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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

Joint Traffic Scheduling and Resource Allocations for Traffic Offloading with Secrecy-Provisioning

Yuan Wu *Senior Member IEEE*, Jianchao Zheng *Member IEEE*, Kuanyang Guo, Li Ping Qian *Senior Member IEEE*, Xuemin (Sherman) Shen *Fellow IEEE*, Yueming Cai *Senior Member IEEE*

***Abstract*—The recent paradigm of small cell dual-connectivity (DC) provides a promising solution to facilitate mobile users' (MUs') traffic offloading in heterogeneous networks. With DC, a MU can flexibly schedule its traffic to macrocell base station (mBS) and offload data to small-cell access point (sAP). However, a malicious node might intentionally eavesdrop the MU's offloaded data, which could lead to the secrecy exposure. In this paper, we investigate the optimal resource allocation for the MUs' traffic offloading via DC with guaranteed secrecy. First, we study a single-MU single-sAP case and formulate a joint optimization of the MU's traffic scheduling, power allocation, and bandwidth usage for traffic offloading, which aims to minimize the MU's overall resource usage including the power consumption and bandwidth usage. Although the joint optimization problem is nonconvex, we propose an efficient algorithm to obtain the optimal offloading solution. Second, by using the single-MU's optimal offloading solution, we study the multi-MU multi-sAP case and formulate an optimal offloading-selection problem that aims to maximize the overall served MUs' throughput with guaranteed secrecy, while taking into account the mBS's and sAPs' limited bandwidths and the sAPs' limited backhaul capacities. Despite the NP-hardness of the formulated offloading-selection problem, we propose an efficient heuristic algorithm to achieve the offloading-selection solution. Numerical results are provided to validate the performance gain of the proposed traffic offloading schemes with guaranteed secrecy.**

I. Introduction

The rapidly increasing penetration of smart mobile devices running a vast array of data-hungry mobile applications has yielded an exponential growth of mobile traffic. Actively offloading mobile users' (MUs') tremendous traffic demands to small cells (such as femtocells and picocells) has emerged as an effective and cost-efficient manner to tackle with the rapid traffic growth in cellular networks. Thanks to the close-proximity between the MUs and small cells, traffic offloading via small cells can bring multi-folded benefits, such as reducing the power consumption and enhancing the throughput. Thus, traffic offloading, especially the associated radio resource management, has attracted lots of interests [1]. Recently, traffic offloading via small-cell dual-connectivity (DC) emerges as a new paradigm advocated in the 3GPP standards [2]. The key feature of the DC is that it allows a MU to simultaneously communicate with a macrocell base station (mBS) and offload data to a small-cell access point (sAP), which thus enables a flexible traffic offloading. For instance, the MU can offload its delay-tolerable traffic to the sAP while sending the delay-sensitive traffic to the mBS. Nevertheless, due to activating two interfaces simultaneously, we need to properly allocate the MU's radio resource usages (e.g., the bandwidth usage and the transmit-power) at two radio interfaces in order to match with the MU's scheduled traffic to the mBS and sAP. Although there exist several studies about resource management for the DC-assisted traffic offloading [3]–[7], it is still an open issue about how to jointly optimize the MU's offloading decisions, i.e., the traffic scheduling, power allocation, and bandwidth usage at the mBS and sAP, for minimizing the MU's overall resource consumption.

In addition to the radio resource consumption, there also exists secrecy exposure issue in the MU's traffic offloading. Specifically, due to the growing need for using unlicensed spectrums to support traffic offloading (such as the emerging paradigm of LTE over unlicensed bands (LTE-U) and the prevalent WiFi systems operating on unlicensed bands), a malicious node might exploit the open-access nature of unlicensed bands and eavesdrop the MU's offloaded traffic (which is delivered over the unlicensed channel) to the sAP. As a result, if a MU blindly offloads too much data, it might suffer from a risk of being eavesdropped. However, securing the MU's offloading is challenging, since it strongly depends on the MU's offloading decisions, e.g., traffic scheduling, power allocation, and bandwidth usage. Moreover, even with a given upper-limit for the MU's secrecy outage, choosing the *right* secrecy-level for the MU is still complicated. Specifically, choosing a too strong secrecy-level consumes the MU a heavy resource consumption, while choosing a too weak secrecy-level yields a large loss in the MU's effective (secure) traffic due to being overheard. Viewing the even crowded and congested licensed bands, the DC-assisted traffic offloading through small cells that use unlicensed bands will be an effective and resource-efficient approach to accommodate the tremendous traffic demands. Accordingly, we need a proper design of resource management to secure the MU's traffic offloading and to optimize the associated resource consumption, which thus motivates our study with the following contributions.

(Single-MU Single-sAP Case): We first consider a single-MU single-sAP case and formulate a novel joint optimization of the MU's traffic scheduling, power allocation, and bandwidth usage for traffic offloading via DC. The formulation is to minimize the MU's overall resource consumption while satisfying the MU's traffic demand and secrecy requirement.

(Effective Algorithm Design): Despite the nonconvexity of the

Y. Wu, K. Guo, and L. Qian are with College of Information Engineering, Zhejiang University of Technology, Hangzhou, China (emails: e-cewuy@alumni.ust.hk, lpqian@zjut.edu.cn). Y. Wu is also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xian, 710071, P.R. China. L. Qian is the corresponding author.

J. Zheng and Y. Cai are with Institute of Communications Engineering, PLA University of Science and Technology, Nanjing, P.R. China (emails: longxingren.zjc.s@163.com, caiym@vip.sina.com).

X. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (email: xshen@bbcr.uwaterloo.ca).

This work was supported in part by the National Natural Science Foundation of China under Grant 61572440, in part by Zhejiang Provincial Natural Science Foundation of China under Grants LR17F010002 and LR16F010003, and in part by the Natural Sciences and Engineering Research Council, Canada. This work was presented, in part, in the 8th International Conference on Wireless Communications and Signal Processing 2016.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

joint optimization problem, we propose an effective algorithm to compute the optimal offloading solution. Exploiting the layered structure of the joint optimization problem, we identify the hidden convexity of the traffic scheduling and power allocation subproblem and the monotonicity of the bandwidth usage subproblem, and efficiently compute the optimal offloading solution.

(Multi-MU Multi-sAP Case): We further consider a multi-MU multi-sAP case. By using the single-MU's optimal offloading solution computed before, we formulate an offloading-selection problem regarding which MUs select which sAPs to form DC-pairs, with the objective of maximizing the overall successfully served MUs' throughput. The formulation guarantees to meet each served MU's traffic and secrecy requirements, while taking into account the mBS's and sAPs' limited bandwidths and the sAPs' limited backhaul capacities. Despite the NP-hardness of the formulated problem (as a multiple-resource generalized assignment problem [41]), we propose an efficient heuristic algorithm to compute the offloading-selection solution.

(Performance Evaluations): We perform extensive numerical tests to evaluate our proposed algorithms and offloading schemes. The results validate the accuracy of our algorithms to compute the offloading solutions. The results verify the performance gains of our proposed offloading scheme with guaranteed secrecy.

The rest of this paper is organized as follows. After reviewing the related studies in Section II, we study the single-MU single-sAP case from Sections III to VI. We then study the multi-MU multi-sAP case in Section VII. We show the numerical results in Section VIII and conclude this work in Section IX.

II. Related Studies

(Studies on Traffic Offloading): Resource allocations for traffic offloading in small cells have gained lots of interests [1]. The related studies can be categorized according to different paradigms, namely, i) offloading via low-powered small cells [8]–[17], ii) offloading via WiFi networks [18], [19], and iii) traffic offloading via users' device-to-device (D2D) cooperation [20]. We mainly focus on the first group in the following, since our study here belongs to the first group. Properly offloading mobile users (or traffic) from macrocells to small cells and allocating radio resources are critical to the success of traffic offloading via small cells. In [8], Wang *et al.* considered users' different QoS requirements and investigated the optimal user-association with small cells and the consequent antenna allocations in massive MIMO systems. Energy-efficiency oriented traffic offloading has been studied in [9], [10]. In [9], Chen *et al.* proposed an energy-efficient oriented dynamic optimization model for offloading users by controlling the on-off states of small cells. In [10], Zhang *et al.* considered the renewable energy powered small cells and proposed a corresponding energy-efficient scheme to offload traffic to small cells. The issues of inter-cell interference mitigation and handover management for offloading traffic to small cells have been studied in [11]. Stochastic geometric model has been used to analyzed the performance (e.g., spectrum efficiency) of traffic offloading through small cells [12], [13]. Recently, an adaptive medium access control scheme based on traffic load has been proposed in [14]. Limited backhaul capacity is an important issue that influences the performance of traffic offloading [15], [16]. In [15], Yang *et al.* assumed the fixed backhaul capacity of small cells and proposed an efficient joint user-admission and power allocation scheme for traffic offloading. In [16], Liu *et al.* proposed the joint backhaul capacity allocation and power allocation for small cell networks. In addition, different incentive-based schemes for motivating traffic offloading have been studied, e.g., [17].

(Studies on Traffic Offloading via DC): The paradigm of data offloading via DC has attracted particular interests since it can improve mobile user's throughput by simultaneously using two radio interfaces. However, due to activating two radio interfaces, the associated resource allocations (or resource-splitting for two interfaces) necessitate a proper design [2]. In [3], Jha *et al.* provided a survey on the DC-assisted traffic offloading in heterogenous networks and carried out simulations to evaluate the influence of resource-splitting on the user's throughput. In [4], to facilitate the MU's uplink data offloading through DC, Liu *et al.* proposed a scheme to split the MU's power capacity for two radio interfaces. In [5], Pan *et al.* proposed a downlink traffic scheduling scheme to maximize the overall throughput. In [6], targeted for the downlink traffic offloading via DC, Wang *et al.* proposed a traffic scheduling scheme by taking into account the backhaul delay. In [7], we proposed a joint traffic scheduling and power allocation scheme for MUs' uplink data offloading via DC to minimize all MUs' economic cost of mobile data. Despite sharing the similarity as resource allocation for multi-RAT (e.g., [21]), the feature of offloading via DC lies in that we need to schedule the MUs' traffic to small cells and macrocells, and allocate radio resources accordingly based on the MUs' QoS and secrecy requirements, which are addressed by our study.

We finally provide a brief review about the related studies on physical layer security, since we will adopt it to model how secure the MU's traffic offloading is. The physical layer secrecy capacity, defined as the difference between the channel capacity of legitimate channel and that of the eavesdropper's channel, provides a fundamental measure of the achievable capacity for a point-to-point link which is impossible to be eavesdropped [22]. With the physical layer secrecy capacity as a meaningful measure, many studies have been carried out to investigate secrecy based communications for cellular systems [23], [24]. For instance, in [23], Wang *et al.* analyzed the achievable secure rate of an arbitrary downlink transmission in cellular networks. In [24], Zhu *et al.* analyzed the secure transmission rate in multi-cell massive MIMO systems. By further accounting for the unknown and random locations of eavesdroppers, the secrecy outage probability, which is based on the physical layer secrecy capacity, emerges as a useful measure to analyze the secrecy based transmission in wireless networks [25], [26]. The secrecy outage probability essentially quantifies the probability (or equivalently, the ratio) of the user's traffic overheard by an eavesdropper at a given transmission rate. For instance, in [25], Wu *et al.* derived the secrecy outage probability in downlink transmission in multi-RAT networks with the BS-association constraint. In [26], Yue *et al.* adopted the secrecy outage probability and proposed the secrecy-based power control scheme for underlaying D2D communications. In this paper, we use the secrecy outage probability to characterize how secure it is when the MU offloads (and sends) data to the sAP (and mBS).

It is worth emphasizing that, in addition to the physical-layer security, many application-oriented secrecy-enhancing schemes (such as authentication for access control and encryption for signalling) have been proposed for LTE systems [27]. Many studies have also been devoted to investigating the security issues in Vehicular Ad-Hoc Networks (VANET) [28]–[33]. For instance, in [28], Qian *et al.* proposed a secure and application-oriented design framework for VANETs. A secure MAC protocol was proposed in [30] to timely and reliably disseminate the safety-related and application-related messages in VANETs.

III. System Model and Problem Formulation for Single-MU Single-sAP Case

Figure 1 shows the considered system model of MUs' traffic offloading with DC. We first study the single-MU single-sAP case in this section (by using MU i and sAP k as an example). We then extend to the multi-MU multi-sAP case in Section VII.

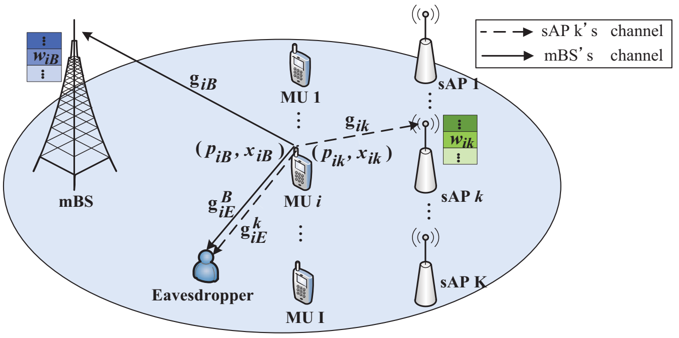


Fig. 1: System model of the MUs' traffic offloading with DC.

We consider that MU i splits its traffic demand to the mBS and offloads the rest to sAP k simultaneously via DC. We use x_{iB} and p_{iB} to denote MU i's traffic rate and the transmit-power to the mBS, respectively (here, the letter "B" stands for "Base Station"). We use w_{iB} to denote MU i's required channel bandwidth at the mBS. Meanwhile, we use x_{ik} and p_{ik} to denote MU i's offloading-rate and the transmit-power to sAP k, respectively, and we use w_{ik} to denote MU i's (required) channel bandwidth at sAP k. We point out that all (x_{iB}, p_{iB}, w_{iB}) and (x_{ik}, p_{ik}, w_{ik}) are the decision variables in our following problem formulation. We assume that the mBS and sAP operate on different spectrums. Thus, there is no interference between MU i's transmissions to the mBS and sAP k.

There exists an eavesdropper which overhears MU i's traffic to sAP k (and the mBS) by receiving MU i's radio signal. To quantify such a secrecy risk, we adopt the physical layer security capacity and derive the MU's secrecy outage probability when transmitting to the sAP and mBS in the next subsection. Note that the considered eavesdropper in this model can be regarded as the one exhibiting the most significant influence on MU i's transmission, if there exist several eavesdroppers.

A. Modeling of MU's Secrecy Outage in Traffic Offloading

According to [22], the achievable secrecy-rate r^{sec}_{ik} from MU i to sAP k can be given by the difference between the channel capacity from MU i to sAP k and that from MU i to the eavesdropper, i.e.,

$$r^{\text{sec}}_{ik} = \max\left\{w_{ik}\log_2\left(1+\frac{p_{ik}g_{ik}}{w_{ik}n_0}\right) - w_{ik}\log_2\left(1+\frac{p_{ik}g^k_{iE}}{w_{ik}n_0}\right), 0\right\},$$

where g_{ik} denotes the channel power gain from MU i to sAP k, and g^k_{iE} denotes the channel power gain from MU i to the eavesdropper over sAP k's channel. Parameter n_0 denotes the density of the background noise at sAP k and the eavesdropper.

