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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

Learning Aided Multiple Time-Scale SON Function Coordination in Ultra-Dense Small Cell Networks

Meng Qin, Qinghai Yang, *Member, IEEE,* Nan Cheng, *Member, IEEE,* Jinglei Li, Weihua Wu, *Member, IEEE,* Ramesh R. Rao, *Fellow, IEEE,* Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—To satisfy the high requirements on operation efficiency in 5G network, self-organizing network (SON) is envisioned to reduce the network operating complexity and costs by providing SON functions which can optimize the network autonomously. However, different SON functions have different time scales and inconsistent objectives, which leads to conflicting operations and network performance degradation, rasing the needs for SON coordination solutions. In this paper, we devise a multiple time-scale coordination management scheme (MTCS) for densely deployed SONs, considering the specific time scales of different SON functions. Specifically, we propose a novel analytical model named M time-scale Markov decision process (MMDP), where SON decisions made in each time-scale consider the impacts of SON decisions in other $M-1$ time scales on the network. Furthermore, in order to manage the network more autonomously and efficiently, a Q-learning algorithm for SON functions (QSON) in the proposed MTCS scheme is proposed to achieve a stable control policy by learning from history experience. To improve the energy efficiency, we then evaluate the proposed MTCS scheme with two functions of mobility load balancing (MLB) and energy saving management (ESM) with designed network utility. Simulation results show that the proposed SON coordination scheme significantly improves the network utility with different quality of experience (QoS) requirements, while guaranteeing stable operations in wireless networks.**

***Key words*: Self-Coordination, Self-Organizing Network (SON), Multiple Time-scale, Markov Decision Process.**

I. Introduction

The tremendous growth of demands for diverse services is leading to complex wireless network infrastructures, making the operating complexity and operating expenses one of the significant challenges in 5G networks [2]. To overcome this problem, self-organizing networks (SON) are envisioned to be a promising technology for wireless network operation management in 5G networks, which is a key driver for improving Operations, Administration, and Maintenance (OAM) activities [3]-[5]. This has motivated network operators to employ SON functions incorporating intelligence to realize autonomous management for the upcoming 5G networks.

In wireless network optimization management, the operators need to continuously adjust parameter configurations based on measurements such as key performance indicators (KPIs) including network performance, capital expenditure (CAPEX), and operational expenditure (OPEX), whereas all those operations must be self-organized to realize self-optimizing for 5G networks. Hence, a lot of SON functions are easily available, which are deployed to meet the management goals of network operators, by replacing time consuming manual optimization tasks. Some typical SON functions are described follows:

- Coverage and Capacity Optimization (CCO): the objective of CCO is to provide optimal coverage and capacity for wireless networks by optimizing transmit power or antenna tilt, which should work on a rather long time-scale of days or weeks to cope with the changes in wireless environment and load imbalance in the long term.
- Mobility Load Balancing (MLB): MLB is enabled to detect changes of loaded conditions or the handoff environment, then determines an optimal policy to improve the network performance, which is executed in several minutes or hours.
- Energy Saving Management (ESM): ESM is defined to reduced energy costs and improve the network energy efficiency for network operators in the large time-scale (serval days or months).

Specifically, different SON functions have different execution time scales. For instance, MLB and MRO (Mobility Robust Optimization) take 5-10 minutes to optimize the network, while it takes a few days or months to realize the ESM function. Consequently, for achieving different network operator management goals, many SON functions with different time-scale may be activated concurrently.

Although SON functions have been deployed in 5G networks, the 5G networks management still face some new challenges. First, during the autonomous management period, different SON functions with contradictive goals in 5G networks may lead to both oscillating function execution and

* This work was accepted in part by the 2018 IEEE Global Communications Conference (2018 IEEE GLOBECOM) [1]. This research was supported in part by NSF China (61471287, 61701371,91638204), 111 Project (B08038), NRF of Korea (MSIP) (NRF-2017R1A2B2012337), China Scholarship Council and by the Natural Sciences and Engineering Research Council (NSERC), Canada.

M. Qin, Q. Yang, W. Wu and J. Li are with State Key Lab. of ISN, School of Telecom. Engineering, and also with Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, No. 2 Taibainan-lu, Xi'an, 710071, Shaanxi, China. (Email: mengqin@stu.xidian.edu.cn, {jlli68, qhyang}@xidian.edu.cn).

N. Cheng is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and School of Telecom. Engineering, Xidian University, No. 2 Taibainan-lu, Xian, 710071, Shaanxi, China. (Email: n5cheng@uwaterloo.ca). N. Cheng is the corresponding author.

R. R. Rao is the Director of CALIT2, University of California at San Diego, La Jolla, CA 920093 USA. (Email: rrao@ucsd.edu).

X. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. (E-mail: sshen@uwaterloo.ca).

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

poor network performance when executing simultaneously in specific cases, causing parametric or objective-based conflicts. Second, as the demands of the capacity in 5G networks keep increasing, more small cells with SON functions are deployed in 5G wireless network, and higher operation cost and computation power is required to manage the SON functions. In addition, the specific optimization algorithms in SON functions are needed to improve the network performance in 5G networks, which are currently usually taken as black boxes for network management. As a summary, considering self-operation for autonomous management in SON-aided 5G networks, the coordination of SON function is necessary to prevent or resolve SON conflicts and to improve network performance in SON-aided 5G networks. Furthermore, in order to reduce OPEX and improve the network efficiency, the design of an appropriate optimization algorithm for SON functions will be extremely challenging.

There have been extensive research efforts on the coordination of multiple SON functions based on reinforcement learning and control theory in [5]-[19]. From a macroscopic perspective, the preventive policy-based decision coordination mechanisms have already been developed in [6], [7]. An SON coordination prototype has been developed in [8], [9]. From the microcosmic point of view, a joint optimization scheme has been proposed for the conflict between MRO function and MLB function, in which the handover problems were prevented [10], [11].

Nevertheless, in the literatures, the SON coordination decisions are made only based on the instantaneous inputs of individual SON functions in the same time-scale, ignoring the spatial and temporal impact of the SON functions when making the optimization decisions. However, a large number of SON functions are parametrically or logically highly interdependent in the time domain, where the dependencies and conflicts among them may have a negative impact on the operation of the network behaviors, leading to poor network performance. Furthermore, due to the increasing dynamics of SON implementations in ultra-dense small cell networks, the future SON coordination is necessarily required to accommodate the SON conflicts with different time-scale by learning from previous experiences of SON functions to improve future decisions of network operators. To the best of our knowledge, few works have taken the SON execution time into account, which is an indispensable part of SON function coordination, especially in the SON-intensive wireless networks[1].

In this paper, we propose a novel multiple time-scale coordination scheme (**MTCS**) for SON functions to ensure steady wireless network operation. Specifically, the self-organizing mechanisms in SON functions are modeled as Markov control loops and the stability conditions of control loops with different time scales are derived. Furthermore, we devise an analytical model, multiple time-scale Markov decision process (MMDP) for hierarchically M SON coordination decision making processes, where SON coordination decisions in each level (each time-scale) are made in a specific time-scale, and the Q-learning algorithm is developed to solve the proposed model, considering the dynamics of wireless network environments. In addition, the QSON algorithm is designed to realize the SON functions for efficient SON management. Then, in order to improve energy efficiency of the wireless networks, we study a MLB-ESM use case with two typical SON functions (MLB and ESM) with designed network utility in ultra-dense small cell networks and develop a novel algorithm for the MLB-ESM SON functions. At last, the coordination capabilities of the proposed **MTCS** scheme is evaluated by the simulation results. The main contributions of this paper are summarized as follows:

- We devise a multiple time-scale Markov decision process model for the coordination of SON functions in wireless networks, considering the time domain characteristics of SON functions.
- We propose a novel **MTCS** scheme for the SON coordination to improve the network performance, while ensuring the stable network operation. In particular, the Q-learning based algorithm is developed to achieve the optimal SON coordination policy in the **MTCS** scheme by learning dependencies impacts on different time scales from history experiences of SON coordinations. In addition, the QSON algorithm is designed to realize SON function for efficient SON management.
- We identify the MLB-ESM use case with MLB function (fast time-scale) and ESM function (slow time-scale) in ultra-dense small cell networks to validate the performance of **MTCS** scheme. Furthermore, the network utility is designed to improve network energy efficiency with different quality of experience (QoS) requirements, and the theoretical analysis and simulation verification of the QSON algorithm is developed to implement SON functions in the proposed **MTCS** scheme.

The remainder of this paper is organized as follows. Section II gives the related works. Section III describes the system model and assumptions. The formulation of SON coordination problem is presented in Section IV. In Section V, we derive the **MTCS** scheme with MMDP model. The solutions for energy efficiency based on Q-learning approach are described in Section VI. Simulation results are given in Section VII. Section VIII concludes this paper. The list of key mathematical notations used in this paper are summarized in Table I.

