China, in 2013. Since then, he has been working toward a Ph.D. degree at BUPT. His research interests include M2M networks and software defined radio.

Lu Hou received his B.S. degree from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), China, in 2014. He is now a candidate for Ph.D. in the Intelligent Computing and Communication (IC^2) Lab, Key Lab of Universal Wireless Communications, Ministry of Education, BUPT. Nowadays, he mainly focuses on resource allocation and security in mobile cloud computing.

Kan Zheng (S'02-M'06-SM'09) currently a Professor in Beijing University of Posts and Telecommunications (BUPT), China. His current research interests lie in the field of wireless communications, with an emphasis on performance analysis and optimization of heterogeneous networks and 5G networks. He has published more than 200 papers in IEEE conferences and transactions. Dr. Zheng is a also an IET Fellow. He holds editorial board positions for several journals. He has organized several special issues in famous journals including IEEE Communications Surveys & Tutorials, IEEE Communication Magazine, IEEE System Journal, IEEE Network and so on. He has also served in the Organizing/TPC Committees for more than 20 conferences such as IEEE PIMRC'2013, IEEEE SmartGrid'2015 and IEEE ICT'2016.

Wei Xiang (S'00-M'04-SM'10) received the B.Eng. and M.Eng. degrees, both in electronic engineering, from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004.

He is currently Foundation Professor and Head of Electronic Systems and Internet of Things Engineering in the College of Science and Engineering at James Cook University, Cairns, Australia. During 2004 and 2015, he was with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, Australia. He is an IET Fellow, a Fellow of Engineers Australia, and an Editor for IEEE Communications Letters. He was a co-recipient of three Best Paper Awards at 2015 WCSP, 2011 IEEE WCNC, and 2009 ICWMC. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow (2010-2011) by the Queensland Government of Australia, an Endeavour Research Fellow (2012-2013) by the Commonwealth Government of Australia, a Smart Futures Fellow (2012-2015) by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science (2014-2015). In 2008, he was a visiting scholar at Nanyang Technological University, Singapore. During Oct. 2010 and Mar. 2011, he was a visiting scholar at the University of Mississippi, Oxford, MS, USA. During Aug. 2012 and Mar. 2013, he was an Endeavour visiting associate professor at the University of Hong Kong. He has published over 160 papers in peer-reviewed journal and conference

papers. His research interests are in the broad area of communications and information theory, particularly coding and signal processing for multimedia communications systems.

M. Shamim Hossain (SM'09) is an Associate Professor at the King Saud University, Riyadh, KSA. Dr. Shamim Hossain received his Ph.D. in Electrical and Computer Engineering from the University of Ottawa, Canada. His research interests include serious games, cloud and multimedia for healthcare, resource provisioning for big data processing on media clouds and biologically inspired approach for multimedia and software system. He has authored and co- authored around 100 publications including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. He has served as a member of the organizing and technical committees of several international conferences and workshops. He has served as co-chair, general chair, workshop chair, publication chair, and TPC for over 12 IEEE and ACM conferences and workshops. Currently, He serves as a co-chair of the 6th IEEE ICME workshop on Multimedia Services and Tools for E-health MUST-EH 2016. He is on the editorial board of International Journal of Multimedia Tools and Applications. Previously, he served as a guest editor of IEEE Transactions on Information Technology in Biomedicine (currently IEEE JBHI), Springer Multimedia tools and Applications (MTAP), Springer Cluster Computing and International Journal of Distributed Sensor Networks. Currently, he serves as a lead guest editor of IEEE Transactions on Cloud Computing, IEEE Communication Magazine, Future Generation Computer Systems (Elsevier), SENSORS (MDPI), and Computers & Electrical Engineering (Elsevier). Dr. Shamim is a Senior Member of IEEE, a member of ACM and ACM SIGMM.

Sk Md Mizanur Rahman is an Assistant Professor in Information Systems Department in the College of Computer and Information Sciences at King Saud University, KSA. Prior to his current appointment, he worked for several years in cryptography and security engineering in the high-tech industry in Ottawa, Canada. He also worked as a postdoctoral researcher for several years in University of Ottawa, University of Ontario Institute of Technology (UOIT), and University of Guelph, Canada. He completed a Ph.D. in Risk Engineering (Major: Cyber Security Engineering) in the Laboratory of Cryptography and Information Security, Department of Risk Engineering, University of Tsukuba, Japan, on March 2007. Information Processing Society Japan (IPSJ) awarded Dr. Rahman with IPSJ Digital Courier Funai Young Researcher Encouragement Award for his excellent contribution in IT security research. He awarded with Gold Medal for the distinction marks in his undergraduate and graduate program. Primary research interest of Dr. Rahman is on Cryptography, Software Security, Information Security, Privacy Enhancing Technology and Network Security. He has published over 60 peer-reviewed journal and international conference research papers and book chapters.