With advanced wireless transceivers, MU i can get accurate information about g_{ik} to sAP k and g_{iB} to the mBS through proper feedback when the channel conditions change relatively slowly. This is in fact our target application scenario (i.e., the MUs with relatively static channel conditions), as it will be difficult to perform effective traffic offloading for fast moving MUs due to the limited coverage of small cells. However, accurate information about the eavesdropper's channel gain is difficult to obtain, since the eavesdropper might intentionally conceal such information. Hence, we adopt a similar assumption as in [25], [26], [34], [35], namely, only the statistics information about the eavesdropper's channel is available, and the channel power gain g^k_{iE} can be considered to follow an exponential distribution with a mean α_{ik}. Due to the randomness of g^k_{iE}, we can express MU i's secrecy-outage probability $P^k_{\text{out}}(p_{ik}, x_{ik}, w_{ik})$ (as a function of its offloading-rate x_{ik}, transmit-power p_{ik}, and bandwidth usage w_{ik}) when offloading data to sAP k as follows:

$$\begin{aligned} &P^k_{\text{out}}(p_{ik}, x_{ik}, w_{ik}) = 1 - \mathbf{Pr}\Big\{r^{\text{sec}}_{ik} \geq x_{ik} | \\ &w_{ik}\log_2\left(1+\frac{p_{ik}g_{ik}}{w_{ik}n_0}\right) - w_{ik}\log_2\left(1+\frac{p_{ik}g^k_{iE}}{w_{ik}n_0}\right) \geq 0\Big\} \\ &= 1 - \mathbf{Pr}\Big\{g^k_{iE} \leq 2^{-\frac{x_{ik}}{w_{ik}}}g_{ik} - \left(1-2^{-\frac{x_{ik}}{w_{ik}}}\right)\frac{w_k n_0}{p_{ik}} | g^k_{iE} \leq g_{ik}\Big\}. \end{aligned} \tag{1}$$

Eq. (1) takes into account that MU i chooses to offload traffic to sAP k only when it expects to receive a nonnegative secrecy-rate. Since $x_{ik} \geq 0$, we have that $g^k_{iE} \leq 2^{-\frac{x_{ik}}{w_{ik}}}g_{ik} - \left(1-2^{-\frac{x_{ik}}{w_{ik}}}\right)\frac{w_{ik}n_0}{p_{ik}}$ is a sufficient condition to guarantee $g^k_{iE} \leq g_{ik}$. Hence, we can further derive MU i's secrecy-outage probability $P^k_{\text{out}}(p_{ik}, x_{ik}, w_{ik})$ in (1) as follows:

$$P^k_{\text{out}}(p_{ik}, x_{ik}, w_{ik}) = \frac{e^{-\frac{2^{-\frac{x_{ik}}{w_{ik}}}g_{ik} - (1-2^{-\frac{x_{ik}}{w_{ik}}})\frac{w_{ik}n_0}{p_{ik}}}{\alpha_{ik}}} - e^{-\frac{g_{ik}}{\alpha_{ik}}}}{1 - e^{-\frac{g_{ik}}{\alpha_{ik}}}}. \tag{2}$$

Eq. (2) leads to $P^k_{\text{out}}(p_{ik}, 0, w_{ik}) = 0$, i.e., the secrecy outage probability is zero when MU i does not offload any traffic to sAP k, which is consistent with the intuition. In particular, we consider that MU i, when offloading traffic, requires its secrecy-outage to be no larger than a given limit $\epsilon^{\text{max}}_{ik} \in (0, 1]$, i.e., $P^k_{\text{out}}(p_{ik}, x_{ik}, w_{ik}) \leq \epsilon^{\text{max}}_{ik}$.

To make the following formulation general, we consider that MU i's transmission to the mBS suffers from a similar secrecy outage as offloading to sAP k. As we will illustrate in Remark 1, this modeling can capture the situation that MU i's transmission to the mBS is perfectly secure (e.g., MU i transmits to the mBS over a licensed channel which is completely unknown to the eavesdropper). Similar to r^{sec}_{ik} before, we derive the achievable secrecy-rate r^{sec}_{iB} from MU i to the mBS as follows:

$$r^{\text{sec}}_{iB} = \max\left\{w_{iB}\log_2\left(1+\frac{p_{iB}g_{iB}}{w_{iB}n_0}\right) - w_{iB}\log_2\left(1+\frac{p_{iB}g^B_{iE}}{w_{iB}n_0}\right), 0\right\},$$

where g_{iB} denotes the channel power gain from MU i to the mBS, and g^B_{iE} denotes the channel power gain from MU i to the eavesdropper.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

Similar to g_{iE}^k, we assume that g_{iE}^B follows an exponential distribution with mean α_{iB}. Thus, we derive MU i's secrecy-outage probability $P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB})$ when transmitting to the mBS (as function of MU i's data rate x_{iB}, power allocation p_{iB}, and bandwidth usage w_{iB}) as follows:

$$P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB}) = 1 - \mathbf{Pr}\Big\{ r_{iB}^{\text{sec}} \geq x_{iB} | \\ w_{iB}\log_2\left(1+\frac{p_{iB}g_{iB}}{w_{iB}n_0}\right) - w_{iB}\log_2\left(1+\frac{p_{iB}g_{iE}^B}{w_{iB}n_0}\right) \geq 0 \Big\} \\ = \frac{e^{-\frac{2^{-\frac{x_{iB}}{w_{iB}}}g_{iB}-(1-2^{-\frac{x_{iB}}{w_{iB}}})\frac{w_{iB}n_0}{p_{iB}}}{\alpha_{iB}}} - e^{-\frac{g_{iB}}{\alpha_{iB}}}}{1-e^{-\frac{g_{iB}}{\alpha_{iB}}}}. \quad (3)$$

We consider that MU i, when sending data to the mBS, requires the suffered secrecy-outage to be no larger than a given limit $\epsilon_{iB}^{\max} \in (0,1]$, i.e., $P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB}) \leq \epsilon_{iB}^{\max}$.

B. Problem Formulation

Using (2) and (3), we formulate the secrecy-constrained optimization of Traffic Scheduling, Power Allocation and Bandwidth Usage as follows:

$$\begin{aligned}
\underline{\textbf{(TSPABU)}}: \quad & \min \ p_{ik} + p_{iB} + \mu_A w_{ik} + \mu_B w_{iB} \\
\text{Subject to:} \quad & P_{\text{out}}^k(p_{ik}, x_{ik}, w_{ik}) \leq \epsilon_{ik}^{\max}, && (4)\\
& P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB}) \leq \epsilon_{iB}^{\max}, && (5)\\
& \left(1 - P_{\text{out}}^k(p_{ik}, x_{ik}, w_{ik})\right) x_{ik} + \\
& \left(1 - P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB})\right) x_{iB} = R_i^{\text{req}}, && (6)\\
& 0 \leq p_{ik} \leq p_{ik}^{\max}, && (7)\\
& 0 \leq p_{iB} \leq p_{iB}^{\max}, && (8)\\
& w_{ik}^{\min} \leq w_{ik} \leq w_{ik}^{\max}, && (9)\\
& w_{iB}^{\min} \leq w_{iB} \leq w_{iB}^{\max}, && (10)\\
& x_{ik} \geq 0, x_{iB} \geq 0, \\
\text{Variables:} \quad & \boldsymbol{x}_i, \boldsymbol{p}_i, \text{ and } \boldsymbol{w}_i.
\end{aligned}$$

Problem (TSPABU) aims at minimizing the overall cost including MU i's total transmit-power and the bandwidth usage cost, where μ_k and μ_B denote the respective unit costs for the bandwidth usage at sAP k and mBS. In Problem (TSPABU), we jointly optimize: i) MU i's data-rate x_{iB} and the transmit-power p_{iB}, and the bandwidth usage w_{iB} to the mBS, and ii) MU i's offloading-rate x_{ik} and the transmit-power p_{ik}, and the bandwidth usage w_{ik} to sAP k. For simplicity, we use vectors $\boldsymbol{x}_i = \{x_{ik}, x_{iB}\}$, $\boldsymbol{p}_i = \{p_{ik}, p_{iB}\}$, and $\boldsymbol{w}_i = \{w_{ik}, w_{iB}\}$ to denote the decision variables.

In Problem (TSPABU), (4) ensures MU i's secrecy-outage probability when offloading data to sAP k no greater than the upper-limit $\epsilon_{ik}^{\max}$. Similarly, (5) ensures MU i's secrecy-outage probability when transmitting to the mBS not greater than the upper-limit $\epsilon_{iB}^{\max}$. Constraint (6), taking into account both MU i's suffered secrecy-outages to the mBS and sAP k, guarantees that MU i receives an effective secrecy-rate equal to its requirement R_i^{req}. (7) and (8) ensure that MU i's transmit-powers p_{ik} to sAP k and p_{iB} to the mBS cannot exceed their respective upper-bounds $p_{ik}^{\max}$ and $p_{iB}^{\max}$. (9) ensures that MU i's bandwidth usage at sAP k is lower-bounded by $w_{ik}^{\min}$ and upper-bounded by $w_{ik}^{\max}$, and (10) ensures that MU i's bandwidth usage at the mBS is lower-bounded by $w_{iB}^{\min}$ and upper-bounded by $w_{iB}^{\max}$.

C. Decomposition Structure of Problem (TSPABU)

Problem (TSPABU) is a nonconvex optimization problem, due to the nonconvex feasible region yielded by (4), (5), and (6). Moreover, (4), (5), and (6) strongly couple the decision variables $\boldsymbol{x}_i$,$\boldsymbol{p}_i$, and $\boldsymbol{w}_i$, which prevent us using conventional Lagrangian method. To solve Problem (TSPABU), it is necessary to exploit the unique properties of Problem (TSPABU).

The key idea to solve Problem (TSPABU) is based on the following important finding. Suppose that MU i's bandwidth usage $\boldsymbol{w}_i$ is given, the resulting subproblem regarding MU i's traffic scheduling $\boldsymbol{x}_i$ and power allocation $\boldsymbol{p}_i$ exhibits a hidden convexity (after executing some equivalent transformations), based on which we can design an efficient algorithm to solve it (we will illustrate the details in Section IV). Based on this finding, we vertically decompose Problem (TSPABU) into two problems to solve as follows.

The problem on the top-level is the MU's bandwidth usage optimization to optimize $\boldsymbol{w}_i$ as follows:

$$\begin{aligned}
\underline{\textbf{(BU)}}: \quad & \min \ C(\boldsymbol{w}_k) + \mu_k w_{ik} + \mu_B w_{iB} \\
\text{Subject to:} \quad & \text{constraints } (9) \text{ and } (10), \\
\text{Variables:} \quad & \boldsymbol{w}_i.
\end{aligned}$$

Notice that in the objective function of Problem (BU), we need to evaluate function $C(\boldsymbol{w}_k)$ for each given $\boldsymbol{w}_k$. We can evaluate $C(\boldsymbol{w}_k)$ as follows.

The problem in the bottom-level is the MU's traffic scheduling and power allocation to optimize $\boldsymbol{x}_i$ and $\boldsymbol{p}_i$ as follows:

$$\begin{aligned}
\underline{\textbf{(TSPA)}}: \quad & C(\boldsymbol{w}_i) = \min \ p_{ik} + p_{iB} \\
\text{Subject to:} \quad & \text{constraints } (4), (5), (6), (7), \text{ and } (8), \\
\text{Variables:} \quad & \boldsymbol{x}_i \text{ and } \boldsymbol{p}_i.
\end{aligned}$$

Recall that $\boldsymbol{w}_i$ is given in (4), (5), (6) in Problem (TSPA).

As we will illustrate shortly, the key advantage of decomposition into Problem (TSPA) and Problem (BU) is that we can exploit the special properties of Problem (TSPA) and Problem (BU) for computing the optimal solution.

D. A Further Decomposition of Problem (TSPA)

Even with a given $\boldsymbol{w}_i$, Problem (TSPA) is still difficult to solve, again due to the nonconvexity in (4), (5), and (6).

The key to solve Problem (TSPA) is as follows. Suppose that we fix $P_{\text{out}}^k(p_{ik}, x_{ik}, w_{ik})$ and $P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB})$ in advance, e.g., let us say $P_{\text{out}}^k(p_{ik}, x_{ik}, w_{ik}) = \epsilon_{ik}$ and $P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB}) = \epsilon_{iB}$. Then, the resulting problem can be equivalently transformed into a convex optimization problem, which is easy to solve. Specifically, the resulting subproblem under a given $(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$, with $\boldsymbol{\epsilon}_i = (\epsilon_{iB}, \epsilon_{ik})$, is as follows.

$$\begin{aligned}
\underline{\textbf{(TSPA-Sub)}}: \quad & V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i) = \min \ p_{ik} + p_{iB} \\
\text{Subject to:} \quad & P_{\text{out}}^k(p_{ik}, x_{ik}, w_{ik}) = \epsilon_{ik}, && (11)\\
& P_{\text{out}}^B(p_{iB}, x_{iB}, w_{iB}) = \epsilon_{iB}, && (12)\\
& (1-\epsilon_{ik})x_{ik} + (1-\epsilon_{iB})x_{iB} = R_i^{\text{req}}, && (13)\\
& \text{constraints } (7) \text{ and } (8), \\
\text{Variables:} \quad & \boldsymbol{p}_i \text{ and } \boldsymbol{x}_i.
\end{aligned}$$

We use function $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ to denote the optimal objective value of Problem (TSPA-Sub) which depends on the given $\boldsymbol{w}_i$ and $\boldsymbol{\epsilon}_i$. We will show how to solve Problem (TSPA-Sub) in Section IV.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

By using $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ and further executing a two dimensional line-search for $(\epsilon_{ik}, \epsilon_{iB}) \in [0, \epsilon_{ik}^{\max}] \times [0, \epsilon_{iB}^{\max}]$, we can thus solve the original Problem (TSPA) and obtain the value of $C(\boldsymbol{w}_i)$. This corresponds to solve the following top problem to optimize $\boldsymbol{\epsilon}_i = (\epsilon_{ik}, \epsilon_{iB})$ as follows

$$\textbf{(TSPA-Top):} \quad C(\boldsymbol{w}_i) = \min \quad V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$$

$$\text{Subject to:} \quad 0 \le \epsilon_{ik} \le \epsilon_{ik}^{\max}, \tag{14}$$

$$0 \le \epsilon_{iB} \le \epsilon_{iB}^{\max}, \tag{15}$$

$$\text{Variables:} \quad \boldsymbol{\epsilon}_i.$$

The optimal objective value $C(\boldsymbol{w}_i)$ of Problem (TSPA-Top) is exactly the optimal function value of Problem (TSPA). $C(\boldsymbol{w}_i)$ will be used in Problem (BU).