TABLE I
SUMMARY OF KEY NOTATIONS

Notation	Description
$i, j \in \boldsymbol{\Re}$	state space of MDP in slow time scale
$\lambda \in \boldsymbol{\Lambda}$	action space of MDP in slow time scale
$\varpi, \kappa \in \boldsymbol{\aleph}$	state space of MDP in fast time scale
$a \in \boldsymbol{A}$	action space of MDP in fast time scale
G^L	MDP decision rule in fast time scale
G^U	MDP decision rule in slow time scale
Φ	MDP action function
$\hbar$	constant parameter
Δ	bounded function over $\Re \times \boldsymbol{\aleph}$
$\alpha \in [0, 1]$	constant parameter in MDP
π	MDP decision policy
T	scale factor between fast time scale and slow time scale

[1]Part of this work that considering the time characteristics has been previously published in IEEE GLOBECOM 2018 [1].

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

II. Related Work

With the increase of SON deployments in wireless networks, a comprehensive scheme of coordination among SONs is an increasing challenge for wireless network management [9-18]. Some literatures have studied for the problem of SON function coordination. In [9], a detailed classification of SON function conflicts was proposed, which provided a principle for designing suitable conflict solutions among SON functions. The conflicts between SON functions included the following different types: i) Configuration Conflicts: they mean that different SON functions operate on the same wireless configuration parameters for different SON functions such as the conflict between MLB and MRO functions. ii) Measurement Conflicts: SON functions influence the same performance data such as the same KPI values. iii) Characteristic Conflicts: SON functions have the different impacts on the attributes that cannot be directly measured and expressed through KPIs, which may be influenced through a fully disjoint set of configuration parameters, like ESM and MLB SON functions.

A generalized machine learning framework for the SON functions of MRO and MLB to optimize both handover performance and load balancing was proposed in [10], [14]. SON functions of cell ID assignment, coverage adjustment and idle mode control for multi-vendor networks were coordinated for balancing operation and further reducing requirements for manual conflict resolution in [15]. In [16], to enforce network stability and achieve network performance improvements, a distributed approach based on concave games for coordinating SON functions was proposed. In [17], a SON coordination scheme based on reinforcement learning approach was proposed, which allows the network to learn from previous experiences for improving future network decisions.

Specially, the high dynamics of ultra-dense small cell wireless environments in 5G networks, will lead to the network performance fluctuation and the increasing power consumption, then trigger different SON functions in different time scales. The dynamics of environments range from microseconds (e.g., fast fading channel), seconds (e.g., user mobility) to days and months (e.g., traffics over weekends), the SON functions are dynamically amplified during different time scales such as (M-RO and MLB in fast time-scale) and (ESM and CCO in slow time-scale) [18]-[19], as shown in Fig. 1. In such complicated environments with various SON functions, resolving possible SON conflicts and increasing energy efficiency are becoming much difficult and require manual updates by skilled experts, which leads to the increase of management complexity, energy consumption, CAPEX and OPEX.

To meet the requirements of efficient management in wireless networks, only a few authors pay attentions to the very important aspects of the dynamics of mobile data traffic and energy efficiency. From the energy saving perspective, a hierarchical management framework was designed in [20], where energy management and traffic steering were jointly optimized in different time scales. To reduce the information exchange cost and computational complexity, a dynamic pricing approach was proposed for power control problem in ultra-dense small cell networks in [21]. By jointly considering the satisfaction of users and the revenue of network operators, an auction game-based method was proposed in [22]. From the load balancing perspective, the problem of joint load balancing and interference mitigation was studied with the stochastic optimization, considering the dynamic scheduling in 5G networks in [23]. The author in [24] proposed a framework for context-aware self-optimization in small cell environments and load balancing was used to evaluate the satisfaction of users. A load-balanced algorithm for small-cell environments was proposed, considering traffic distribution in [25]. To fully incorporate the green energy utilization in traffic load balancing strategies, a traffic load balancing framework which can obtain the tradeoff between network utilities of the average traffic delivery latency and the green energy utilization was proposed in [26]. To minimize the total discounted energy consumption, Q-learning approach was developed, taking full use of the global network state information obtained from the network controller in [27]. A balanced dynamic planning approach based on Markov decision process was proposed in [28], while saving energy and satisfying the users' QoS requirements. However, most of the research efforts in wireless networks have been focused on dealing with the inefficiencies and impairments arising from very short-term dynamics. It is notable that much more system level efficiency can be harnessed by developing SON coordination solutions for longer time scales. From a 5G networks perspective, there is a increasing challenge to design an efficient SON coordination management scheme for reducing manual operation cost, complexity of the envisioned network architecture and improving network performance for small cell networks. In particular, we make a detailed comparison of these literatures, the differences between our proposed work and other related works are summarized in Table II.

III. System Model

In a typical SON function, there are three kinds of SON functions, self-configuration, self-optimization and self-healing, which replace the classic manual configuration, post deployment optimization, and maintenance in traditional wireless networks, respectively. We consider M $(M \geq 2)$ SON functions operating simultaneously in small cell networks as shown in Fig. 1. Then, we introduce the SON control model, SON operation model and SON performance metric model.

A. Control Model

A SON function is effectively realized by means of specific SON mechanisms (SON_m), which relates to particular SON use cases defined by the 3rd Generation Partnership Project (3GPP). Focusing on self-optimization, the SON_m aims at maintaining relevant KPIs above or below a specific value by actuating over an appropriate set of input parameters to achieve predefined objectives of operators. Typically, the SON mechanism is represented by several control loops where controllable input parameters are dynamically adjusted according to output metrics and corresponding objective requirements. From an implementation point of view, a SON controller is considered to be implemented responding to strategies of

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

TABLE II
THE SUMMARY OF SON FUNCTION COORDINATION WORKS

SON Functions	Timescale	SON Model	Network Objective	Related Work
ESM and MLB	multiple	multiple MDP model	Network energy efficiency	Our Work
CCO and MLB	same	linear ordinary differential equation	outage probability	[6]
MLB and MRO	same	Graph-Theoretic Representation	Handover rate	[8]
MLB and MRO	same	extreme values model	call dropping	[11][17]
MRO and ICIC	same	single MDP model	network utility	[9][14][18]

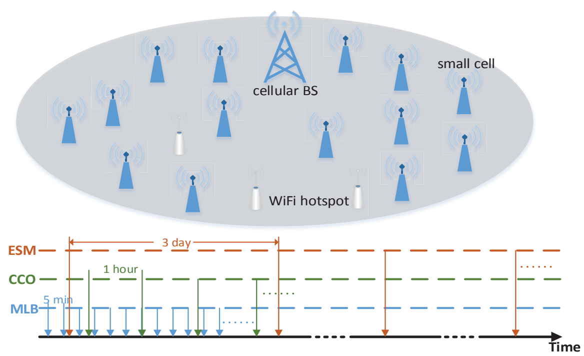


Fig. 1. SON functions with multiple timescales in ultra-dense network

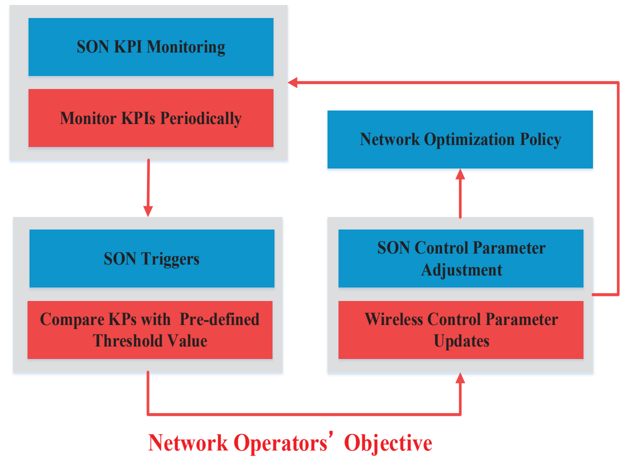


Fig. 2. SON Function Operation Flow for Autonomous Management

network operators, deciding which SON_m actions have higher priority with respect to others at a given time period. In the context of heterogeneous scenarios, we consider a universal framework where stand-alone SON_m reacts to the fluctuation of KPIs by deciding to either increase or decrease corresponding influential parameters, aimed at maintaining particular metrics or KPIs. The SON controller is needed to provide the means for SON_m coordination and potential conflict management arising from possible conflicting decisions by different SON_m.

B. SON Operation Model

SON function is basically a control loop that autonomously tunes RAN parameters to adapt the RAN to variations of the wireless environment and of the traffic according to objectives of network operators. It makes it possible to autonomously configure newly deployed base stations or nodes for self-configuration, adjusting parameters to improve network performance, which can be seen as a self-optimization process, and root faults and compensation of faulty network nodes to realize self-healing functionality as shown in Fig. 2. Specially, different SON functions operate in different time scales. Some SON solutions have the time-scale of hours to days such as CCO and ESM. Meanwhile, some other SON functions operate at the time-scale of seconds to minutes such as MRO and MLB [11]. Hence, there are mainly two SON operation models for SON functions ensuring stable network operation: i) SON functions work in different time scales; ii) SON functions work in the same time-scale. The multiple time-scale model of SON functions is studied in this paper.

C. SON Performance Metric Model

We consider the ultra-dense small cell network scenario consisting of U users and C small cells. The signal to interference plus noise power ratio of the user u from small cell c is given by

$$SINR_{c,u} = \frac{P_{c,u}}{\sum\limits_{q \neq c} P_{q,u} + P_{M,u} + N_0}, \qquad (1)$$

where $P_{c,u}$ represents the received power of the user u from the small cell c, $P_{q,u}$ is the interference power from small cell q, N_0 is the background noise. $P_{M,u}$ is the interference power from Marcocell.