Figure 2(a) shows the decomposition of Problem (TSPABU) into Problems (BU) and (TSPA), which is further decomposed into Problems (TSPA-Top) and (TSPA-Sub). We solve Problems (TSPA-Sub), (TSPA-Top), and (BU) in the next Sections IV-VI.

IV. Algorithm to Solve (TSPA-Sub)

We first solve Problem (TSPA-Sub) in this section. The key is to equivalently transform Problem (TSPA-Sub) into a convex optimization problem (namely, Problem (TSPA-Sub-E)) as shown in subsections IV-A and IV-B. Figure 2(a) shows the connection between Problems (TSPA-Sub) and (TSPA-Sub-E).

A. Transformations at sAP-side & Effective Secrecy-Rate

By putting (2) into (11) and performing some transformations, MU i's offloading-rate to the sAP should satisfy

$$x_{ik} = w_{ik} \log_2\left(\frac{p_{ik} g_{ik} + w_{ik} n_0}{p_{ik}\theta_{ik} + w_{ik} n_0}\right), \tag{16}$$

where θ_{ik} is a newly introduced parameter and is given by

$$\theta_{ik} = -\alpha_{ik} \ln\left(1 - (1 - e^{-\frac{g_{ik}}{\alpha_{ik}}})(1 - \epsilon_{ik})\right). \tag{17}$$

Parameter θ_{ik} represents an aggregate strength of security-requirement, which captures the MU's channel gain g_{ik}, the eavesdropper's average channel strength α_{ik}, and MU i's secrecy-outage limit ϵ_{ik}. Notice that $\theta_{ik} > 0$ always holds. Since θ_{ik} plays a key role in the following analysis, we provide three important properties for it as follows.

Lemma *1: (Properties of Parameter θ_{ik})* The following three properties always hold: i) θ_{ik} is decreasing in ϵ_{ik}; ii) θ_{ik} is increasing in α_{ik}; iii) $\theta_{ik} \le g_{ik}$ always holds.

Proof: Properties i) and iii) can be proved based on the definition of θ_{ik} in (17). We skip the details. We prove Property ii) as follows. Specifically, we can derive

$$\frac{d\ \theta_{ik}}{d\ \alpha_{ik}} = -\ln\left(1 - (1 - e^{-\frac{g_{ik}}{\alpha_{ik}}})(1 - \epsilon_{ik})\right) - \frac{e^{-\frac{g_{ik}}{\alpha_{ik}}} g_{ik}(1 - \epsilon_{ik})}{\alpha_{ik}\left(1 - (1 - e^{-\frac{g_{ik}}{\alpha_{ik}}})(1 - \epsilon_{ik})\right)}. \tag{18}$$

Using (18), we can further derive

$$\frac{d^2\ \theta_{ik}}{d\ \alpha_{ik}^2} = -\frac{e^{-\frac{g_{ik}}{\alpha_{ik}}} g_{ik}(1 - \epsilon_{ik})}{\alpha_{ik}^2\left(1 - (1 - e^{-\frac{g_{ik}}{\alpha_{ik}}})(1 - \epsilon_{ik})\right)^2} \cdot \frac{g_{ik}\epsilon_{ik}}{\alpha_{ik}} < 0,$$

which implies that $\frac{d\ \theta_{ik}}{d\ \alpha_{ik}}$ (given in (18)) is decreasing in α_{ik}. Furthermore, due to $\lim_{\alpha_{ik} \to +\infty} \frac{d\ \theta_{ik}}{d\ \alpha_{ik}} = 0$, we conclude $\frac{d\ \theta_{ik}}{d\ \alpha_{ik}} \ge 0$, meaning that θ_{ik} is increasing in α_{ik}. ■

Lemma 1 matches with the intuition. Since θ_{ik} captures the strength of security requirement, it is reasonable that: i) θ_{ik} decreases with a larger ϵ_{ik} (i.e., a less stringent limit on the secrecy-outage), and ii) θ_{ik} increases with a larger α_{ik} (i.e., a stronger capability of eavesdropper to overhear the data).

In particular, by taking into account the given outage-probability ϵ_{ik}, we denote the MU's effective secrecy-rate $r_{ik} \in [0, R_i^{\text{req}}]$ offloaded to sAP k as follows:

$$r_{ik} = (1 - \epsilon_{ik}) x_{ik} = (1 - \epsilon_{ik}) w_{ik} \log_2\left(\frac{p_{ik} g_{ik} + w_{ik} n_0}{p_{ik}\theta_{ik} + w_{ik} n_0}\right). \tag{19}$$

Furthermore, by using (16), p_{ik} can be expressed as:

$$p_{ik} = w_{ik} n_0 \frac{F(r_{ik}) - 1}{g_{ik} - \theta_{ik} F(r_{ik})}, \tag{20}$$

where $F(r_{ik})$ is defined as follows:

$$F(r_{ik}) = 2^{\frac{r_{ik}}{(1 - \epsilon_{ik}) W_{ik}}}. \tag{21}$$

By substituting (20) into (7), and using that $0 \le r_{ik} \le R_i^{\text{req}}$, we can derive the upper-limit of r_{ik} as follows:

$$r_{ik} \le r_{ik}^{\text{upper}} = \min\left\{R_i^{\text{req}}, (1 - \epsilon_{ik}) w_{ik} \log_2\left(\frac{p_{ik}^{\max} g_{ik} + w_{ik} n_0}{p_{ik}^{\max}\theta_{ik} + w_{ik} n_0}\right)\right\}. \tag{22}$$

B. Transformations at mBS-side & Effective Secrecy-Rate

With the similar procedures for the sAP-side, we further perform the transformations for the mBS-side. By putting (3) into (12), we derive the MU's rate to the mBS as:

$$x_{iB} = w_{iB} \log_2\left(\frac{p_{iB} g_{iB} + w_{iB} n_0}{p_{iB}\theta_{iB} + w_{iB} n_0}\right), \tag{23}$$

where θ_{iB} is a newly introduced parameter given by:

$$\theta_{iB} = -\alpha_{iB} \ln\left(1 - (1 - e^{-\frac{g_{iB}}{\alpha_{iB}}})(1 - \epsilon_{iB})\right). \tag{24}$$

Parameter θ_{iB}, which has the similar meaning as θ_{ik} (at the sAP-side), represents an aggregate strength of security-requirement when MU i transmits to the mBS. θ_{iB} captures the MU's channel gain g_{iB}, the eavesdropper's average channel strength to the mBS α_{iB}, and the MU's secrecy-outage limit ϵ_{iB}. Similar to Lemma 1, we have the following properties of θ_{iB}.

Lemma *2: (Properties of Parameter θ_{iB})* The following three properties always hold: i) θ_{iB} is decreasing in ϵ_{iB}; ii) θ_{iB} is increasing in α_{iB}; iii) $\theta_{iB} \le g_{iB}$ always holds.

Proof: The proof is same as that for Lemma 1. ■

Remark 1: (Perfectly Secure Situation) Eq. (23) can capture the situation that MU i's transmission to the mBS is perfectly secure. Specifically, if we consider that MU i's transmission to the mBS is perfectly secure, we can set $\epsilon_{ik} = 1$ (as there is no need to provide any secrecy-protection in the perfectly secure case). This consequently leads to that $\theta_{ik} = 0$ according to (24) and $x_{iB} = w_{iB} \log_2\left(1 + \frac{p_{iB} g_{iB}}{w_{iB} n_0}\right)$ which is the conventional channel capacity formula from MU i to the mBS.

Eq. (13) leads to the MU's effective secrecy-rate to mBS as:

$$r_{iB} = x_{iB}(1 - \epsilon_{iB}) = R_i^{\text{req}} - r_{ik}. \tag{25}$$

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

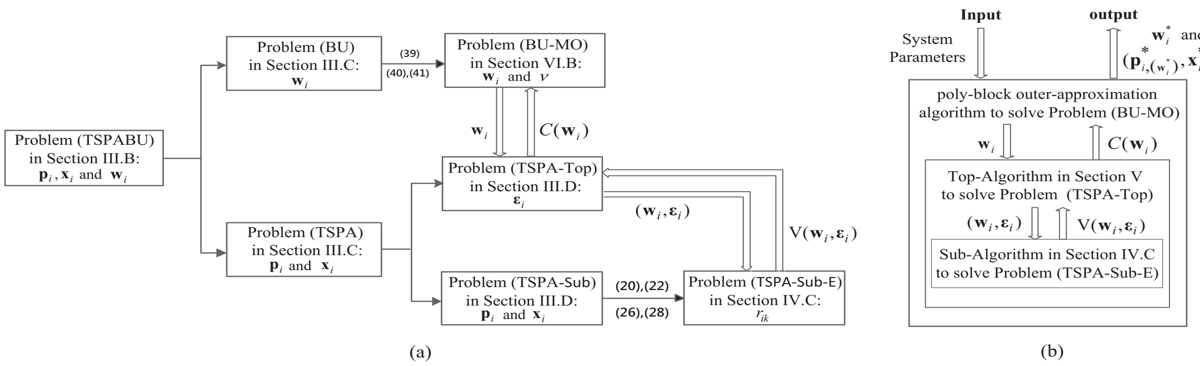


Fig. 2: Subfigure (a): Relationship among Problems (TSPABU), (BU), (TSPA-Top), and (TSPA-Sub). Subfigure (b): The layered structure of the proposed algorithm according to the relationship shown in Subfigure (a).

Further using (19) and (23), we can derive p_{ik} as follows:

$$p_{iB} = w_{iB} n_0 \frac{G(r_{ik}) - 1}{g_{iB} - \theta_{iB} G(r_{ik})}, \qquad (26)$$

where $G(r_{ik})$ is defined as

$$G(r_{ik}) = 2^{\frac{R_i^{\text{req}} - r_{ik}}{(1-\epsilon_{iB}) w_{iB}}}. \qquad (27)$$

Putting (26) into (8), we derive the lower-limit of r_{ik} as:

$$r_{ik} \geq r_{ik}^{\text{lower}} = \max \Bigg\{ 0, \; R_i^{\text{req}} - (1-\epsilon_{iB}) w_{iB} \log_2 \left(\frac{p_{iB}^{\max} g_{iB} + w_{iB} n_0}{p_{iB}^{\max} \theta_{iB} + w_{iB} n_0} \right) \Bigg\}, \qquad (28)$$

C. Equivalent Problem (TSPA-Sub) and Optimal Solution

Using (20), (22), (26), and (28), we equivalently transform Problem (TSPA-Sub) into the following one (the "E" denotes "Equivalence")

(TSPA-Sub-E): $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i) =$

$$\min \; w_{ik} n_0 \frac{F(r_{ik}) - 1}{g_{ik} - \theta_{ik} F(r_{ik})} + w_{iB} n_0 \frac{G(r_{ik}) - 1}{g_{iB} - \theta_{iB} G(r_{ik})}$$

$$\text{Subject to:} \quad r_{ik}^{\text{lower}} \leq r_{ik} \leq r_{ik}^{\text{upper}}, \qquad (29)$$

Variables: r_{ik}.

Problem (TSPA-Sub-E) only uses r_{ik} (i.e., MU i's effective secrecy-rate to sAP k) as the decision variable. Note that given $(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$, we can uniquely determine $(r_{ik}^{\text{lower}}, r_{ik}^{\text{upper}})$ according to (28) and (22).

Figure 2(a) shows the connection between Problem (TSPA-Sub) and Problem (TSPA-Sub-E). To solve Problem (TSPA-Sub-E), we identify the following property.

Proposition 1: *(Convexity)* Problem (TSPA-Sub-E) is a strictly convex optimization problem with respect to r_{ik}.

Proof: Let $\varphi(r_{ik})$ denote the first order derivative of the objective function of Problem (TSPA-Sub-E). We have

$$\varphi(r_{ik}) = w_{ik} n_0 \frac{(g_{ik} - \theta_{ik}) F'(r_{ik})}{(g_{ik} - \theta_{ik} F(r_{ik}))^2} + w_{iB} n_0 \frac{(g_{iB} - \theta_{iB}) G'(r_{ik})}{(g_{iB} - \theta_{iB} G(r_{ik}))^2}. \qquad (30)$$

where $F'(r_{ik})$ and $G'(r_{ik})$ are respectively given by:

$$F'(r_{ik}) = 2^{\frac{r_{ik}}{(1-\epsilon_{ik}) W_k}} \frac{\ln 2}{(1-\epsilon_{ik}) W_k}, \qquad (31)$$

$$G'(r_{ik}) = -2^{\frac{R_i^{\text{req}} - r_{ik}}{(1-\epsilon_{iB}) W_B}} \frac{\ln 2}{(1-\epsilon_{iB}) W_B}. \qquad (32)$$

Since both $F(r_{ik})$ in (21) and $F'(r_{ik})$ in (31) are positive and increasing in r_{ik}, the first item in the right-hand side of (30) (i.e., $w_{ik} n_0 \frac{(g_{ik} - \theta_{ik}) F'(r_{ik})}{(g_{ik} - \theta_{ik} F(r_{ik}))^2}$) is increasing in r_{ik}. Moreover, since $G(r_{ik})$ in (42) is positive and decreasing in r_{ik}, while $G'(r_{ik})$ in (32) is negative and increasing in r_{ik}, the second item in the right-hand side of (30) (i.e., $w_{iB} n_0 \frac{(g_{iB} - \theta_{iB}) G'(r_{ik})}{(g_{iB} - \theta_{iB} G(r_{ik}))^2}$) is increasing in r_{ik}. Therefore, $\varphi(r_{ik})$ is increasing in r_{ik}. ■

Proposition 1 enables us to use the Karush-Kuhn-Tucker (KKT) conditions, a set of sufficient conditions for achieving the optimum of convex optimization problems [36], to determine the optimal solution for Problem (TSPA-Sub-E). Specifically, we use $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)}$ to denote the optimal solution of Problem (TSPA-Sub-E), which depends on the given $(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$. According to the KKT conditions, $\varphi(r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)}) = 0$ is a sufficient condition for achieving the optimality of Problem (TSPA-Sub-E). Using this condition and the increasing property of function $\varphi(r_{ik})$, we propose the following **Sub-Algorithm**, which works in the manner of bisection search, to compute $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)}$ for MU i.

The key of Sub-Algorithm includes the following steps.

- Step 3: Given $(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$, MU i uses (28) and (22) to compute r_{ik}^{lower} and r_{ik}^{upper}, respectively.
- Steps 8 to 11: If $\varphi(r_{ik}^{\text{upper}}) \leq 0$, MU i directly sets $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)} = r_{ik}^{\text{upper}}$ since the objective function is decreasing. If $\varphi(r_{ik}^{\text{lower}}) \geq 0$, MU i directly sets $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)} = r_{ik}^{\text{lower}}$, since the objective function is increasing.
- Steps 12 to 26: If $\varphi(r_{ik}^{\text{lower}}) < 0 < \varphi(r_{ik}^{\text{upper}})$, then MU i executes a bisection search to determine $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)}$ such that $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)} = 0$ holds.

Because of the efficient operations of bisection search, Sub-Algorithm can compute $r^*_{ik,(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)}$ within $\log_2 \left(\frac{r_{ik}^{\text{upper}} - r_{ik}^{\text{lower}}}{\gamma} \right)$ rounds of iterations, and thus is computationally efficient. In Figure 2(b), we illustrate how Sub-Algorithm is used as a subroutine to compute $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$.

V. Solving Problem (TSPA-Top)

By using Sub-Algorithm to compute $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$, we then solve Problem (TSPA-Top) (recall that Figure 2(a) shows the connections between Problem (TSPA-Top) and Problem (TSPA-Sub)). The difficulty in solving Problem (TSPA-Top) is that we cannot derive $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ analytically. Fortunately, the feasible region $(0, \epsilon_{ik}^{\max}] \times (0, \epsilon_{iB}^{\max}]$ is usually small, since $\epsilon_{ik}^{\max}$ and $\epsilon_{iB}^{\max}$ are small. Thus, to solve Problem (TSPA-Top), we use a small step-size Δ to execute a two-dimensional line-search for $\boldsymbol{\epsilon}_i$

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

Sub-Algorithm: to compute $r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)}$ and $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$

Input: $(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$.
MU i sets a tolerance γ for computational error and sets flag $=1$.
MU i computes r_{ik}^{upper} and r_{ik}^{lower} according to (28) and (22).
if $r_{ik}^{\text{lower}} > r_{ik}^{\text{upper}}$ **then**
  MU i sets $V(\boldsymbol{\epsilon}_i)=\infty$.
else
  MU i computes $\varphi(r_{ik}^{\text{upper}})$ and $\varphi(r_{ik}^{\text{lower}})$ according to (30).
  if $\varphi(r_{ik}^{\text{upper}}) \leq 0$ **then**
    MU i sets $r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)} = r_{ik}^{\text{upper}}$ and flag $=0$.
  else if $\varphi(r_{ik}^{\text{lower}}) \geq 0$ **then**
    MU i sets $r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)} = r_{ik}^{\text{lower}}$ and flag $=0$.
  else
    MU i sets $\underline{r}_{ik} = r_{ik}^{\text{lower}}$ and $\overline{r}_{ik} = r_{ik}^{\text{upper}}$, respectively.
    while flag $=1$ **do**
      MU i sets $r_{ik}^{\text{cur}} = (\overline{r}_{ik}+\underline{r}_{ik})/2$.
      if $(\overline{r}_{ik}-\underline{r}_{ik}) \leq \gamma$ **then**
        MU i sets $r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)} = r_{ik}^{\text{cur}}$ and flag $=0$.
      else
        if $\varphi\left(r_{ik}^{\text{cur}}\right) > 0$ **then**
          MU i sets $\overline{r}_{ik} = r_{ik}^{\text{cur}}$.
        else
          MU i sets $\underline{r}_{ik} = r_{ik}^{\text{cur}}$.
        end if
      end if
    end while
  end if
  MU i computes $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i) = w_{ik}n_0\frac{F(r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)})-1}{g_{ik}-\theta_{ik}F(r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)})} + w_{iB}n_0\frac{G(r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)})-1}{g_{iB}-\theta_{iB}G(r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)})}$.
end if
Output: $r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)}$ and $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$.

within $(0,\epsilon_{ik}^{\max}] \times (0,\epsilon_{iB}^{\max}]$ (i.e., Steps 2 to 9 in the following proposed Top-Algorithm). Based on this rationale, we propose the following **Top-Algorithm** to find the optimal solution of Problem (TSPA-Top). Specifically, we use $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ to denote the optimal solution of Problem (TSPA-Top), since Problem (TSPA-Top) depends on the given $\boldsymbol{w}_i$.

The key of Top-Algorithm is Step 4. Given the currently enumerated $\boldsymbol{\epsilon}_i$, MU i uses Sub-Algorithm to compute $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$. Based on $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$, MU i updates its current best value (CBV) and current best solution (CBS) in Step 6. Given MU i's bandwidth usage $\boldsymbol{w}_i$, Top-Algorithm outputs the optimal secrecy level $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ for MU i (yet subject to its secrecy requirement) to minimize its total power consumption. Figure 2(b) shows how Top-Algorithm uses Sub-Algorithm.

We have the following proposition about the optimality of $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$, which depends on the chosen Δ.

Proposition 2: $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ is guaranteed to be the global optimum of Problem (TSPA-Top) as Δ approaches to zero.

Proof: As Δ approaches to zero, Top-Algorithm enumerates all feasible $(\epsilon_{ik},\epsilon_{iB})$. For each $(\epsilon_{ik},\epsilon_{iB})$, Sub-Algorithm yields the global optimum of Problem (TSPA-Sub-E) according to Proposition 1. Hence, by comparing all feasible $(\epsilon_{ik},\epsilon_{iB})$, $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ is guaranteed to be the global optimum of (TSPA-Top). ■

Fortunately, as $\epsilon_{ik}^{\max}$ and $\epsilon_{iB}^{\max}$ are small, we only need to set Δ moderately small such that Top-Algorithm can yield $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ sufficiently close to the global optimum while consuming a very short computational time. We will verify this in Section VIII.

Using $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ output by Top-Algorithm, we can compute the optimal solution of the original Problem (TSPA) as follows.

Proposition 3: *(Optimal Solution of Problem (TSPA)):* Using

Top-Algorithm: to compute $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ and $C(\boldsymbol{w}_i)$

1: MU i sets $CBV=\infty$ and **CBS** $=(\emptyset,\emptyset)$.
2: **for** $\epsilon_{ik}=\Delta:\Delta:\epsilon_{ik}^{\max}$ **do**
3:   **for** $\epsilon_{iB}=\Delta:\Delta:\epsilon_{iB}^{\max}$ **do**
4:     Given $(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$, MU i uses Sub-Algorithm to get $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$.
5:     **if** $V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i) < CBV$ **then**
6:       MU i updates **CBS** $=\boldsymbol{\epsilon}_i$, $CBV=V(\boldsymbol{w}_i,\boldsymbol{\epsilon}_i)$.
7:     **end if**
8:   **end for**
9: **end for**
10: **if** $CBV=\infty$ **then**
11:   **Output**: Problem (TSPA-Top) (or Problem (TSPA)) is infeasible.
12: **else**
13:   **Output**: $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)} = (\epsilon^*_{ik,(\boldsymbol{w}_i)},\epsilon^*_{iB,(\boldsymbol{w}_i)})$ = **CBS** and $C(\boldsymbol{w}_i)$ = **CBV** for Problem (TSPA-Top).
14: **end if**

$\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$ output by Top-Algorithm, MU i can compute the optimal traffic scheduling and power allocation for Problem (TSPA) as follows (the subscript $\boldsymbol{w}_i$ is included in all the following optimal solutions, since Problem (TSPA) is subject to the given $\boldsymbol{w}_i$)

$$x^*_{ik,(\boldsymbol{w}_i)} = \frac{r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})}}{(1-\epsilon^*_{ik,(\boldsymbol{w}_i)})}, \tag{33}$$

$$p^*_{ik,(\boldsymbol{w}_i)} = w_{ik}n_0\frac{F(r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})})-1}{g_{ik}-\theta_{ik}F(r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})})}, \tag{34}$$

$$x^*_{iB,(\boldsymbol{w}_i)} = \frac{R_i^{\text{req}} - r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})}}{(1-\epsilon^*_{iB,(\boldsymbol{w}_i)})}, \tag{35}$$

$$p^*_{iB,(\boldsymbol{w}_i)} = w_{iB}n_0\frac{G(r^*_{iB,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})})-1}{g_{iB}-\theta_{iB}G(r^*_{iB,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})})}. \tag{36}$$

Proof: Given $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$, MU i's optimal effective secrecy-rate $r^*_{ik,(\boldsymbol{w}_i,\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)})}$ can be obtained by using Sub-Algorithm. Therefore, we can derive MU i's optimal offloading-rate to sAP k in (33) according to (19), and its corresponding optimal transmit-power to sAP k in (34) according to (20). We can derive MU i's optimal rate to the mBS in (35) according to (25), and derive the optimal transmit-power to the mBS in (36) according to (26). ■

Until now, we complete solving Problem (TSPA). Using $\boldsymbol{\epsilon}^*_{i,(\boldsymbol{w}_i)}$, we can compute $C(\boldsymbol{w}_i)$ for each given $\boldsymbol{w}_i$ as

$$C(\boldsymbol{w}_i) = p^*_{ik,(\boldsymbol{w}_i)} + p^*_{iB,(\boldsymbol{w}_i)}. \tag{37}$$

VI. Solving Problem (BU)

Using $C(\boldsymbol{w}_i)$, we continue to solve Problem (BU) to determine MU i's optimal bandwidth usage in this section (Figure 2(a) shows the connection between Problem (BU) and Problem (TSPA)). The key difficulty in solving Problem (BU) is that we cannot analytically derive $C(\boldsymbol{w}_i)$. Thus, Problem (BU) does not have an analytical objective function, which prevents us from using gradient-based approach to solve it. Nevertheless, for each given $\boldsymbol{w}_i$, we can compute $C(\boldsymbol{w}_i)$ by using our Top-Algorithm, Proposition 3, and (37). Moreover, as we will show in this section, Problem (BU) can be equivalently transformed into a monotonic optimization problem [37] [38]. Using this important property and our Top-Algorithm to compute $C(\boldsymbol{w}_i)$, we can efficiently solve Problem (BU). The details are as follows.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

A. Introduction to Monotonic Optimization

We provide an introduction to the monotonic optimization theory, since we will use it to solve Problem (BU). Two important definitions are given as follows.

Definition 1: (Definition of Normal Set). A set $\mathcal{G} \subset \mathcal{R}_+^n$ is normal, if for any two points x and $x' \in \mathcal{R}_+^n$ with $x' \leq x$ and $x \in \mathcal{G}$, we always have $x' \in \mathcal{G}$.

Definition 2: (Definition of Reverse Normal Set). A set $\mathcal{H} \subset \mathcal{R}_+^n$ is a reversed normal set, if for two points x and $x' \in \mathcal{R}_+^n$ with $x' \geq x$ and $x \in \mathcal{H}$, we always have $x' \in \mathcal{H}$.

Based on the above definitions, a canonic form of the monotonic optimization problem is as follows.

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}), \text{ subject to: } \boldsymbol{x} \in \mathcal{G} \cap \mathcal{H}, \tag{38}$$

where $f(\boldsymbol{x}) : \mathcal{R}_+^n \rightarrow \mathcal{R}$ is an increasing function[1]. Set $\mathcal{G} \subset [\mathbf{0}, \boldsymbol{b}]$ is a normal set with nonempty interior, and set $\mathcal{H}$ is a reverse normal set on $[\mathbf{0}, \boldsymbol{b}]$, with vector $\boldsymbol{b}$ denoting a point in $\mathcal{R}_+^n$.

As shown in (38), a monotonic optimization problem involves maximizing a monotonic objective function subject to a feasible region constructed by the intersection of a normal set and a reversed normal set [37] [38]. The monotonicity inherent in the problem enables us to design a very efficient algorithm to solve the problem. Specifically, using the monotonicity of the constraints, one can iteratively construct a group of poly-blocks to approximate the feasible region with increasing precisions. Furthermore, because of the monotonicity of the objective function, the optimal solution is guaranteed to lie at one of the vertices of the constructed poly-blocks, as long as this vertex falls within (or very close to) the feasible region. Based on this idea, one can design the poly-block outer-approximation algorithm, which can efficiently search for the globally optimal solution of a monotonic optimization problem.

B. Monotonicity of Problem (BU)

To solve Problem (BU), we identify the following proposition.

Proposition 4: *(Monotonicity of $C(\boldsymbol{w}_i)$):* $C(\boldsymbol{w}_i)$ is nonincreasing in $\boldsymbol{w}_i$. Specifically, given two different $\boldsymbol{w}_i$ and $\boldsymbol{w}_i'$ with $\boldsymbol{w}_i \geq \boldsymbol{w}_i'$ (i.e., $w_{ik} \geq w_{ik}'$ and $w_{iB} \geq w_{iB}'$), there always exists $C(\boldsymbol{w}_i) \leq C(\boldsymbol{w}_i')$.

Proof: Please refer to Appendix I. ■

Proposition 4 matches the intuition very well. $C(\boldsymbol{w}_i)$ (i.e., the optimal objective value of Problem (TSPA)) represents MU i's minimum total power consumption to meet its traffic demand requirement and secrecy requirement. Thus, with a non-smaller bandwidth usage at the mBS and sAP k, MU i's minimum total power consumption at least will not increase. Nevertheless, the technical proof of Proposition 4 is very challenging as we present in Appendix I.

The objective function of Problem (BU) can be expressed as a difference between two parts as follows:

$$\begin{aligned} C(\boldsymbol{w}_i) + \mu_k w_{ik} + \mu_B w_{iB} = \\ -\left(-C(\boldsymbol{w}_i) - (\mu_k w_{ik} + \mu_B w_{iB})\right), \end{aligned} \tag{39}$$

where $-C(\boldsymbol{w}_i)$ is nondecreasing in $\boldsymbol{w}_i$ (i.e., Proposition 4), and $\mu_k w_{ik} + \mu_B w_{iB}$ is increasing$\boldsymbol{w}_i$. These two properties enable us to equivalently transform Problem (BU) in a monotonic optimization problem. We first introduce an auxiliary variable ν:

$$\mu_k w_{ik} + \mu_B w_{iB} + \nu = \mu_k w_{ik}^{\max} + \mu_B w_{iB}^{\max}, \tag{40}$$

where the feasible interval of ν is:

$$0 \leq \nu \leq \mu_k w_{ik}^{\max} + \mu_B w_{iB}^{\max} - (\mu_A w_{ik} + \mu_B w_{iB}). \tag{41}$$

Using ν, (39), (40), and (41), we equivalently transform Problem (BU) into the following one ("MO" means "Monotonic Optimization"):

$$\begin{aligned} &\underline{\textbf{(BU-MO)}} \quad \max -C(\boldsymbol{w}_i) + (\nu - \mu_k w_{ik}^{\max} - \mu_B w_{iB}^{\max}) \\ &\text{Subject to:} \quad (\boldsymbol{w}_i, \nu) \in (\mathcal{G} \cap \mathcal{H}) \\ &\text{Variables:} \quad \boldsymbol{w}_i \text{ and } \nu, \end{aligned}$$

where set $\mathcal{G}$ and set $\mathcal{H}$ are respectively given by:

$$\begin{aligned} \mathcal{G} = \Big\{ &(\boldsymbol{w}_i, \nu) | 0 \leq w_{ik} \leq w_{ik}^{\max}, 0 \leq w_{iB} \leq w_{iB}^{\max}, \\ &\nu + \mu_k w_{ik} + \mu_B w_{iB} \leq \mu_k w_{ik}^{\max} + \mu_B w_{iB}^{\max} \Big\}, \end{aligned} \tag{42}$$

$$\mathcal{H} = \left\{ (\boldsymbol{w}_i, \nu) | w_{ik}^{\min} \leq w_{ik}, w_{iB}^{\min} \leq w_{iB}, 0 \leq \nu \right\}. \tag{43}$$

Let $\boldsymbol{w}_i^*$ and ν^* denote the optimal solution of Problem (BU-MO). Then, $\boldsymbol{w}_i^*$ suffices to be the optimal solution of Problem (BU). To solve Problem (BU-MO), we identify the following important proposition.