The total data rate of the ultra-dense small cell network is defined as

$$R_{sum} = \sum_{c=0}^{C} \sum_{u=1}^{U} B_{c,u} \log_2(1 + SINR_{c,u}), \qquad (2)$$

where $B_{c,u}$ represents the bandwidth allocated to user u from small cell c, and $c \in \{0, 1, 2, \cdots, C\}$, the $c = 0$ represents the Macrocell for unification of expression for the wireless network sum rate.

For ESM function, small cells can be switched to the sleep mode for saving power or switched on for improving the capacity and QoS requirements of users. Therefore, the small cell c power consumption can be modeled as

$$P_c = \begin{cases} P_c^O + P_c^T, & \gamma_c = 1, \\ P_c^O, & \gamma_c = 0, \end{cases} \qquad (3)$$

where the P_c^O is basic consumption of circuit power for small cells and P_c^T is the transmission power of small cells,

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

respectively. $\gamma_c \in \{0, 1\}$ is the indicator, $\gamma_c = 1$ when the small cell c is switched on, and $\gamma_c = 0$ when the small cell c is switched to sleep mode. Consequently, the sum energy consumption of the wireless network is formulated as

$$P_{sum} = P_{c=0}^{T} + P_{c=0}^{O} + \sum_{c=1}^{C} (\gamma_c P_c^T + P_c^O) \tag{4}$$

where $P_{c=0}^{T}$ is the transmission power of Marcocell and $P_{c=0}^{O}$ is the circuit power of Marcocell, respectively.

Furthermore, we assume that each SON mechanism SON_m sends an update request targeting to modify the K tuned parameters, which runs onto optimize the performance of wireless networks. Let $\mathcal{P}$ be the set of possible values of wireless network parameters, such as the cell individual offsets and transmission power. The update requests consist of K tuple $\Psi = \{0, -1, +1\}^K$ as shown

$$\Psi_k = \begin{cases} +1 & \uparrow, \\ 0 & \rightarrow, \\ -1 & \downarrow, \end{cases} \tag{5}$$

where $\Psi_k = +1$, $\Psi_k = 0$ and $\Psi_k = -1$ means to increase $\uparrow$, maintain $\rightarrow$ and decrease $\downarrow$ the value of parameter k, respectively. For instance, the MLB SON function instances tune the Cell Individual Offset (CIO) parameter in order to optimize the cell load or the transmission power, and the CIO values can choose from the set of $\{-9, -6, -3, 0\}$ dB when the MLB SON function receives requests such as off-load, on-load or keep the load unchanged according different network goals. For the CCO SON function instances, it tunes the electronic antenna tilt with the set $\{0, 1, 2, 3, \cdots, 90\}$, which can optimize the network capacity and coverage.

In particular, for the SON function operation in this paper, we assume that the instantaneous signal strength from each small cell can be measured by each user with pilot detection. The measurement results are sent to its serving small cell within uplink data transmission or by periodical report.

IV. Multiple Time-Scales SON Coordination

The goals of specific SON_m are to be translated into specific KPI metric requirements in terms of satisfactory metric values, which represent a high-level specific requirements by the network operator. Once different metrics have been identified, we need to find those parameters, which have the influence on the specific metrics and use corresponding SON functions.

For SON functions, every decision corresponds to an action taken towards a represented state of the process, helping the wireless network to evaluate its condition. For such a dynamic environment of wireless networks, the network dynamics are modeled using a mathematical framework named Markov decision process (MDP) to optimize desired objectives of the network operators in the process of SON functions. We define the underlying MDP for each SON mechanism SON_m, which consists of the SON_m state space $\mathcal{S}$, SON_m action space $\mathcal{A}$, and the SON_m reward function R as

- SON_m state space: $S = \mathcal{P} \times \Psi$. The state $s \in \mathcal{S}$ of MDP model contains the parameter configurations and update requests.
- SON_m action space: $\Psi_S = \{0, -1, +1\}^K$. The action of MDP model operates to increase, decrease or maintain the value of wireless parameter k when there exists a network optimization request during the SON optimization period.

$$a_k = \begin{cases} \uparrow & \text{if} \quad \Psi_{\mathrm{s,k}} = +1, \\ \rightarrow & \text{if} \quad \Psi_{\mathrm{s,k}} = 0, \\ \downarrow & \text{if} \quad \Psi_{\mathrm{s,k}} = -1, \end{cases} \tag{6}$$

- SON_m reward: $R = r(s, a)$. A reward set containing a r value for each action. The reward is defined as the network utility, which will be defined later.

In practice, specific metrics for a SON user case can be detected by a SON controller, and it can figure out key parameters for the SON user case. The SON controller is taking charge of enforcing (or not) the SON_m decisions to keep the metrics below some specific thresholds. The network operator receives a SON optimal policy which is the one that maximizes the measure of long-run expected rewards. Then, it optimizes the network according to this recommendation, considering different network requirements.

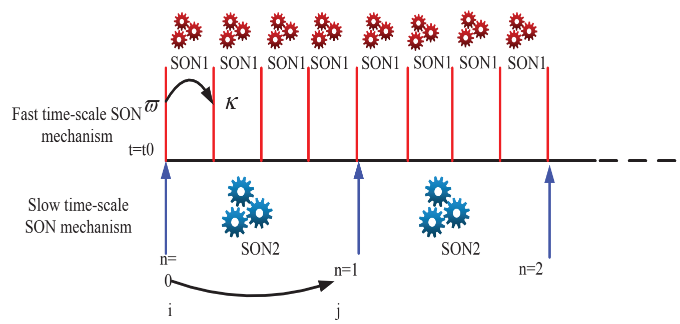


Fig. 3. Multiple time-scale SON coordination management framework

First, we take two time-scale MDP model ($M = 2$) as an example to elaborate the SON coordination scheme. It can be extended to multiple time-scale MDP model for SON coordination with low complexity, which will be discussed later. We consider that the MDP of the upper SON mechanism with slow time-scale has a state space $\mathfrak{R}$ and an action space $\boldsymbol{\Lambda}$. At each decision time $n \in \{0, 1, \cdots, \}$ and at a state $i_n \in \mathfrak{R}$, an action $\lambda_n \in \boldsymbol{\Lambda}$ is taken, and the state i_n transits to state $i_{n+1} \in \mathfrak{R}$, according to the probability $P_{SON}^{U}(i_{n+1} | i_n, \lambda_n)$. During the period of the upper SON mechanism, once the state of upper SON mechanism was chosen and the action of the upper SON mechanism was taken, the states and actions for the lower SON mechanism are chosen based on the upper SON mechanism as shown in Fig. 3. In addition, we consider that MDPs in the lower SON mechanism have the same finite SON actions denoted by $\boldsymbol{A}$ and the same SON states denoted by $\aleph$. There is no intersection between the lower SON mechanism and the upper SON mechanism both in states and actions, i.e., $\mathfrak{R} \cap \aleph = \emptyset$, $\boldsymbol{A} \cap \boldsymbol{\Lambda} = \emptyset$. For the SON mechanism with fast time-scale, we denote the time by $t \in \{t_0, t_1, \cdots\}$. We also define T as the scale factor between the upper SON mechanism and the lower SON mechanism, and thus we obtain $t_{nT} = n, n \in \{0, 1, \cdots\}$,

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

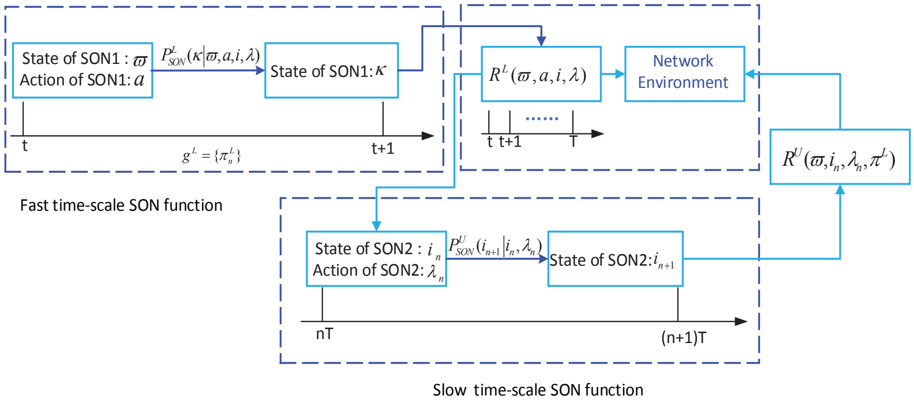


Fig. 4. The Flowchart of Multiple Time-scale SON coordination Framework

which is the representation of multiple time-scale for SON functions.

Second, we initialize the state of the lower SON mechanism level MDP $\varpi \in \aleph$ and we define $i \in \Re$ as the initial state for the upper SON mechanism with slow time-scale, and thus the state can be $\varpi_{t_0} = \varpi$ and $i_0 = i, n = 0$. Then, the MDP process of the lower SON mechanism is evolved to achieve the optimal network policy.