Proposition 5: Problem (BU-MO) is a monotonic optimization problem with respect to $(\boldsymbol{w}_i, \nu)$.

Proof: It can be verified that $\mathcal{G}$ in (42) is a normal set according to Definition 1, and $\mathcal{H}$ in (43) is a reversed normal set according to Definition 2. Based on Proposition 4, the objective function of Problem (BU-MO) is increasing in $\boldsymbol{w}_i$ and ν. Thus, based on the monotonic optimization theory [37] [38], Problem (BU-MO) is a monotonic optimization problem. ■

Proposition 5 enables us to use the standard poly-block outer-approximation algorithm to solve Problem (BU-MO) and obtain the optimal solution $(\boldsymbol{w}_i^*, \nu^*)$. Due to the limited space, we skip the detailed descriptions about the operations of the polyblock outer-approximation algorithm (interested readers can refer to [37]). We emphasize that in executing the poly-block outer-approximation algorithm, for each evaluated $\boldsymbol{w}_i$, we need to use Top-Algorithm to obtain $C(\boldsymbol{w}_i)$. Figure 2(b) illustrates the whole algorithm to solve Problem (BU), which uses Top-Algorithm a subroutine to compute $C(\boldsymbol{w}_i)$ under a given $\boldsymbol{w}_i$.

Finally, after obtaining $\{\boldsymbol{w}_i^*\}$, we can compute the optimal solution of the original Problem (TSPABU) as follows. After obtaining $\boldsymbol{w}_i^*$, MU i uses Top-Algorithm to obtain $\boldsymbol{\epsilon}_{i,(\boldsymbol{w}_i^*)}^*$. Afterwards, MU i can compute its traffic scheduling $(x_{ik,(\boldsymbol{w}_i^*)}^*, x_{iB,(\boldsymbol{w}_i^*)}^*)$ and power allocation $(p_{ik,(\boldsymbol{w}_i^*)}^*, p_{iB,(\boldsymbol{w}_i^*)}^*)$ according to (33), (34), (35), and (36) in Proposition 3, which completes solving Problem (TSPABU).

VII. Extension to Multiple-MU Multiple-sAP Case

A. Multi-MU Multi-sAP Model

The optimal MU's traffic offloading solution for the single-MU single-sAP case provides us an important step to study more complicated scenarios. As a result, we consider the multi-MU multi-sAP case (as shown in Figure 1) and investigate how the

[1]Specifically, given two different $\boldsymbol{x}$ and $\boldsymbol{x}'$ with $x_k \geq x_k', \forall k$ and $x_j > x_j'$ for at at least one index j, there always exists $f(\boldsymbol{x}) < f(\boldsymbol{x}')$.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

MUs, by using the optimal offloading solutions with respect to different sAPs, properly select different sAPs to form the DC-pairs for traffic offloading. We consider a mBS, a group $\mathcal{K} = \{1, ..., K\}$ of sAPs, and a group $\mathcal{I} = \{1, ..., I\}$ of MUs under the mBS's coverage and the sAPs' coverage. Each MU i can form the DC with the mBS and a selected sAP to offload its traffic demand. Meanwhile, we consider that there exists a malicious eavesdropper which overhears the MUs' traffic to the mBS and sAPs. Thus, it is interesting to study how different MUs properly select the sAPs to form DC-pairs for traffic offloading.

To model the MUs' selections, we use $s_{ik} = 1$ to denote that MU i chooses sAP j to offload data (in other words, to form a DC with the mBS and sAP k), and $s_{ik} = 0$ otherwise. Moreover, supposing that $s_{ik} = 1$, MU i will set its bandwidth usage, traffic scheduling, and power allocation to sAP k and mBS according to the optimal offloading solution of Problem (TSPABU), i.e., the bandwidth usage (w_{ik}^*, w_{iB}^*), the traffic scheduling $(x_{ik,(\boldsymbol{w}_i^*)}^*, x_{iB,(\boldsymbol{w}_i^*)}^*)$, and the power allocation $(p_{ik,(\boldsymbol{w}_i^*)}^*, p_{iB,(\boldsymbol{w}_i^*)}^*)$. Notice that using the optimal offloading solution matches MU i's interest, since it minimizes MU i's overall radio resource usage while satisfying both MU i's traffic demand and secrecy-requirement. As summarized in Figure 2(b), we have proposed an efficient algorithm to successfully compute (w_{ik}^*, w_{iB}^*), $(x_{ik,(\boldsymbol{w}_i^*)}^*, x_{iB,(\boldsymbol{w}_i^*)}^*)$, and $(p_{ik,(\boldsymbol{w}_i^*)}^*, p_{iB,(\boldsymbol{w}_i^*)}^*)$ for $s_{ik} = 1$ (i.e., MU i chooses sAP k to form a DC-pair (with the mBS) for traffic offloading).

In this section, to differ the optimal solutions for different MU-sAP pairs, we denote MU i's optimal offloading solution for $s_{ik} = 1$ as follows:

$$\text{Bandwidth Usage: } (w_{ik}^*, w_{iB(k)}^*) = (w_{ik}^*, w_{iB}^*)$$
$$\text{Traffic Scheduling:} (x_{ik}^*, x_{iB(k)}^*) = (x_{ik,(\boldsymbol{w}_i^*)}^*, x_{iB,(\boldsymbol{w}_i^*)}^*)$$
$$\text{Power Allocation: } (p_{ik}^*, p_{iB(k)}^*) = (p_{ik,(\boldsymbol{w}_i^*)}^*, p_{iB,(\boldsymbol{w}_i^*)}^*)$$

In addition to choosing a sAP to form a DC-pair for traffic offloading, we also allow the possibility that MU i can choose to transmit to the mBS alone (i.e., without offloading any data). We use $s_{i0} = 1$ to denote this case, and $s_{i0} = 0$ otherwise. Correspondingly, we denote MU i's optimal offloading solution when $s_{i0} = 1$ by $(w_{i0}^*, w_{iB(0)}^*)$, $(x_{i0}^*, x_{iB(0)}^*)$, and $(p_{i0}^*, p_{iB(0)}^*)$[2]. Notice that there always exists $w_{i0}^* = 0$, $x_{i0}^* = 0$, and $p_{i0}^* = 0$, no traffic or radio resource involved at the sAP-side.

We consider that the mBS and sAPs use frequency division to serve different MUs. Thus, there is no interference among different MUs which offload data to same sAP, and no interference among the MUs when transmitting to the mBS. Nevertheless, we consider that the mBS has the maximum bandwidth limit denoted by $W_B^{\max}$, and each sAP k has the maximum bandwidth limit denoted by $W_k^{\max}$ for serving the MUs.

B. Problem Formulation

We next formulate the MUs' optimal offloading-selection problem. First, we consider that each MU i has two radio-interfaces and can either choose one sAP to form the DC-pair (together the mBS) or choose to transmit to the mBS only, i.e.,

$$s_{i0} + \sum_{k\in\mathcal{K}} s_{ik} \leq 1, \forall i \in \mathcal{I}. \tag{44}$$

From MU i's perspective, MU i can set $s_{ik} = 1$ (i.e., choosing sAP k and mBS to form DC), only if MU i's required optimal power allocations p_{ik}^* and $p_{iB(k)}^*$ no larger than the respective limits $p_{ik}^{\max}$ and $p_{iB}^{\max}$. This leads to the following constraint[3]:

$$s_{ik} \leq \mathbb{I}(p_{ik}^* \leq p_{ik}^{\max}) \cdot \mathbb{I}(p_{iB(k)}^* \leq p_{iB}^{\max}), \forall i \in \mathcal{I}, k \in \mathcal{K}. \tag{45}$$

Similarly, MU i can set $s_{i0} = 1$ (i.e., only sending traffic to the mBS), only if MU i's required optimal power allocation $p_{iB(0)}^*$ no larger than the limit $p_{iB}^{\max}$. This leads to

$$s_{i0} \leq \mathbb{I}(p_{iB(0)}^* \leq p_{iB}^{\max}), \forall i \in \mathcal{I}. \tag{46}$$

Second, from sAP k's perspective, all MUs' bandwidth usage at sAP k cannot exceed sAP k's bandwidth capacity $W_k^{\max}$, i.e.,

$$\sum_{i\in\mathcal{I}} s_{ik} w_{ik}^* \leq W_k^{\max}, \forall k \in \mathcal{K}. \tag{47}$$

Moreover, we consider that each sAP k has a limited backhaul capacity denoted by $C_k^{\max}$. Thus, all MUs' offloaded data to sAP k cannot exceed $C_k^{\max}$, i.e.,

$$\sum_{i\in\mathcal{I}} s_{ik} x_{ik}^* \leq C_k^{\max}, k \in \mathcal{K}. \tag{48}$$

Finally, from the mBS's perspective, all MUs' total bandwidth usage at the mBS cannot exceed $W_B^{\max}$, i.e.,

$$\sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}} s_{ik} w_{iB(k)}^* + \sum_{i\in\mathcal{I}} s_{i0} w_{iB(0)}^* \leq W_B^{\max}. \tag{49}$$

In (49), the first part denotes all MUs' bandwidth usage at the mBS when forming the DC-pairs, and the second part denotes all MUs' bandwidth usage at the mBS when the MUs choose to send traffic the mBS only.

For notational simplicity, we define an auxiliary group $\widetilde{\mathcal{K}} = \mathcal{K} \cup \{0\}$, and moreover, we set $W_0^{\max}$ and $C_0^{\max}$ as arbitrary positive values[4]. We formulate the following MUs' Offloading-Selection Problem.

$$\underline{\textbf{(OSP)}} \quad \max \sum_{i\in\mathcal{I}} \sum_{k\in\widetilde{\mathcal{K}}} s_{ik} R_i^{\text{req}}$$
$$\text{Subject to: } \sum_{k\in\widetilde{\mathcal{K}}} s_{ik} \leq 1, \forall i \in \mathcal{I}, \tag{50}$$
$$s_{ik} \leq \mathbb{I}(p_{ik}^* \leq p_{ik}^{\max}) \cdot \mathbb{I}(p_{iB(k)}^* \leq p_{iB}^{\max}), \forall i \in \mathcal{I}, k \in \widetilde{\mathcal{K}}, \tag{51}$$
$$\sum_{i\in\mathcal{I}} s_{ij} w_{ik}^* \leq W_k^{\max}, \forall k \in \widetilde{\mathcal{K}}, \tag{52}$$
$$\sum_{i\in\mathcal{I}} s_{ik} x_{ik}^* \leq C_k^{\max}, k \in \widetilde{\mathcal{K}}, \tag{53}$$
$$\sum_{i\in\mathcal{I}} \sum_{k\in\widetilde{\mathcal{K}}} s_{ik} w_{iB(k)}^* \leq W_B^{\max}, \tag{54}$$
$$\text{Variables: } s_{ik} = \{0, 1\}, \forall i \in \mathcal{I}, k \in \widetilde{\mathcal{K}}.$$

Constraints (50)-(54) are equivalent to (44)-(49), due to $\widetilde{\mathcal{K}} = \mathcal{K} \cup \{0\}$. The objective is to maximize the total MUs' traffic demands which are successfully served. Here, the successful serve means that if MU i is selected to be served (i.e., $s_{ik} = 1$ for one $k \in \widetilde{\mathcal{K}}$), we guarantee to satisfy MU i's traffic demand and secrecy requirement, because of using the optimal solution of Problem

[2]Notice that by fixing $r_{ik} = 0$ in Problem (TSPA-Sub-E), we can easily compute $(w_{i0}^*, w_{iB(0)}^*)$, $(x_{i0}^*, x_{iB(0)}^*)$, and $(p_{i0}^*, p_{iB(0)}^*)$ for each MU i with the similar algorithm as we have described in Sections IV-VI.

[3]Indicator function $\mathbb{I}(x) = 1$ if condition x is true, and $\mathbb{I}(x) = 0$ otherwise.

[4]In $\widetilde{\mathcal{K}}$, sAP 0 is a virtual sAP and denotes the mBS, with $W_0^{\max}$ and $C_0^{\max}$ being arbitrary positive values.

(TSPABU). We emphasize that in Problem (OSP), the values of $(w^*_{ik}, w^*_{iB(k)})$, $(x^*_{ik}, x^*_{iB(k)})$, and $(p^*_{ik}, p^*_{iB(k)})$ are known after we solve Problem (TSPABU) for the pair of MU i and sAP k. In Problem (OSP), by properly adjusting the offloading-selection $\{s_{ik}\}_{i\in\mathcal{I},k\in\mathcal{K}}$, we can guarantee constraints (52)-(54).

Problem (OSP) corresponds to a complicated multiple-resource generalized assignment problem (MRGAP) [41]. The objective of Problem (OSP) is to properly assign the MUs in $\mathcal{I}$ (i.e., tasks) to different sAPs in $\mathcal{K}$ (i.e., agents). The associated costs with each pair of MU i and sAP k are given by the optimal bandwidth usage $(w^*_{ik}, w^*_{iB(k)})$, traffic scheduling $(x^*_{ik}, x^*_{iB(k)})$, and power allocation $(p^*_{ik}, p^*_{iB(k)})$ for Problem (TSPABU) with respect to the specific pair of MU i and sAP k. In Problem (OSP), we are subject to four resource-budget constraints, namely, i) the power-budget constraint (51), ii) sAP k's bandwidth budget (52), iii) sAP k's backhaul capacity budget (53), and iv) the mBS's bandwidth budget (54). Moreover, in Problem (OSP), each agent (sAP k) can serve multiple tasks (i.e., MUs), and the connection limit (50) ensures that each task (MU i) can only be served at most once. Such complicated constraints together yield Problem (OSP) a typical NP-hard problem [42]. According to [41] [42], even GAP allowing an agent to serve multiple different tasks (with each task performed exactly once) is a NP-hard problem.