- For the lower SON mechanism state ϖ at t_0, an action $a \in \boldsymbol{A}$ is taken, when ϖ makes transition to the next state $\kappa \in \aleph$ at time t_1 with the probability $P_{SON}^{L}(\kappa\,|\varpi, a, i, \lambda)$ and the reward $R^L(\varpi, a, i, \lambda)$ in the lower SON mechanism, and continues until it arrives at the time t_{T-1}.
- In the lower level MDP, the states and decisions of the upper SON mechanism bring out the function of state transition and the reward function. The action λ_1 of the upper SON mechanism at time $n = 1$ will be taken and a new MDP determination process will be triggered.

At last, we define a lower SON mechanism level decision rule $g^L = \{\pi_n^L\}$ and $\pi_n^L = \{\Phi_{t_{nT}}, \cdots, \Phi_{t_{(n+1)T-1}}\}$. A nonnegative and bounded reward function R^U is defined for the upper SON level. Given a lower SON mechanism level decision rule $g^L \in G^L$, we define the reward function R^U, which is the total expected reward obtained by the fast time-scale T horizon policy π^L including an instantaneous reward at the slow time-scale

$$\begin{aligned} &R^U(\varpi, i_n, \lambda_n, \pi^L) = R_I^U(i_n, \lambda_n) + \\ &E_{i_n,\lambda_n}^{\varpi}\left\{\sum_{t=t_{nT}}^{t_{(n+1)T-1}} \alpha^t R^L(\varpi_t, i_n, \lambda_n, \Phi_t((\varpi, i_n, \lambda_n))\right\}, \end{aligned} \tag{7}$$

where α^t with $0 \geq \alpha \leq 1$, is taken as a constant parameter for all n, which represents the contributions of the fast time-scale SON function reward for the reward of the slow time-scale SON function at time t. $\Phi_t(\varpi, i_n, \lambda_n)$ is the upper level action function: $\aleph \times \Re \times \boldsymbol{A} \to \boldsymbol{\Lambda}$. The reward R^U for slow time-scale SON function is the total expected reward with policy $\pi^L(i_n, \lambda_n)$ at T horizon, added by an instant reward of the upper SON mechanism. The instant reward function of the upper SON at state i_n is defined as $R_I^U(i_n, \lambda_n)$. The multiple time-scale SON coordination policy is obtained based on the MDP model. Based on the analysis, the multiple time-scale SON coordination framework is given as shown in Fig. 4.

Remark 1: For SON functions, keeping their own network information during the execution time of a SON function is insufficient because of the potential conflicts after the end of the SON execution. To satisfy these additional requirements of wireless network, history network information of SON functions should be taken into account for the conflicting SON functions. Hence, temporal-spatial impacts are constructed specifically for pairs of conflicting SON functions to satisfy the demands of SON coordination with respect to potential conflicts between SON functions with the proposed MMDP model.

V. Solution of Multiple SON Coordinations

The core problem of MMDP modeled SON coordination is to find an optimal policy that specifies the actions of SON functions, which will be chosen when in different network states. Q-learning is a straightforward and model-free reinforcement learning algorithm adopted to define the values of transition probabilities and find an optimal policy which converges to an optimal policy for the given MDP with a finite action-state [10]. Hence, the Q-learning approach is developed to solve this problem.

The single-step reward for the upper level SON is achieved with the given reward at the upper SON mechanism, and we have the total expected reward achieved by the low SON mechanism level T horizon nonstationary policy π^L. Our goal is to obtain a SON mechanism coordination decision rule pair of $g^L \in G^L$ and $g^U \in G^U$ that achieves the following functional values defined as

$$
\begin{aligned}
&V^*_{SON}(\varpi, i) = \\
&\max_{g^U\in G^U} \max_{g^L\in G^L} E^{\varpi,i}\left\{\sum_{n=0}^{\infty}\gamma^n R^U(\varpi_{t_{nT}}, i_n, g^U(\varpi_{t_{nT}}, i_n), \pi^L)\right\}\\
&= \max_{g^U\in G^U} \max_{g^L\in G^L} E^{\varpi,i}\left\{\sum_{n=0}^{\infty}\gamma^n \left(E^{\varpi_{t_{nT}}}_{i_n,\lambda_n}\right.\right.\\
&\left[\sum_{t=t_{nT}}^{t_{(n+1)T-1}}\alpha^{(t)}R^L(\varpi_t, i_n, g^U(\varpi_{t_{nT}},\varpi_n), \Phi_t((\varpi, i_n, g^U(\varpi_{t_{nT}},\varpi_n)))\right.\\
&+\left.\left. R^U_I(i_n,\lambda_n)\right]\right\},
\end{aligned}
\tag{8}
$$

where the V^*_{SON} is defined as the optimal discounted value function over $\aleph\times\Re$ in two time-scale. Based on the slow time-scale decision rule g^u, the slow time-scale SON will make the decisions depending on the states of the fast time-scale SON functions.

Furthermore, given the fixed pair of the slow time-scale SON mechanism state i and action λ, we define a set $\Pi^L_{SON}[i,\lambda]$ of all possible T horizon policies as

$$
\Pi^L_{SON}[i,\lambda] := \left\{\pi^L[i,\lambda]\,\middle|\,\pi^L[i,\lambda] := \{\Phi^{i,\lambda}_{t_0},\cdots,\Phi^{i,\lambda}_{t_{T-1}}\}\right\}. \tag{9}
$$

For achieving the optimal policy, the Bellman's optimal equation and the optimal decision rule by obtaining the optimal value of each state are given as shown in Theorem 1.

Theorem 1: For all $\varpi\in\aleph$ and $i\in\Re$, we obtain

$$
\begin{aligned}
V^*_{SON}(\varpi,i) = \max_{\lambda\in\Lambda}\Bigg(&\max_{\pi^L[i,\lambda]\in\Pi^L_{SON}[i,\lambda]}\Big\{R^U(\varpi,i,\lambda,\pi^L[i,\lambda])\\
&+\gamma\sum_{\kappa\in\aleph}\sum_{j\in\Re}P^T_{\varpi\kappa}(\pi^L[i,\lambda])P^U_{SON}(j\,|i,\lambda)V^*_{SON}(\kappa,i)\Big\}\Bigg),
\end{aligned}
\tag{10}
$$

where V^*_{SON} is the unique solution of (10) in the proposed scheme. Then, we achieve the equation as λ^* and $\Pi^*[i,\lambda]=\{\Phi^*\}$. We set the decision rule $g^U(\varpi,i)=\lambda^*$ and set the decision rule π^L for g^L with $\Phi(\varpi,i,\lambda^*)=\Phi^*(\varpi,i,\lambda^*)$. The optimal V^*_{SON} is achieved based on the decision rule g^U in slow time-scale and the decision rule g^L in fast time-scale.

Proof: Using the principle of optimality, V^* is the optimal solution. Specifically, the principle of optimality shows that an optimal policy can be constructed in a piecemeal fashion. Thus, we first construct an optimal policy for the "tail subproblem" involving the last stage V^*_T, and then extend the optimal policy to the "tail subproblem" involving the last two stages, V^*_T, V^*_{T-1}. Continuing this manner until an optimal policy for the entire problem is constructed. Hence, the Theorem 1 is proved. ■

Particularly, we define an initialization function Γ_{SON} such that we initialize $\varpi_{t_{nT}}, n=1,2,\cdots$ by Γ_{SON}, which gives a probability distribution. Based on the MDP theory, we achieve the optimal SON coordination decision rule for each SON mechanism level in MMDP. To obtain the decision, we first define an operator Ξ such that for a function V defined over $\aleph\times\Re$ as

$$
\begin{aligned}
\Xi(V_{SON})(\varpi,i) = \max_{\lambda\in\Lambda}\Bigg(&\max_{\pi^L[i,\lambda]\in\Pi^L[i,\lambda]}\Big\{R^U(\varpi,i,\lambda,\pi^L[i,\lambda])\\
&+\gamma\sum_{\kappa\in\aleph}\sum_{j\in\Re}\Gamma^{\pi^L}_{SON}(\varpi,i,\lambda)[\kappa]P^U_{SON}(j\,|i,\lambda)V_{SON}(\kappa,i)\Big\}\Bigg),
\end{aligned}
\tag{11}
$$

where $\Xi(V_{SON})$ is a $\gamma - contraction - mapping$ in sup-norm as shown in [30], and $\Gamma^{\pi^L}_{SON}(\varpi,i,\lambda)[\kappa]$ is the probability defined on $\kappa\in\aleph$.

Then, we obtain the function V for all $\varpi\in\aleph$ as

$$
\lim_{n\to\infty}\left\{\Xi^n(V_{SON})(\varpi,i)-\Xi^{n-1}(V_{SON})(\varpi,i)\right\}\to\hbar, \tag{12}
$$

where $\hbar$ is constant. With any fixed state pair $\kappa\in\aleph$ and $j\in\Re$, we then achieve

$$
\lim_{n\to\infty}\{\Xi^n(V_{SON})(\varpi,i)-\Xi^n(V_{SON})(\kappa,j)\}\mapsto\Delta(\varpi,i), \tag{13}
$$

where $\Delta(\varpi,i)$ is defined as a bounded function over $\Re\times\aleph$. In particular, the MDP solution of the upper SON mechanism should relay on the MDP solution of the lower SON mechanism. We achieve the optimal value of the **MTCS** algorithm by applying the iteration method for solving MMDP. In particular, the flowchart of multiple time SON coordination framework is given in Fig. 4.