C. Solving Problem (OSP)

Viewing the NP-hardness of Problem (OSP), we propose a low-complexity heuristic algorithm (i.e., **LiMo-Algorithm**) to solve it. The key idea is as follows. Given the currently available MUs in group $\mathcal{I}^{\text{cur}}$, the mBS's backhaul capacity w_B^{cur}, the available sAPs' bandwidth capacities, and the backhaul capacities $\{w_k^{\text{cur}}, c_k^{\text{cur}}\}_{k\in\widetilde{\mathcal{K}}}$, we solve Problem (OSP-Li), a Linear-relaxation of Problem (OSP). Let $\{\hat{s}_{ik}\}_{i\in\mathcal{I}^{\text{cur}},k\in\widetilde{\mathcal{K}}}$ denote the optimal solution of Problem (OSP-Li). Intuitively, the larger $\hat{s}_{ik}$ (e.g., closer to one), the more beneficial to myopically fix the choice that MU i chooses sAP k to offload traffic via DC. This rationale yields the following While-loop (Steps 5-20) in LiMo-Algorithm in which we fix one MU-sAP pair in each round.

- Step 6: Given the available group $\mathcal{I}^{\text{cur}}$ of MUs, w_B^{cur}, and $\{w_k^{\text{cur}}, c_k^{\text{cur}}\}_{k\in\widetilde{\mathcal{K}}}$, we solve Problem (OSP-Li) shown below which corresponds to the linear relaxation of Problem (OSP) under the current $\mathcal{I}^{\text{cur}}$, w_B^{cur}, and $\{w_k^{\text{cur}}, c_k^{\text{cur}}\}_{k\in\widetilde{\mathcal{K}}}$. We use $\{\hat{s}_{ik}\}_{i\in\mathcal{I}^{\text{cur}},k\in\widetilde{\mathcal{K}}}$ to denote the optimal solution of Problem (OSP-Li).

$$\underline{\textbf{(OSP-Li)}} \quad \max \sum_{i\in\mathcal{I}^{\text{cur}}}\sum_{k\in\widetilde{\mathcal{K}}} s_{ik}R_i^{\text{req}}$$

$$\text{Subject to:} \quad \sum_{k\in\widetilde{\mathcal{K}}} s_{ik} \leq 1, \forall i\in\mathcal{I}^{\text{cur}},$$

$$s_{ik} \leq \mathbb{I}(p^*_{ik} \leq p_{ik}^{\max})\cdot\mathbb{I}(p^*_{iB(k)} \leq p_{iB}^{\max}), \forall i\in\mathcal{I}^{\text{cur}}, k\in\widetilde{\mathcal{K}},$$

$$\sum_{i\in\mathcal{I}^{\text{cur}}} s_{ij}w^*_{ik} \leq w_k^{\text{cur}}, \forall k\in\widetilde{\mathcal{K}},$$

$$\sum_{i\in\mathcal{I}^{\text{cur}}} s_{ik}x^*_{ik} \leq c_k^{\text{cur}}, \forall k\in\widetilde{\mathcal{K}},$$

$$\sum_{i\in\mathcal{I}^{\text{cur}}}\sum_{k\in\widetilde{\mathcal{K}}} s_{ik}w^*_{iB(k)} \leq w_B^{\text{cur}},$$

$$\text{Variables:} \quad 0 \leq s_{ik} \leq 1, \forall i\in\mathcal{I}^{\text{cur}}, k\in\widetilde{\mathcal{K}}.$$

- Step 7: Based on $\{\hat{s}_{ik}\}_{i\in\mathcal{I}^{\text{cur}},k\in\widetilde{\mathcal{K}}}$, we set group $\mathcal{Z}$ to include the largest element(s) which meet the conditions $\frac{w^*_{ik}}{w_k^{\text{cur}}} \leq 1$, $\frac{w^*_{iB(k)}}{w_B^{\text{cur}}} \leq 1$, $\frac{x^*_{ik}}{c_k^{\text{cur}}} \leq 1$ (notice that meeting these conditions mean that it is feasible to set $\hat{s}_{ik} = 1$ directly). Group $\mathcal{Z}$ includes the candidate MU-sAP pairs that potentially can help increase the objective function value.
- Step 12: If there exist several candidate MU-sAP pairs in group $\mathcal{Z}$, we choose the pair, denoted by $(i^{\text{opt}}, k^{\text{opt}})$, which leads to the largest $R_{\hat{i}}^{\text{req}}/(\frac{w^*_{\hat{i}\hat{k}}}{w_{\hat{k}}^{\text{cur}}} + \frac{w^*_{\hat{i}B(\hat{k})}}{w_B^{\text{cur}}} + \frac{x^*_{\hat{i}\hat{k}}}{c_{\hat{k}}^{\text{cur}}})$. This pair $(i^{\text{opt}}, k^{\text{opt}})$ can be considered as the most beneficial one to increase the objective value but using the least resources.
- Step 17: We set $s^*_{i^{\text{opt}},k^{\text{opt}}} = 1$, namely, MU i^{opt} determines to use sAP k^{opt} to form a DC-pair (i.e., $k^{\text{opt}} \geq 1$) or only transmit to the mBS (i.e., $k^{\text{opt}} = 0$). Accordingly, we update I^{cur}, $w_{k^{\text{opt}}}^{\text{cur}}$, $c_{k^{\text{opt}}}^{\text{cur}}$, and w_B^{cur}.

LiMo-Algorithm: to solve Problem (OSP) and find $\{s^*_{ik}\}_{i\in\mathcal{I},k\in\widetilde{\mathcal{K}}}$

1: Input: $(w^*_{ik}, w^*_{iB(k)})$, $(x^*_{ik}, x^*_{iB(k)})$, $(p^*_{ik}, p^*_{iB(k)})$ for each MU $i\in\mathcal{I}$ and each sAP $k\in\widetilde{\mathcal{K}}$.
2: Initialize $s^*_{ik} = 0, \forall i\in\mathcal{I}, k\in\widetilde{\mathcal{K}}$.
3: Set $w_k^{\text{cur}} = W_k^{\max}$ and $c_k^{\text{cur}} = C_k^{\max}$ for each sAP $k\in\widetilde{\mathcal{K}}$. Set $w_B^{\text{cur}} = W_B^{\max}$.
4: Set $\mathcal{I}^{\text{cur}} = \mathcal{I}$, and set $\{\Omega_k^{\text{opt}}\}_{k\in\widetilde{\mathcal{K}}} = \emptyset$ for each sAP $k\in\widetilde{\mathcal{K}}$.
5: **while** $\mathcal{I}^{\text{cur}} \neq \emptyset$ **do**
6: Solve Problem (OSP-Li) and denote the optimal solution by $\{\hat{s}_{ik}\}_{i\in\mathcal{I}^{\text{cur}},k\in\widetilde{\mathcal{K}}}$.
7: Set $\mathcal{Z} = \Big\{(\hat{i},\hat{k}) = \arg\max_{i\in\mathcal{I}^{\text{cur}},k\in\widetilde{\mathcal{K}}}\{\hat{s}_{ik}\} \big| \frac{w^*_{ik}}{w_k^{\text{cur}}} \leq 1, \frac{w^*_{iB(k)}}{w_B^{\text{cur}}} \leq 1, \frac{x^*_{ik}}{c_k^{\text{cur}}} \leq 1, \forall i\in\mathcal{I}^{\text{cur}}, k\in\widetilde{\mathcal{K}}\Big\}$.
8: **if** $\mathcal{Z} = \emptyset$ **then**
9: Break the While-Loop and go to Step 21.
10: **else**
11: **if** $|\mathcal{Z}| > 1$ **then**
12: $(i^{\text{opt}}, k^{\text{opt}})$=$\arg\max_{(\hat{i},\hat{k})\in\mathcal{Z}}\big\{R_{\hat{i}}^{\text{req}}/(\frac{w^*_{\hat{i}\hat{k}}}{w_{\hat{k}}^{\text{cur}}} + \frac{w^*_{\hat{i}B(\hat{k})}}{w_B^{\text{cur}}} + \frac{x^*_{\hat{i}\hat{k}}}{c_{\hat{k}}^{\text{cur}}})\big\}$.
13: **else**
14: $(i^{\text{opt}}, k^{\text{opt}}) = (\hat{i}, \hat{k})$.
15: **end if**
16: **end if**
17: Set $s^*_{i^{\text{opt}},k^{\text{opt}}} = 1$.
18: Update $I^{\text{cur}} = I^{\text{cur}}\backslash\{i^{\text{opt}}\}$.
19: Update $w_{k^{\text{opt}}}^{\text{cur}} = w_{k^{\text{opt}}}^{\text{cur}} - w^*_{i^{\text{opt}}k^{\text{opt}}}$, $c_{k^{\text{opt}}}^{\text{cur}} = c_{k^{\text{opt}}}^{\text{cur}} - x^*_{i^{\text{opt}}k^{\text{opt}}}$, and $w_B^{\text{cur}} = w_B^{\text{cur}} - w^*_{i^{\text{opt}}B(k^{\text{opt}})}$.
20: **end while**
21: **Output:** $\{s^*_{ik}\}_{\forall i\in\mathcal{I},k\in\widetilde{\mathcal{K}}}$.

LiMo-Algorithm terminates when the currently available MU-group $\mathcal{I}^{\text{cur}} = \emptyset$ or when we cannot find any feasible MU-sAP pair to update the optimal solution $\{s^*_{ik}\}_{i\in\mathcal{I},k\in\widetilde{\mathcal{K}}}$. The time complexity of LiMo-Algorithm can be analyzed as follows. The key complexity of LiMo-Algorithm stems from solving the linear programming problem (OSP-Li) in each round of iteration (Step 6 to step 19). Specifically, in the q-th iteration, the number of variables in Problem (OSP-Li) is equal to $(I+1-q)(K+1)$, which thus requires a time complexity of $O\big(((I+1-q)(K+1))^{3.5}\big)$ supposing that the interior point method is used to solve Problem (OSP-Li) [43]. Therefore, by taking into account that LiMo-Algorithm requires no more than I round of iterations, i.e., $q \leq I$, the overall time complexity of LiMo-Algorithm is equal to $O\big(\frac{1}{3}(K+1)^{3.5}I^{4.5}\big)$.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

TABLE I: Top-Algorithm to solve Problem (TSPA) ($w_{ik} = 20$MHZ, $w_{iB} = 5$MHZ, $\epsilon_{ik}^{\max} = \epsilon_{iB}^{\max} = 0.4$)

	$R^{\rm req} = 11$Mbps	$R^{\rm req} = 12$Mbps	$R^{\rm req} = 13$Mbps	$R^{\rm req} = 14$Mbps	$R^{\rm req} = 15$Mbps	Ave. Diff.
$\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/1000$	$3.7448 * 10^{-4}$, 171s	$4.4254 * 10^{-4}$, 177s	$5.2224 * 10^{-4}$, 185s	$6.1660 * 10^{-4}$, 202s	$7.2976 * 10^{-4}$, 206s	N/A
$\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/50$	$3.7450 * 10^{-4}$, 0.70s	$4.4257 * 10^{-4}$, 0.72s	$5.2224 * 10^{-4}$, 0.73s	$6.1662 * 10^{-4}$, 0.80s	$7.2985 * 10^{-4}$, 0.89	0.003%
$\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/20$	$3.7470 * 10^{-4}$, 0.10s	$4.4263 * 10^{-4}$, 0.11s	$5.2224 * 10^{-4}$, 0.12s	$6.1680 * 10^{-4}$, 0.16s	$7.3014 * 10^{-4}$, 0.17s	0.032%
$\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/10$	$3.7497 * 10^{-4}$, 0.05s	$4.4293 * 10^{-4}$, 0.05s	$5.2244 * 10^{-4}$, 0.06s	$6.1682 * 10^{-4}$, 0.06s	$7.3039 * 10^{-4}$, 0.06s	0.076%

VIII. Numerical Results

A. Single-MU Single-sAP Case

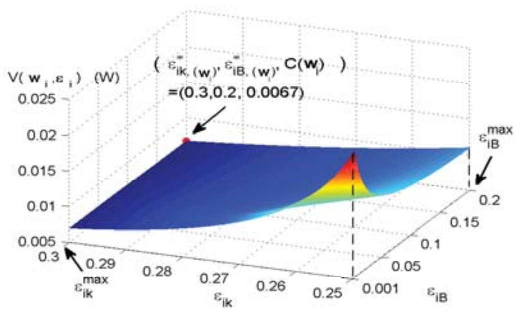
We setup the network scenario for the single-MU single-sAP case as follows.

(Network topology and channel gains): We consider a scenario where the mBS is located at the origin $(0\text{m}, 0\text{m})$, the sAP is located at $(250\text{m}, 0\text{m})$, and MU i is located at $(220\text{m}, 0\text{m})$. We set the channel power gain g_{iB} from MU i to the mBS according to the path-loss model, namely, $g_{iB} = \lambda d_{iB}^{-\kappa}$, in which d_{iB} denotes the distance between MU i and the mBS, κ denotes the scaling-parameter (we set $\kappa = 2.5$), and λ follows an exponential distribution with unit mean for capturing the impact of channel fading. The channel gain g_{ik} from MU i to sAP k is generated similarly. With this scheme, the randomly generated channel power gains are $g_{iB} = 4.14 \times 10^{-7}$ and $g_{ik} = 7.86 \times 10^{-5}$. We set $\alpha_{ik} = 2 \times 10^{-5} > \alpha_{iB} = 1 \times 10^{-7}$ to simulate that the eavesdropper hides around the sAP and thus has a much stronger capability of overhearing the MU's transmission to the sAP.

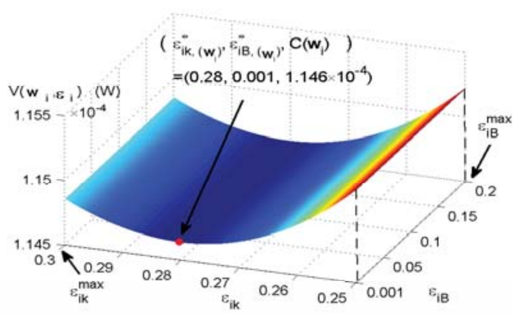
(System resources): We set $w_{iB}^{\min} = 0.01$MHz and $w_{iB}^{\max} = 5$MHz (close to a standard WCDMA channel) regarding MU i's bandwidth usage limits at the mBS. We set $w_{ik}^{\min} = 0.01$MHz and $w_{ik}^{\max} = 20$MHz (i.e., an IEEE 802.11a/b/g channel) regarding MU i's bandwidth usage limits at sAP k. We set MU i's $p_{iB}^{\max} = 0.3$W and $p_{ik}^{\max} = 0.25$W [39], and $n_0 = 10^{-15}$W/Hz.