In traditional management of network operators, SON functions are typically built in a stand-alone manner, i.e. without considering the existence of other SON functions in the network. From the practical perspective, the network will run one or several instances of each SON function. Each SON instance optimizes a set of cells. Having several SON instances in the network will create conflicts and lead to poor network KPIs. In this paper, we propose an universal SON management framework, which can be extended to other SON functions. The Q-learning approach is proposed to learn from the past experience, which is beneficial for improving our future decisions. Without this information, the SON coordination would suffer from considerable shortcomings in anticipating the impact of the decisions. The SON functions are seen as black-boxes by the SON coordination scheme, knowing nothing about the algorithm running inside them. It only knows current update requests and current network parameter configuration. The **MTCS** scheme provides an equitable conflict resolution. A step by step **MTCS** scheme description is given in **Algorithm 1**.

In the line 1, we initialize the state of the lower SON mechanism level MDP $\varpi\in\aleph$ and we define $i\in\Re$ as the initial state for the upper SON mechanism with slow time-scale, and thus the state can be $\varpi_{t_0}=\varpi$ and $i_0=i, n=0$. From line 2 to line 13, the proposed scheme is solved numerically by iteration computation method with the maximum iterations $I_{\max}$. From line 3 to line 5, it is the SON1 iteration loop. With the wireless configuration parameters $\mathcal{P}$ and update requests Ψ, the optimal policy of SON1 π^{L^*} is obtained. In line 6, the defined reward function R^U of the proposed model is calculated based on the current state and the optimal policy of fast time scale SON1 function. In line 7 and line 8, if the condition that $\lim_{n\to\infty}\left\{\Xi^n(V_{SON})(\varpi,i)-\Xi^{n-1}(V_{SON})(\varpi,i)\right\}\to\hbar$ can be

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

Algorithm 1 MTCS Algorithm

1: **Initialization** Set maximum iteration number $I_{\max}$, $R^L = 0$, $R^U = 0$, T and $\hbar$; Observe current network wireless configuration parameters $\mathcal{P}$ and update requests Ψ; begin iteration (SON1 Outer Loop);
2: **for** $1 \le n \le I_{\max}$ **do**
3:   **for** $1 \le t \le t_{nT}$ **do**
4:     Solve SON2 Function with $\mathcal{P}$ and Ψ;
5:     obtain SON1 policy π^{L^*} (SON2 Inner Loop);
6:   **end for**
7:   Calculate R^U with the current action in the network;
8:   **if** $\lim\limits_{n\to\infty} \left\{\Xi^n(V_{SON})(\varpi, i) - \Xi^{n-1}(V_{SON})(\varpi, i)\right\} \to \hbar$ **then**
9:     Take current action as the optimal policy π^{U^*} in the wireless network.
10:       **break**
11:   **else**
12:     Set $t = t + 1$;
13:   **end if**
14: **end for**

satisfied, we can obtain the optimal policy π^{U^*} by taking current actions. From line 9 to line 13, if the condition that $\lim\limits_{n\to\infty} \left\{\Xi^n(V_{SON})(\varpi, i) - \Xi^{n-1}(V_{SON})(\varpi, i)\right\} \to \hbar$ cannot be satisfied, we set $t = t+1$, and then the algorithm executes the next iteration loop. Based on these steps, the optimal SON function coordination policy can be obtained.

Remark 2: In particular, the M MDP models are ordered sequentially from fast time-scale to slow time-scale. For the $M > 2$ time-scale model of SON coordination, we can see that the states and the actions of the Mth MDP model in the slow time-scale depend on the the states and the fast time-scale MDP models $(1, 2, \cdots, M-1)$, and the reward function of the Mth MDP model is calculated with the states and the actions of $M-1$ MDP models. Hence, the multiple time-scale optimal value function can be obtained based these reward functions $(1, 2, \cdots, M-1)$, which is similar with the proposed **MTCS** scheme. Particularly, in the value based iteration solutions, the maximized complexity is $\mathcal{O}(|S|^2 |A|)$, because the value based solution takes the optimization bellman equation to update the value function, so the value function is optimal when the algorithm converges with the corresponding optimal policy. More importantly, the time scales M of SON functions are generally limited to a few (seconds, minutes, hours, days, etc.) in real systems. Therefore, the complexity of the proposed scheme is acceptable with good scalability, which is very practical in real wireless systems.

VI. Q-Learning based energy efficiency management with SON

In order to guarantee stable and desired network operation, SON functions have attracted significant attentions from industry. Furthermore, energy efficiency has also been a key performance metric for the requirements of SON function in 5G networks. At the same time, smart load balancing is required to automatically achieve optimal network performance and fulfill QoS requirements of users in wireless networks. Hence, to improve the energy efficiency, the network utility is designed properly and two different time-scale SON functions, i.e., MLB and ESM SON functions in 3GPP, are considered for energy efficiency management with the proposed SON coordination management scheme.

MLB function is purposed to cope with the unequal traffic load and to achieve the load balancing in 3GPP [31]-[32]. The thresholds of MLB function are defined by related load performance KPIs, which trigger an offloading operation in wireless networks. The self-organized process of MLB function can adapt network capacity or load to the current traffic demand and QoS requirements of users by shifting users or traffics among the small cell networks. In addition, to facilitate energy saving in wireless networks, ESM function is currently discussed in 3GPP standardization [31]. The self-organized cell activation or sleeping off of ESM function is a straightforward process as a small cell can determine the behavior of cell activation or sleeping off. If a particular cell has been switched off to reduce energy consumption, it notifies nearby cells via a deactivation indication over X2. In addition, a cell can request a neighbor cell to re-activate the cells with the network demands.

In particular, for ESM functions, the dynamic architectures that vary with the dynamic traffic, passive sleep mode and offloading operation have been considered. It will lead to unbalanced load and handover problems, which caused the potential conflicts between ESM and MLB functions, declining wireless network performance. The two SON instances are seen as black-boxes by the operators, since the network does not know the algorithm running inside them. It only knows the immediate update requests and the immediate network parameter configurations. For the network deployed with both MLB and ESM functions, the task of SON coordination scheme is to find an efficient conflict solution. If these two SON functions are independent of each other, both slow time-scale and fast time-scale mechanisms run to achieve their optimal solution independently. But it leads to the oscillation problem and increases the network management cost when the base station trigger these SON functions as shown in Table III (× means that there is a conflict and ✓ means that there is no conflict) [33].

TABLE III
The Potential Conflicts of Different SON Functions

SON Functions	Input Parameters Conflicts	Output Parameters Conflicts	Measurement Conflicts
ESM and CCO	✓	×	×
ESM and MLB	✓	✓	×
MLB and MRO	✓	×	×
ESM and MRO	✓	✓	×
MLB and CCO	✓	✓	×

To achieve energy efficiency, first, larger amounts of management information that have been reported periodically are collected to calculate the energy efficiency of ultra-dense small cell network. Then, when the energy efficiency is less than a preset threshold which is defined based on different objectives of network operators, our proposed **MTCS** scheme

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

is triggered to make the optimal decision policy for wireless networks. Consequently, new configuration parameters are determined for network operators reconfiguration to improve energy efficiency. In particular, the SON mechanisms of MLB and ESM in **MTCS** scheme are proposed. The network utility is designed as the network performance to achieve optimal energy efficiency, considering the SON conflicts between MLB and ESM functions.

Q-learning is based on the agents that works by learning an action-value function, denoted by $\mathbf{Q}(s,a)$, reflecting potential gains in an action-state pair [32]. Considering the dynamic environment of small cell networks, MLB and ESM functions are taken as agents, using the feedback of wireless networks and learning from the effects of their actions. Each SON agent, by following the Q-learning algorithm, has to interact in real time with the environment in order to achieve the goal of learning an optimal energy efficiency management policy. The SON agent aims to maximize its total reward by learning which action is optimal for each network state with the highest long-term reward. The reward is calculated as a weighted sum of the expected values of the rewards of all future steps starting from the current state. Therefore, we define the $\mathbf{Q}^{son}$ value for SON as $\mathbf{Q}^{son}:\mathbf{S}_{son}\times\mathbf{A}_{son}\rightarrow\mathbb{R}_{son}$.

The SON agent chooses an action $a_{son,t}$ and achieves the reward $r_{son,t}$ at time t, as mentioned in section IV. In Q-learning process, the SON agent obtains an optimal state-action policy by learning from the wireless environment. The Q-learning algorithm is solved by the value iteration method [33] as

$$\begin{aligned}\mathbf{Q}^{son}_{t+1}(s_{son,t},a_{son,t}) &= (1-\eta)\mathbf{Q}^{son}_{t}(s_{son,t},a_{son,t}) \\ &+\eta(\mathbf{r}_{son,t}+\kappa^{son}\max\mathbf{Q}(s_{son,t+1},a_{son})),\end{aligned} \tag{14}$$

where η is the learning rate ($0<\eta\leq 1$) and κ^{son} is the discount factor, respectively. $\mathbf{r}_{son,t}$ is the SON reward at time t. It ultimately gives the expected utility of taking a given action a_{son} in a given state s_{son} based on the Q-learning algorithm for SON function.