1) Accuracy of Top-Algorithm and Sub-Algorithm to solve Problem (TSPA-Top) Table I validates the accuracy of Top-Algorithm (together with Sub-Algorithm) to solve Problem (TSPA). As an illustrative example, we set $w_{ik} = 20$MHz, $w_{iB} = 5$MHz, and $\epsilon_{ik}^{\max} = \epsilon_{iB}^{\max} = 0.4$. We first use an extremely small $\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/1000$ in Top-Algorithm to solve Problem (TSPA) and obtain the global optimum of Problem (TSPA) as a benchmark (the 2nd row in Table I). Next, we vary $\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/50$, $\max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/20$, and $\max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/10$ in Top-Algorithm and show the corresponding solution (i.e., the 3rd to 5th rows in Table I). Specifically, in each cell of Table I, the first number denotes the optimal objective value of Problem (TSPA), i.e., MU i's minimum total power consumption $C(\boldsymbol{w}_i)$ (given $\boldsymbol{w}_i$), and the second number denotes the corresponding computational time[5]. The results match with Proposition 2, and in particular, the results indicate that that using a moderately small $\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/20$ in Top-Algorithm suffices to achieve the solutions extremely close to the global optimum (with the average relative difference no greater than 0.05% compared to the benchmark solution), while consuming a very short computational time. In the rest of the paper, we thus fix $\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/20$ to strike a good balance between the accuracy and computational time.

Figure 3 shows the importance of properly selecting $(\epsilon_{ik}, \epsilon_{iB})$ to solve Problem (TSPA-Top). We enumerate $(\epsilon_{ik}, \epsilon_{iB})$ and use the 3-D plots to illustrate the consequently obtained $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$. Recall that $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ is obtained by using Sub-Algorithm to solve Problem (TSAP-Sub), and the optimal objective value of Problem (TSPA-Top) $C(\boldsymbol{w}_i) = \min_{\boldsymbol{\epsilon}_i} V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$. The comparison between Figures 3(a) and 3(b) shows the following indication. When the traffic demand is large (e.g., $R_i^{\rm req} = 15$Mbps in Figure 3(a)), MU i will choose to suffer the largest secrecy-outage probabilities $\epsilon_{ik}^{\max}$ and $\epsilon_{iB}^{\max}$, since this choice reduces MU i's power consumption (but with a sacrifice of MU i's effective secrecy-rate). However, when $R_i^{\rm req}$ is small in Figure 3(b), MU i will choose $(\epsilon_{ik,(\boldsymbol{w}_i)}^*, \epsilon_{iB,(\boldsymbol{w}_i)}^*)$ strictly lower than $(\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max})$, since this choice improves MU i's effective secrecy-rate (in spite of increasing MU i's required transmit-power). Eventually, MU i's total power consumption is reduced.



(a) $R_i^{\rm req} = 25$Mbps



(b) $R_i^{\rm req} = 5$Mbps

Fig. 3: Illustration of the importance of properly choosing $(\epsilon_{ik}, \epsilon_{iB})$ to solve Problem (TSPA-Top) with $\epsilon_{ik}^{\max} = 0.3$ and $\epsilon_{iB}^{\max} = 0.2$. Subplot (a): the case of the optimum $(\epsilon_{ik,(\boldsymbol{w}_i)}^*, \epsilon_{iB,(\boldsymbol{w}_i)}^*) = (\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max})$; Subplot (b): the case of the optimum $(\epsilon_{ik,(\boldsymbol{w}_i)}^*, \epsilon_{iB,(\boldsymbol{w}_i)}^*) < (\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max})$;

2) Optimal Offloading Solution for Problem (TSPABU) Figure 4 shows the accuracy and efficiency of using the poly-block outer-approximation algorithm (that uses Top-Algorithm as a subroutine) to solve Problem (BU). Specifically, the left-subplot of Figure 4 plots MU i's minimum overall cost obtained by using our proposed algorithm (i.e., the whole one shown in Figure 2(b)). To validate the accuracy, we also use LINGO[6] to solve Problem (TSPABU) and obtain MU i's global minimum

[5]We perform the numerical simulations on a PC with Intel(R) Core(TM) Core(TM) i7-4610 CPU @ 3.00GHz.

[6]LINGO is a an optimization software to solve complicated optimization problems [44]. LINGO provides an integrated package that can solve linear, convex, nonconvex, second order cone, and integer optimization models and etc.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

cost as a benchmark. Since Problem (TSPABU) is a nonconvex optimization problem, we need to use LINGO's global-solver to solve Problem (TSPABU). LINGO's global-solver converts the original nonconvex problem into several subproblems, and then uses the branch-and-bound technique to exhaustively search over these subproblems for the global solution. The downside of using LINGO's global-solver is that it might consume a long computational time. As shown in the left-subplot of Figure 4, our proposed algorithm achieves the results extremely close to those obtained by LINGO's global-solver in all tested cases, and the right-subplot of Figure 4 shows that our algorithm significantly reduces the computational time in comparison with LINGO. These results validate our analysis regarding the hidden monotonic property of Problem (BU) (in Proposition 5), and the effectiveness and efficiency of our proposed whole algorithm to compute the optimal offloading solution of Problem (TSPABU).

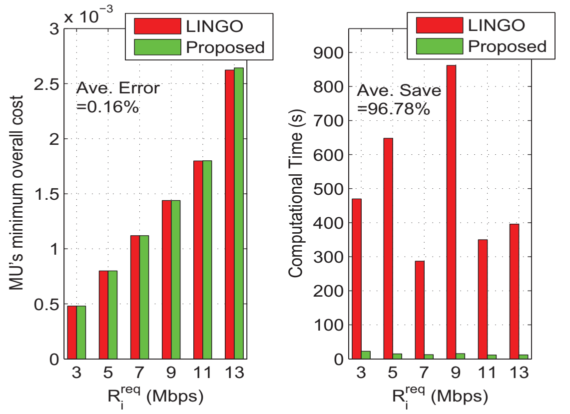


Fig. 4: Performance of the proposed algorithm to solve Problem (BU-MO). We set $\epsilon_{ik}^{\max} = 0.1$, $\epsilon_{iB}^{\max} = 0.15$, $w_{ik}^{\max} = 20$MHz, $w_{iB}^{\max} = 5$MHz, and $w_{ik}^{\min} = w_{iB}^{\min} = 0.01$MHz. We use $\Delta = \max\{\epsilon_{ik}^{\max}, \epsilon_{iB}^{\max}\}/20$.

Figure 5 further illustrates the optimal offloading solution under MU i's different traffic demands. Subfigure 5(a) plots MU i's optimal traffic scheduling to sAP k and the mBS. Subfigure 5(b) plots MU i's optimal bandwidth usage at sAP k and the mBS. Subfigure 5(c) plots MU i's optimal power allocation. We explain the results in Figure 5 as follows. When MU i's traffic demand R_i^{req} is small (i.e., $R_i^{\text{req}} = 3, 6, 9,$ and 12Mbps), it is optimal for MU i to offload its entire demand to sAP k. This is because that MU i is geographically close to sAP k, and it is beneficial for MU i to offload as much traffic as possible to sAP k for the perspective of saving MU i's energy consumption. In addition, as MU i's traffic demand R_i^{req} increases, both MU i's bandwidth usage and transmit-power to sAP k gradually increase. However, when MU i's traffic demand is large (i.e., $R_i^{\text{req}} = 15, 18$Mbps), it is optimal for MU i to properly offload part of its traffic demand to sAP k and send the remaining part to the mBS. This is because when MU i's traffic demand is large (e.g., $R_i^{\text{req}} = 15$MHz) and MU i's bandwidth usage at sAP k reaches the upper-bound $W_{iA}^{\max}$, MU i needs to consume a large transmit-power for offloading at a large rate to sAP k. Thus, it is beneficial for MU i to properly send part of its traffic demand to the mBS such that it achieves the best balance between the resource consumption (including the bandwidth usage and power consumption) at sAP k and at the mBS.

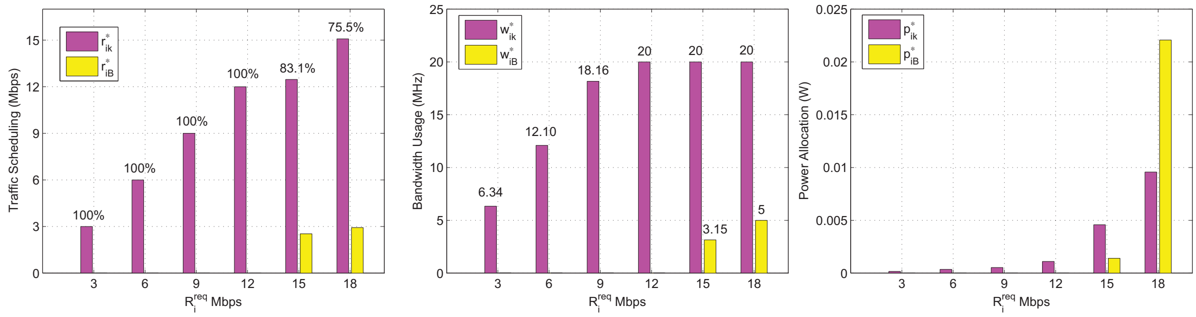
3) Influence of MU's secrecy-requirement on the MU's offloading solution Figure 6 shows the influence of MU i's secrecy requirement on the optimal offloading solution. We vary $\epsilon_{ik}^{\max}$ from 0.015 to 0.055 (while fixing $\epsilon_{iB}^{\max} = 0.1$) and plot the corresponding MU's optimal traffic scheduling to sAP k and mBS in Figure 6(a), the bandwidth usage in Figure 6(b), the power allocation in Figure 6(c), and the overall cost in Figure 6(d). Figure 6(a) shows that as $\epsilon_{ik}^{\max}$ increases (meaning that the secrecy requirement on MU i's offloaded data becomes weaker), MU i tends to offload more data to sAP k and send less data to the mBS. Accordingly, Figure 6(b) shows that MU i's optimal bandwidth usage at the mBS and sAP k gradually decrease. Specifically, the mBS's bandwidth usage decreases due to MU i's smaller rate to the mBS. The bandwidth usage at sAP k gradually decreases due to MU i's weaker secrecy-requirement. Figure 6(c) shows the exactly same trend in MU i's power allocation as Figure 6(b). Again, the power allocation to the mBS shows a significant decrease, since less traffic is sent to the mBS. The power allocation to sAP k decreases mainly due to MU i's weaker secrecy requirement. Finally, Figure 6(d) shows that the MU i's overall cost decreases.

B. Numerical Results for the Multiple-MU Multiple-sAP

We further consider the multiple-MU and multiple-sAP case and evaluate the performance of our LiMo-Algorithm, which is used to solve Problem (SOP) and find the optimal offloading-selection. We setup the network topology as follows. The mBS is again located at (0m,0m), and the group $\mathcal{I}$ of MUs are randomly and uniformly located within a disk with the center of (220m, 0m) and a radius of 30m. Meanwhile, the group $\mathcal{K}$ of sAPs are evenly placed on a circle with the center of (220m,0m) and a radius of 30m. Such a setting means that the MUs are relatively closer to the sAPs than the mBS, which is the targeted scenario for offloading the MUs' data. Regarding radio resources, for each sAP $k \in \mathcal{K}$, we set $W_k^{\max} = 20$MHz, and we set $W_B^{\max} = 15$MHz. For each MU $i \in \mathcal{I}$, we set MU i's $p_{iA}^{\max} = 0.25$W and $p_{iB}^{\max} = 0.3$W. Meanwhile, we set each MU i's R_i^{req} to be uniformly random within $[5, 10]$Mbps, and set MU i's security-outage limits $\epsilon_{iA}^{\max}$ and $\epsilon_{iB}^{\max}$ to be uniformly random within $[0.04, 0.06]$ and $[0.1, 0.15]$ respectively. To capture the uncertainty in the eavesdropper's information with respect to different MUs, we set the average channel power gain α_{ik} from MU i to the eavesdropper (on sAP k's channel) to be uniformly random within $[1 \times 10^{-5}, 3 \times 10^{-5}]$, and the average channel power gain α_{iB} from MU i to the eavesdropper (on the mBS's channel) to be uniformly random within $[1 \times 10^{-7}, 2 \times 10^{-7}]$.

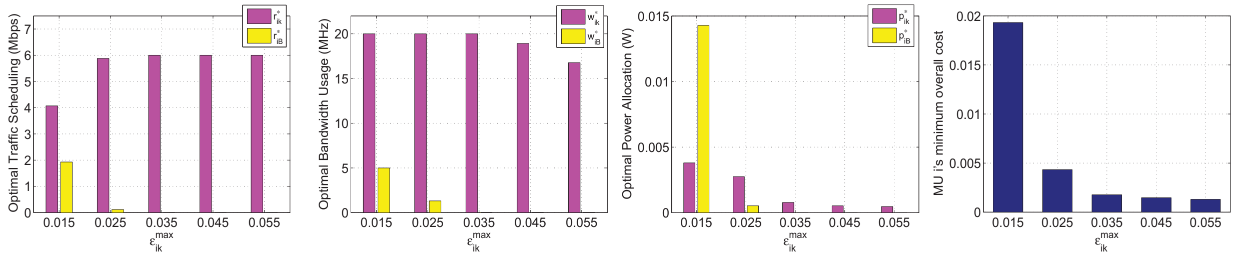
Figure 7 shows the performance of using LiMo-Algorithm to compute the total successfully served MUs' throughput. We consider 6-MU, 8-MU, 10MU, and 12-MU cases, and for each case, we vary the numbers of the sAPs from 2 to 6. Figure 7 shows the total successfully served throughput obtained by our LiMo-Algorithm. Each point denotes the average result of 20 different random realizations. It is reasonable to observe in Figure 7 that given the number of MUs to be served, the total served throughput first increases in the number of the sAPs, since more sAPs can provide a larger offloading capacity to accommodate the MUs' data. Moreover, as the number of the sAPs further increases, the total successfully served throughput keep almost unchanged, since all MUs' demands have been satisfied. To evaluate the accuracy of our LiMO-Algorithm, we use LINGO's global-solver to directly solve Problem (OSP) and obtain 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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology



(a) (b) (c)

Fig. 5: MU i's optimal offloading solution versus different $R_i^{\rm req}$ under the setting of $\epsilon_{ik}^{\rm max} = 0.1$ and $\epsilon_{iB}^{\rm max} = 0.15$, $w_{ik}^{\rm max} = 20$MHz and $w_{iB}^{\rm max} = 5$MHz, and $w_{ik}^{\rm min} = w_{iB}^{\rm min} = 0.01$MHz: (a) MU i's optimal traffic scheduling; (b) MU i's optimal bandwidth usage; (c) MU i's optimal power allocation.