Furthermore, we define the time averaged power consumption and transmit rate of the wireless network [35]-[38], respectively, as

$$\overline{P_{sum}}=\lim_{T\rightarrow\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\{P_{sum}(t)\}. \tag{15}$$

and

$$\overline{R_{sum}}=\lim_{T\rightarrow\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\{R_{sum}(t)\}. \tag{16}$$

To improve the long-term energy efficiency of ultra-dense small cell networks, we consider the time characteristics of ESM and MLB function. The network utility is defined as the value that the throughput minus the power consumption in this paper, considering the MLB and ESM SON functions, which is the format translation as the ratio of long-term total throughput to the long-term corresponding energy consumption as

$$\boldsymbol{U}_{sum}=W_1\overline{R_{sum}}-W_2\beta\overline{P_{sum}}. \tag{17}$$

where $\boldsymbol{U}_{sum}$ is the network utility, taking as the reward function in the Q-learning process. $\overline{R}_{sum}$ and $\overline{P}_{sum}$ are the long-term total throughput and long-term corresponding energy consumption, respectively. W_1 is the weight of network throughput, W_2 is the weight of power consumption and β is the coefficients for the uniform dimension, which is taken as a constant in this paper. In detail, the states of MLB SON functions are defined as $S_{MLB,t}=\{{P_t}^{cf},D_{s,t},D_{n,t}\}$, ${P_t}^{cf}$ is the wireless configuration parameter set (including the transmission power of the small cells) in time t, $D_{s,t}$ and $D_{n,t}$ are the serving and neighboring cells user distribution at time t, respectively. The action space of MLB SON functions are defined as $A_{MLB}=\{-9,-6,-3,0\}$ dB. For the ESM SON function instance, it tune the small cell operation model and transmission power (we set the wireless operation parameter equal to 1 if the small cell is in active state, otherwise it equals to 0 when it is in sleeping state) to save the energy consumption with the network goal of energy saving in wireless networks. Hence, the states of ESM functions are defined as $S_{ESM,t}=\{{P_t}^{cf},P_E\}$, ${P_t}^{cf}$ is the wireless configuration parameter set (including the transmission power of the small cells) in time t, $P_{E,t}$ are the energy consumption distribution at time t. The action space of ESM SON functions are defined as $A_{ESM}=\{0,1\}$, which is the operation mode of small cells.

For finding the optimal policy of QSON algorithm, ϵ-greedy method is used herein to determine when and how much to explore the action space before exploiting learned knowledge [34]. In addition, the proposed QSON learning algorithm is given step by step in **Algorithm 2** as follows.

Algorithm 2 QSON Energy Efficiency Algorithm

Initialization At beginning of SONs that been triggered, set the $\mathbf{Q}^{son}(s,a)$ as zero, action set $\mathbf{A}_{son}$, state set $\mathbf{S}_{son}$, learning rate η, discount factor κ, probability ϵ and time horizon T_{son};

for $1\leq t\leq T_{son}$ **do**

1. Observe the current state $s_{son,t}$, select and execute an action $a_{son,t}=\arg\max_{a\in A}\{\mathbf{Q}^{son}(s,a)\}$ with probability $1-\epsilon$, otherwise, the QSON selects a random action a_{son} from the action set $\mathbf{A}_{son}$ with probability ϵ;
2. Derive the reward value $r_{son,t}$;
3. For new network state $s_{son,t+1}$, update QSON function value $\mathbf{Q}^{son}_{t+1}(s,a)$ with (14);
4. update the network state $s_{son}=s_{son,t+1}$ and time $t=t+1$;
5. Until it satisfies the conditions and convergence;

end for

Remark 3: To improve the network energy efficiency, the MLB function and ESM function are deployed and running with proposed QSON algorithm respectively if there is no conflict occurs or no network performance decreases. When the energy efficiency value is smaller than the default threshold which is defined based on different objectives of network operators, our proposed **MTCS** scheme is triggered to make the optimal decision policy for wireless networks.

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

TABLE IV
KEY PERFORMANCE INDICATORS

Parameters	**Values**
System bandwidth	10MHz
Number of Small Cell	10
Number of Users	40
Small Cell Transmission power	1W
Marocell Transmission power	40W
Small Cell Circuit power	5W
Marocell Circuit power	50W
Shadowing standard deviation	4dB
Pathloss	$128.1 + 37.6log10(max(d_{Km}))$
Noise Power	-174 dBm/Hz
Learning rate	0.8
Discount factor	0.6
Exploration rate	0.2
QoS requirements	30kbps, 40kbps, 50kbps.

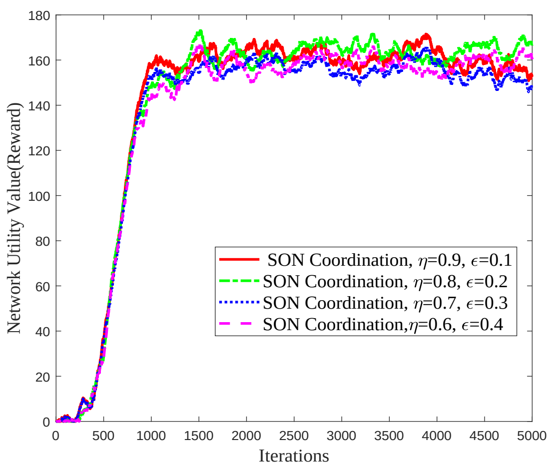


Fig. 5. Network Utility vs. Time iterations

VII. SIMULATION RESULTS

We consider a heterogeneous scenario consisting of one Marcocell, C small cell base stations, and U user equipments (UE), which are randomly distributed in the network area. MATLAB 2016a tool is used as the simulation platform. The standard MLB and ESM instances in 3GPP TS 36.902 are considered in this paper. We make the assumption that there are conflicts between the MLB function and ESM function which are caused by the input wireless parameters or the out wireless parameters as shown in the Table III. Hence, we introduce the case where alone SON scheme is implemented as a benchmark, which means that there is no coordination with the SON functions and the SON functions operate independently. The detailed simulation parameters are given in Table IV.

Fig. 5 shows the policy convergence of **MTCS** scheme with time iterations. We can observe that the policy of **MTCS** scheme has a quick convergence to obtain the optimal load balancing in small time-scale and to achieve the optimal energy saving policy in large time-scale as shown in Fig. 5. Furthermore, it can be shown that the proposed **MTCS** scheme converges under different learning rate η $(0 < \eta \leq 1)$, and the optimal policy of **MTCS** scheme can be derived quickly with $\eta = 0.9$ and $\epsilon = 0.1$, because the **MTCS** scheme obtains good

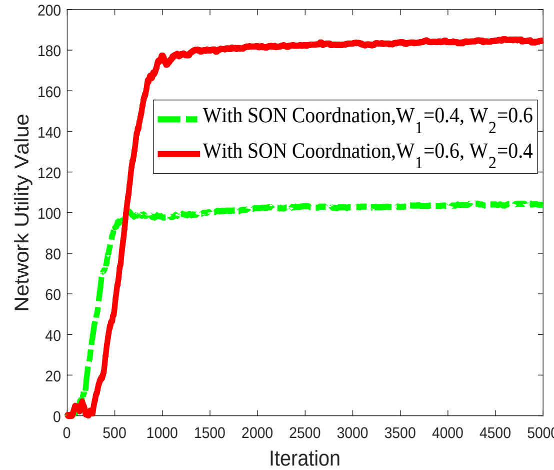


Fig. 6. Network Utility With different weight W_1 and W_2

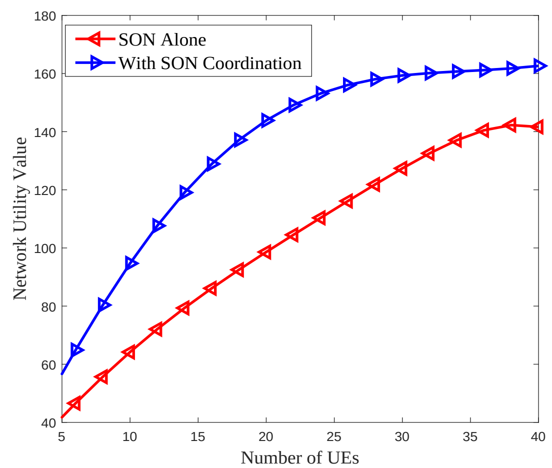


Fig. 7. Network Utility vs. Number of UEs

balance between exploration and exploitation.

Fig. 6 shows the proposed scheme performance with different objective weights. We consider two scenarios as: Capacity Prefer scenario (**CPS**) with $W_1 = 0.6, W_2 = 0.4$ and Energy Prefer scenario (**EPS**) with $W_1 = 0.4, W_2 = 0.6$. It can be seen that the proposed **MTCS** scheme converges quickly both in **CPS** scenario and in **EPS** scenario. Based on the utility function (17), in **CPS**, the network operators prefer to improve the network capacity as the network performance. In **EPS**, the energy saving is taken as the primary network performance of network operators to manage the wireless network, which achieves a lower function value compared with the **CPS**. As a consequence, the proposed **MTCS** algorithm is reasonable and extensible in real systems based on different objectives of network operators, which is suitable for the dynamics of the wireless environment.

We use the SON alone scheme as the benchmark, which means that the ESM and MLB run respectively with no coordination between them. Fig. 7 depicts the network utility

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

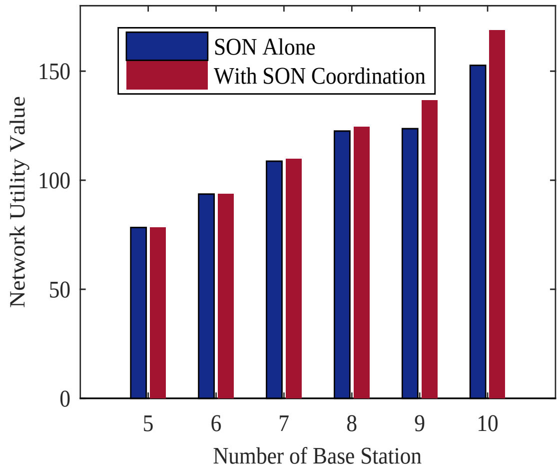


Fig. 8. Network Utility vs. Number of SCBS

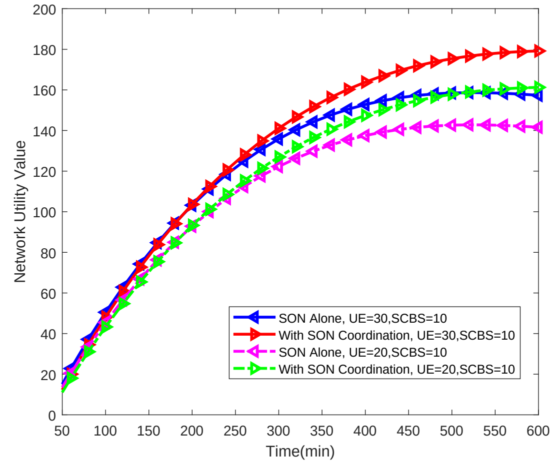


Fig. 10. Network Utility with Times(min)

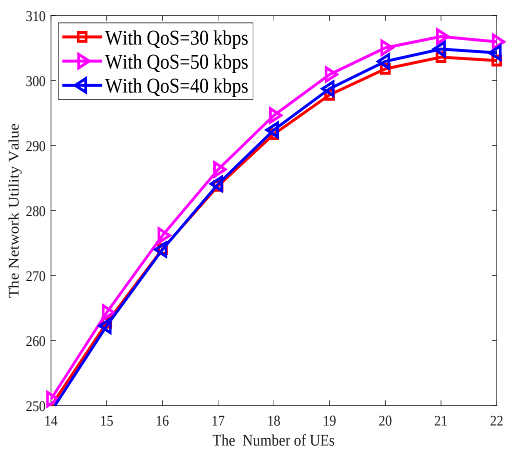


Fig. 9. Network Utility with Different QoS Requirements

in terms of the number of users with $C = 10$. It shows that the performance of proposed algorithm is superior to that of the SON alone scheme. The reason is that the MLB SON is implemented more frequently with the increase of the number of UEs, which leads to more potential conflicts between MLB and ESM, and degrades the contribution of SONs on the network performance improvement. Hence, without **MTCS** coordination algorithm, the wireless network performance increases slowly due to the SON confliction.

Fig. 8 shows the network utility in terms of the number of small cell base stations and investigates the impact of the density of small cells with $U = 40$. It can be seen that the proposed **MTCS** scheme achieves better performance than the SON alone scheme with large amount of small cell base stations, because more potential conflicts are effectively avoided in ultra-dense small cell networks with the SON **MTCS** scheme. Furthermore, the dynamic of small cell network is considered in the propose scheme by learning the history knowledge from the wireless environment, and it shows that the **MTCS** scheme has practical significance in real system.

Fig. 9 shows the network utility in terms of different QoS requirements of users. We set the different QoS requirement scenarios as QoS=30 kbps, QoS=40 kbps and QoS=50 kbps. It can be seen that the proposed **MTCS** scheme obtains good performance with different QoS requirements and the network utility increases along with the QoS requirements. The **MTCS** scheme can meet different users' QoS requirements, which is flexible in dynamic small cell networks.

Furthermore, Fig. 10 shows the network utility of the proposed scheme in terms of the time with $C = 10$. We set that the time-scale of MLB function is 5 minutes and the time-scale of ESM function is 5 hours. It can be seen that the proposed SON coordination scheme derives better performance compared with the SON alone scheme. In addition, we can also see that the performance of **MTCS** scheme achieves better gain than the SON alone scheme with the increase of the number of UEs. In particular, there is no much difference in the performance at beginning. After 300 minutes, the performance is getting much better than SON alone scheme. Because that both MLB (with fast time-scale) and ESM (with slow time-scale) SON functions are triggered and potential conflicts problems arise, the proposed **MTCS** scheme can achieve an optimal policy compared with the SON alone scheme, considering the time domain characteristic of multiple time-scale. In addition, it shows that the **MTCS** scheme has achieved good effects on the network performance with long-term dynamic small cell networks.

VIII. Conclusions

We have developed a comprehensive self-coordination management scheme for SON functions to ensure stable wireless network operation in ultra-dense small cell networks. A novel **MTCS** scheme for SON coordination based on the multiple time-scale MDP model has been proposed to accommodate the SON conflicts with different time-scale by learning from previous experiences, which can improve the network operation efficiency while ensuring a stable operation. In particular, to improve the energy efficiency, a Q-learning algorithm for SON

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/TWC.2019.2898002, IEEE Transactions on Wireless Communications

functions (QSON) has also been proposed to enrich the stable control policy with designed network utility, which is practical and flexible in real system for network operators. For our future work, we will study the big data based programmable SON framework complemented by efficiency improvements brought by artificial intelligence.

References

[1] M. Qin, J. Li, Q. Yang, N. Cheng, K. S. Kwak, and X. Shen,"Multiple Time-Scale SON Function Coordination in Ultra-Dense Small Cell Networks", *In Proc. IEEE GLOBECOM*, Dec. 2018.

[2] N. Zhang, N. Cheng, A. Gamage, K. Zhang, J. W. Mark, and X. Shen, "Cloud assisted HetNets toward 5G Wireless Networks", *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 59-65, Jun. 2015.

[3] A. Imran, A. Zoha, and A. Dayya, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G", *IEEE Netw.* vol. 28, no. 6, pp. 27-33, Nov. 2014.

[4] L. Lyu, C. Chen, S. Zhu, and X. Guan, "5G enabled Co-design of Energy Efficient Transmission and Estimation for Industrial IoT Systems", *IEEE Trans. Ind. Informat.*, DOI:10.1109/TII.2018.2799685, 2018.

[5] N. Cheng, F. Lyu, J. Chen, W. Xu, H. Zhou, S. Zhang, and X. Shen, "Big data Driven Vehicular Networks", *IEEE Netw.*, vol. 32, no. 6, pp. 160-167, Aug. 2018.

[6] P. Klaine, M. Imran, O. Onireti, and R. Souza,"A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks", *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392-2431, Jul. 2017.

[7] A. Tall, R. Combes, Z. Altman, and E. Altman, "Distributed Coordination of Self-Organizing Mechanisms in Communication Networks", *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 4, pp. 328-337, Dec. 2014.

[8] O. Aliu, A. Imran, M. Imran, and B. Evans, "A Survey of Self Organisation in Future Cellular Networks", *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 336-61, Feb. 2013.

[9] X. Gelabert and B. Sayrac, "A Heuristic Coordination Framework for Self-Optimizing Mechanisms in LTE HetNets", *IEEE Trans. Veh. Technol.* vol. 63, no. 3, pp. 1320-1334, Mar. 2014.

[10] J. Moysen, M. Garcia-Lozano, L. Giupponi, and S. Ruiz, "Conflict Resolution in Mobile Networks: A Self-Coordination Framework Based on Non-Dominated Solutions and Machine Learning for Data Analytics", *IEEE Comput. Intell. Mag.*, vol. 13, no. 2, pp. 52-64, Apr. 2018.

[11] S. Mwanje, H. Sanneck, and A. Mitschele-Thiel, "Synchronized Cooperative Learning for Coordinating Cognitive Network Management Functions", *IEEE Trans. Cognitive Commun. Netw.*, DOI: 10.1109/TCCN.2018.2805349, 2018.

[12] M. Hasan, S. Kwon and J. Na, "Adaptive Mobility Load Balancing Algorithm for LTE Small-Cell Networks", *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2205-2217, Jan. 2018.

[13] F. Alfarhan and A. Alsohaily,"Self-Organizing Wireless Nnetwork Parameter Optimization through Mixed Iinteger Programming", *In Proc. of IEEE ICC*, May. 2017, pp. 1-6.

[14] L. Jorguseski, A. Pais, F. Gunnarsson, A. Centonza, and C. Willcock, "Self-Organizing Networks in 3GPP: Standardization and Future Trends", *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 28-34, Dec. 2014,

[15] H. Lateef, A. Imran, M. Imran, L. Giupponi, and M. Dohler, "LTE-Advanced Self-Organizing Network Conflicts and Coordination Algorithms", *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 108-117, Jun. 2015.

[16] A. Zakrzewska, L. Ho, H. Gacanin and H. Claussen, "Coordination of SON Functions in Multi-Vendor Femtocell Networks", *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 165-171, Jul. 2017.

[17] S. Mwanje, L. Schmelz, and A. Thiel, "Cognitive Cellular Networks: A Q-learning Framework for Self-organizing Networks", *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 1, pp. 85-98, Jan. 2016.

[18] P. Munoz, R. Barco, and S. Fortes, "Conflict Resolution between Load Balancing and Handover Optimization in LTE Networks", *IEEE Commun. Lett.* vol. 18, no. 10, pp. 1795-1798, Oct. 2014.

[19] O. Iacoboaiea and B. Sayrac, "SON Coordination in Heterogeneous Networks: A Reinforcement Learning Framework", *IEEE Wireless Commun.*, vol. 15, no. 9, pp. 5835-5847, May 2016.

[20] N. Zhang, S. Zhang, J. Zheng, X. Fang, J. W. Mark, and X. Shen, "QoE Driven Decentralized Spectrum Sharing in 5G Networks: Potential Game Approach", *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7797-7808, Mar. 2017.

[21] S. Zhang, N. Zhang, X. Fang, P. Yang, and X. Shen,"Self-Sustaining Caching Stations: Toward Cost-Effective 5G-Enabled Vehicular Networks", *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 202-208, Aug. 2017.

[22] J. Zheng, Y. Wu, N. Zhang, H. Zhou, Y. Cai, and X. Shen,"Optimal Power Control in Ultra-Dense Small Cell Networks: A Game-Theoretic Approach", *IEEE Wireless Commun.*, vol. 16, no. 7, pp. 4139-4150, Dec. 2016.

[23] N. Cheng, N. Lu, N. Zhang, X. Zhang, X. Shen, and J. W. Mark. "Opportunistic WiFi Offloading in Vehicular Environment: A Game-theory Approach", *IEEE Trans. Intell. Transport. Syst.*, vol. 17, no. 7, pp. 1944-1955, Jul. 2016.

[24] T. Vu, M. Bennis, S. Samarakoon, M. Debbah, and M. Aho," Joint Load Balancing and Interference Mitigation in 5G Heterogeneous Networks", *IEEE Wireless Commun.*, vol. 16, no. 9, pp. 6032-6046, Sept. 2017.

[25] A. Garcia, S. Fortes, A. Duran, and R. Barco,"Context-Aware Self-Optimization: Evolution Based on the Use Case of Load Balancing in Small-Cell Networks", *IEEE Trans. Veh. Technol. Mag.*, vol. 11, no. 1, pp. 86-95, Mar. 2016.

[26] P. Munoz, R. Barco, J. Aviles, I. Cascales, and A. Garcia, "Fuzzy Rule-based Reinforcement Learning for Load Balancing Techniques in Enterprise LTE Femtocells", *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1962-1973, Jun. 2013.

[27] H. Tao, and N. Ansari. "A Traffic Load Balancing Framework for Software-Defined Radio Access Networks Powered by Hybrid Energy Sources", *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1038-1051, Mar. 2016.

[28] X. Chen, J. Wu, Y. Cai, H. Zhang, and T. Chen. "Energy-efficiency Oriented Traffic Offloading in Wireless Networks: A Brief Survey and a Learning Approach for Heterogeneous Cellular Networks", *IEEE J. Sel. Areas Commun.*, vol. 33, no. 4, pp. 627-640, Jan. 2015.

[29] M. Kashef, M. Ismail, E. Serpedin, and K. Qaraqe,"Balanced Dynamic Planning in Green Heterogeneous Cellular Networks", *IEEE J. Sel. Areas Commun.* vol. 34, no. 12, pp. 3299-3312, Nov. 2016.

[30] H. Chang and P. Fard, "Multi-time Scale Markov Decision Processes", *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 976-987, 2003.

[31] 3GPP TR 36.902 v1.2.0, "Self-configuring and Self-optimizing Network Use Cases and Solutions (Release 9)", Mar. 2011.

[32] E. Aqeeli, A. Moubayed, and A. Shami, "Dynamic SON-Enabled Location Management in LTE Networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 7, pp. 1511-1523, 2018.

[33] M. Bennis, S. Perlaza, P. Blasco, Z. Han, and H. V. Poor. "Self-Organization in Small Cell Networks: A Reinforcement Learning Approach", *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202-3212, 2013.

[34] G. Arslan and S. Yuksel," Decentralized Q-Learning for Stochastic Teams and Games", *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1545-1558, Aug. 2017.

[35] W. Wu, Q. Yang, P. Gong, and K. S. Kwak. "Energy-efficient Traffic Splitting for Time-varying Multi-RAT Wireless Networks", *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6523-6535, 2017.

[36] Q. Ye, W. Zhuang, L. Li, and P. Vigneron, "Traffic Load Adaptive Medium Access Control for Fully-connected Mobile Ad Hoc Networks", *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9358-9371, Nov. 2016.

[37] H. Klessig, D. Ohmann, A. Reppas, H. Hatzikirou, M. Abedi, M. Simsek, and G. Fettweis, "From Immune Cells to Self-Organizing Ultra-Dense Small Cell Networks", *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 800-811, Mar. 2016.

[38] L. Qian, Y. Wu, H. Zhou, and X. Shen,"Joint Uplink Base Station Association and Power Control for Small-Cell Networks With Non-Orthogonal Multiple Access", *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5567-5582, Mar. 2017.

Meng Qin received his B.S. degree in Communication Engineering from Taiyuan University of Technology, China in 2012, M.S. degree and Ph.D degree in Information and Communication Systems from Xidian University, China in 2015 and in 2018, respectively. His research interests include AI-aided wireless network operation and management, machine learning, self-organized network, statistical quality of service (QoS) provisioning and applications of stochastic optimization in intelligent wireless networks.

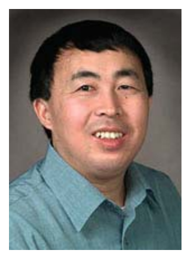
Qinghai Yang received his B.S. degree in Communication Engineering from Shandong University of Technology, China in 1998, M.S. degree in Information and Communication Systems from Xidian University, China in 2001, and Ph. D. in Communication Engineering from Inha University, Korea in 2007 with university-president award. From 2007 to 2008, he was a research fellow at UWB-ITRC, Korea. Since 2008, he is with Xidian University, China . His current research interest lies in the fields of autonomic communication, content delivery networks and LTE-A techniques.

Nan Cheng (S12,M16) received the B.S. degree and the M.S. degree from the Department of Electronics and Information Engineering, Tongji University in 2009 and 2012, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, 2016. He is currently working as a joint Post-doctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto and the Department of Electrical and Computer Engineering, University of Waterloo. He is a joint professor at School of Telecommunication Engineering, Xidian University, China. His research interests include performance analysis, MAC, opportunistic communication for vehicular networks, unmanned aerial vehicles, and cellular traffic offloading.

Jinglei Li received the B.S. degree in Electronic Information Engineering in 2008 from The PLA Information Engineering University, the M.S. and Ph.D. degrees in Communication and Information Systems from Xidian University in 2011 and 2016, respectively. Now he is currently wording at Xidian University. His research interests in wireless network connectivity and node-selfishness management.

Weihua Wu received the B.S. and M.E. degrees in telecommunications engineering and the Ph.D. degree in communication and information systems from Xidian University, China, in 2011, 2014, and 2017, respectively. Since 2018, he has been with the School of Telecommunications Engineering, Xidian University, where he is currently an Assistant Professor. His research interests include wireless resource allocation and stochastic network optimization and their applications in multihoming wireless networks.

Ramesh R. Rao (M'85-SM'90-F'10) has been a faculty member at UC San Diego since 1984, and Director of the Qualcomm Institute, UCSD division of the California Institute for Telecommunications and Information Technology (Calit2), since 2001. He holds the Qualcomm Endowed Chair in Telecommunications and Information Technologies in the Jacobs School of Engineering at UCSD, and is a member of the school's Electrical and Computer Engineering department. Previously, he served as the Director of UCSD's Center for Wireless Communications (CWC). Dr. Rao is an IEEE Fellow and Senior Fellow of the California Council on Science and Technology. Dr. Rao earned his Ph.D. in Electrical Engineering from the University of Maryland, College Park in 1984, after receiving his M.S. from the same institution in 1982. He earned his Bachelors degree in 1980 from the University of Madras (the National Institute of Technology, Tiruchirapalli).



Xuemin (Sherman) Shen (IEEE M'97-SM'02-F09) received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a University Professor and the Associate Chair for Graduate Studies, Department of Electrical and Computer Engineering, University of Waterloo, Canada. Dr. Shen's research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom'16, Infocom'14, IEEE VTC'10 Fall, and Globecom'07, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the General Co-Chair for ACM Mobihoc'15, Chinacom'07 and the Chair for IEEE Communications Society Technical Committee on Wireless Communications. He also serves/served as the Editor-in-Chief for IEEE Internet of Things Journal, IEEE Network, Peerto- Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications; an Associate Editor for IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks, etc.; and the Guest Editor for IEEE JSAC, IEEE Wireless Communications, and IEEE Communications Magazine, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.