(a) (b) (c) (d)

Fig. 6: Impact of the MU's secrecy-limit $\epsilon_{ik}^{\rm max}$ on the optimal offloading solution under the setting of $\epsilon_{iB}^{\rm max} = 0.1$ and $R_i^{\rm req} = 6$Mbps: (a) MU i's optimal traffic scheduling; (b) MU i's optimal bandwidth usage; (c) MU i's optimal power allocation; (d) MU i's minimum overall cost.

globally optimum solution. By using the LINGO's output as a benchmark, we evaluate the relative differences between the results obtained by LiMo-algorithm and the LINGO's output. We mark out these relative differences in Figure 7, i.e., the number marked above each tested case. The results show that LiMO-Algorithm achieves the results very close to the globally optimum solutions (obtained by LINGO's global-solver), with the largest relative difference no greater than 4% in all cases.

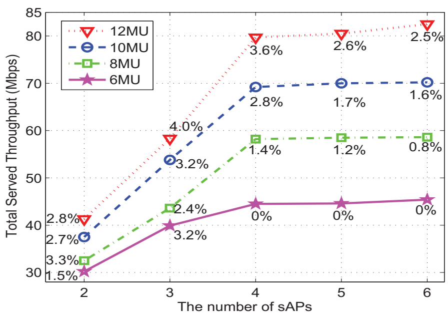
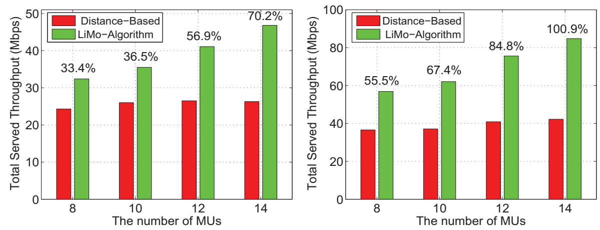


Fig. 7: Performance of LiMo-Algorithm to solve Problem (OSP). The number above each tested case is the relative difference between the result obtained by LiMo-Algorithm and that obtained by LINGO.

Figure 8 and Figure 9 together show the advantage of the optimal offloading-selection solution (obtained by our LiMo-Algorithm) in improving the total served MUs' traffic, in comparison with a (heuristic) distance-based offloading-selection scheme. In the distance-based selection scheme, each MU myopically chooses the closest sAP to offload traffic via DC. Figure 8 shows the results under different settings of the numbers of the MUs and sAPs. Figure 9 shows the results under different MUs' secrecy requirements (but with the setting of 10 MUs and 3sAPs). In both figures, each result denotes the average of 20 random realizations, and the marked number above each case denotes the relative improvement against the distance-based selection. Both Figure 8 and Figure 9 clearly show that using the optimal offloading-selection (obtained by our LiMo-Algorithm) can significantly improve the total served MUs' throughput compared to the distanced-based selection solution. As explained before, in LiMo-Algorithm, we properly select the most beneficial MU-sAP pair (by solving Problem (OSP-Li)) one by one, and take into account the associated radio resource consumptions due to fixing this MU-sAP pair. That is why the optimal offloading-selection achieves such a significant gain in improving the overall served MUs' throughput.



(a) 2 sAPs (b) 4 sAPs

Fig. 8: Advantage of optimal offloading-selections. The number represents the relative improvement against the distance-based selection.

IX. Conclusion

We have studied the resource allocations for the MUs' traffic offloading via DC with guaranteed secrecy. For the single-MU single-sAP case, we proposed an efficient algorithm to compute the MU's optimal traffic scheduling, power allocation,

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

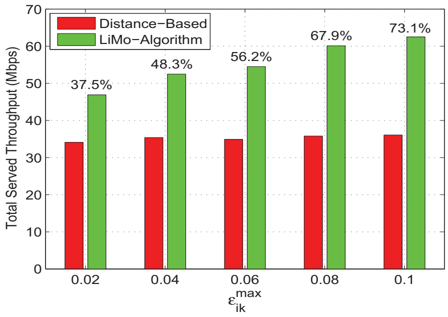


Fig. 9: Impact of the MUs' secrecy-limits on the total served throughput.

and bandwidth usage that together minimize the MU's overall resource consumption, while meeting the MU's traffic and secrecy requirements. Based on the single-MU's optimal offloading solution, we further studied the MUs' optimal offloading-selection problem to maximize the overall successfully served MUs' traffic demands with guaranteed secrecy, and proposed a low-complexity heuristic algorithm that efficiently computes the offloading-selection solution. Numerical results validate the effectiveness of our algorithms and the advantages of our proposed traffic offloading schemes. The proposed MUs' traffic offloading schemes in this work rely on a network-centric implementation. Our future work is to investigate the user-centric implementation of the proposed traffic offloading, in which the MUs are dominant in determining their respective offloading decisions.

APPENDIX I: PROOF OF PROPOSITION 4

To prove Proposition 4, we first show and prove Lemma 3 and Lemma 4 below. Lemma 3 characterizes the monotonicity of x_{iB} in (23) and x_{ik} in (16) with respect to w_{ik}. Lemma 4 characterizes the monotonicity of $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ (i.e., the optimal objective value of (TSPA-Sub)) with respect to w_{ik}.

Lemma *3:* We have the following two properties: (i) x_{iB} in (23) increases in w_{iB}, and (ii) x_{ik} in (16) increases in w_{ik}.

Proof: We first prove Property (i). We use $\frac{\partial x_{iB}}{\partial w_{iB}}$ to denote the first order derivative of (16), which can be given by

$$\frac{\partial x_{iB}}{\partial w_{iB}} = \log_2\left(\frac{p_{iB}g_{iB} + w_{iB}n_0}{p_{iB}\theta_{iB} + w_{iB}n_0}\right) + \frac{w_{iB}n_0p_{iB}}{\ln 2}\frac{(\theta_{iB} - g_{iB})}{(p_{iB}g_{iB} + w_{iB}n_0)(p_{iB}\theta_{iB} + w_{iB}n_0)}. \quad (55)$$

Furthermore, we can derive

$$\frac{\partial^2 x_{iB}}{\partial (w_{iB})^2} = \frac{n_0 p_{iB}}{\ln 2}\frac{(\theta_{iB} - g_{iB})}{(p_{iB}g_{iB} + w_{iB}n_0)(p_{iB}\theta_{iB} + w_{iB}n_0)} \cdot \left(2 - \left(\frac{w_{iB}n_0}{p_{iB}g_{iB} + w_{iB}n_0} + \frac{w_{iB}n_0}{p_{iB}\theta_{iB} + w_{iB}n_0}\right)\right).$$

Notice that $\left(\frac{w_{iB}n_0}{p_{iB}g_{iB}+w_{iB}n_0} + \frac{w_{iB}n_0}{p_{iB}\theta_{iB}+w_{iB}n_0}\right)$ is increasing in w_B, and $\lim_{w_{iB}\to+\infty}\left(\frac{w_{iB}n_0}{p_{iB}g_{iB}+w_{iB}n_0} + \frac{w_{iB}n_0}{p_{iB}\theta_{iB}+w_{iB}n_0}\right) = 2$. Hence $\frac{\partial^2 x_{iB}}{\partial (w_{iB})^2} \leq 0$ always holds. Moreover, since $\lim_{w_{iB}\to+\infty}\frac{\partial x_{iB}}{\partial w_{iB}} = 0$, we conclude that $\frac{\partial x_{iB}}{\partial w_{iB}} \geq 0$. With the similar procedures, we can prove Property (ii). ■

Lemma *4:* Given an arbitrary $\boldsymbol{\epsilon}_i$ and two different $\boldsymbol{w}_i$ and $\boldsymbol{w}'_i$ with $\boldsymbol{w}_i \geq \boldsymbol{w}'_i$, there always exists $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i) \leq V(\boldsymbol{w}'_i, \boldsymbol{\epsilon}_i)$.

Proof: Recall that $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ is the optimal objective value of Problem (TSPA-Sub), which is equivalent to Problem (TSPA-Sub-E). Thus, our focus is to show that for arbitrary given $\boldsymbol{\epsilon}_i$, the optimal objective value $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ of Problem (TSPA-Sub-E) is decreasing in $\boldsymbol{w}_i$. To prove this result, it suffices for us to prove the following two properties:

- Property (i): The objective function of Problem (TSPA-Sub-E) itself is decreasing in $\boldsymbol{w}_i$.
- Property (ii): Constraint (29) leads to a non-smaller feasible interval for r_{ik} when $\boldsymbol{w}_i$ increases.

We next prove these two properties.

Proof of Property (i): First, we prove that $w_{ik}n_0\frac{F(w_{ik})-1}{g_{ik}-\theta_{ik}F(w_{ik})}$ is decreasing in w_{ik}, where $F(w_{ik}) = 2^{\frac{r_{ik}}{(1-\epsilon_{ik})w_{ik}}}$. Specifically, we can derive

$$\frac{d}{d\ w_{ik}}\left(w_{ik}\frac{F(w_{ik})-1}{g_{ik}-\theta_{ik}F(w_{ik})}\right) = \frac{1}{\big(g_{ik}-\theta_{ik}F(w_{ik})\big)^2}\Big(\big(F(w_{ik})-1\big)\big(g_{ik}-\theta_{ik}F(w_{ik})\big) + w_{ik}F'(w_{ik})(g_{ik}-\theta_{ik})\Big).$$

Thus, in order to prove $\frac{d}{d\ w_{ik}}\left(w_{ik}n_0\frac{F(r_{ik})-1}{g_{ik}-\theta_{ik}F(r_{ik})}\right) \leq 0$, we need to prove the following result:

$$\big(F(w_{ik})-1\big)\big(g_{ik}-\theta_{ik}F(w_{ik})\big) + w_{ik}F'(w_{ik})(g_{ik}-\theta_{ik}) \leq 0. \quad (56)$$

For easy presentation, we denote $M = \frac{r_{ik}}{1-\epsilon_{ik}} \geq 0$. Then, we can denote $F(w_{ik}) = 2^{\frac{M}{w_{ik}}}$. To prove (56), we need to prove

$$(2^{\frac{M}{w_{ik}}} - 1)(g_{ik} - \theta_{ik}2^{\frac{M}{w_{ik}}}) - (g_{ik}-\theta_{ik})2^{\frac{M}{w_{ik}}}\frac{M}{w_{ik}}\ln 2 \leq 0.$$

To prove the above inequality, we can first derive

$$\begin{aligned}&(2^{\frac{M}{w_{ik}}} - 1)(g_{ik} - \theta_{ik}2^{\frac{M}{w_{ik}}}) - (g_{ik}-\theta_{ik})2^{\frac{M}{w_{ik}}}\frac{M}{w_{ik}}\ln 2\\ &\leq (2^{\frac{M}{w_{ik}}} - 1)(g_{ik}-\theta_{ik}) - (g_{ik}-\theta_{ik})2^{\frac{M}{w_{ik}}}\frac{M}{w_{ik}}\ln 2\\ &= (g_{ik}-\theta_{ik})\left((2^{\frac{M}{w_{ik}}} - 1) - 2^{\frac{M}{w_{ik}}}\frac{M}{w_{ik}}\ln 2\right).\end{aligned}$$

Due to the fact that $g_{ik} - \theta_{ik} \geq 0$ (Lemma 1), we only need to show $(2^{\frac{M}{w_{ik}}} - 1) - 2^{\frac{M}{w_{ik}}}\frac{M}{w_{ik}}\ln 2 \leq 0$. We denote $t = \frac{M}{w_{ik}} \geq 0$. Then, we only need to prove $2^t t\ln 2 - 2^t + 1 \geq 0$. To show this, we calculate the first order derivative of $(2^t t\ln 2 - 2^t + 1)$ in t:

$$2^t\ln 2 + t\ln 2\cdot 2^t\ln 2 - 2^t\ln 2 \geq 0,$$

which means that $2^t t\ln 2 - 2^t + 1$ is increasing. Moreover, since $2^0\cdot 0\cdot\ln 2 - 2^0 + 1 = 0$, we can conclude that $2^t t\ln 2 - 2^t + 1 \geq 0$ always holds when $t \geq 0$. Therefore, we complete the proof that $w_{ik}n_0\frac{F(r_{ik})-1}{g_{ik}-\theta_{ik}F(r_{ik})}$ is decreasing in w_{ik}.

With the similar arguments and derivations, we can also prove that $w_{iB}n_0\frac{G(r_{ik})-1}{g_{iB}-\theta_{iB}G(r_{ik})}$ is decreasing in w_{iB}.

Therefore, we have completed the proof for Property (i).

Proof of Property (ii): According to (16) and Lemma 3, r_{ik}^{upper} in (22) non-decreasing in w_{ik}. Similarly, according to (23) and Lemma 3, r_{ik}^{lower} in (28) is non-increasing in w_{iB}. Therefore, r_{ik} in Problem (TSPA-Sub-E) experiences a no smaller feasible region when $\boldsymbol{w}_i$ increases. We thus complete the proof.

Thus, we complete the proofs for properties (i) and (ii). As properties (i) and (ii) together are sufficient to ensure that $V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$ decreases in $\boldsymbol{w}_i$. We finish the proof of Lemma 4. ∎

Using Lemmas 3 and 4, we can prove Proposition 4 as follows. Suppose that there exist two different $\boldsymbol{w}_i$ and $\boldsymbol{w}'_i$ with $\boldsymbol{w}_i \geq \boldsymbol{w}'_i$. Corresponding to $\boldsymbol{w}'_i$, we denote $\hat{\epsilon}(\boldsymbol{w}'_i)$ such that $V(\boldsymbol{w}'_i, \hat{\epsilon}(\boldsymbol{w}'_i)) = \min_{\boldsymbol{\epsilon}_i} V(\boldsymbol{w}'_i, \boldsymbol{\epsilon}_i)$. Recall that the optimal function value of Problem (TSPA-Top) $C(\boldsymbol{w}_i) = \min_{\boldsymbol{\epsilon}_i} V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i)$. Thus, we have $C(\boldsymbol{w}'_i) = \min_{\boldsymbol{\epsilon}_i} V(\boldsymbol{w}'_i, \epsilon_i) = V(\boldsymbol{w}'_i, \hat{\epsilon}(\boldsymbol{w}'_i)) \geq V(\boldsymbol{w}_i, \hat{\epsilon}(\boldsymbol{w}'_i)) \geq \min_{\boldsymbol{\epsilon}_i} V(\boldsymbol{w}_i, \boldsymbol{\epsilon}_i) = C(\boldsymbol{w}_i)$, in which the third inequality (i.e., $V(\boldsymbol{w}'_i, \hat{\epsilon}(\boldsymbol{w}'_i)) \geq V(\boldsymbol{w}_i, \hat{\epsilon}(\boldsymbol{w}'_i))$) is based on Lemma 4. Thus, we complete the proof of Proposition 4.

REFERENCES

[1] F. Rebecchi, M. Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data Offloading Techniques in Cellular Networks: A Survey," *IEEE Communication Surveys & Tutorials*, vol.17, no.2, pp.580-603, Second Quarter, 2015.

[2] C. Rosa, K. Pedersen, H. Wang, P. Michaelsen, S. Barbera, E. Malkamaki, T. Henttonen, and B. Sebire, "Dual Connectivity for LTE Small Cell Evolution: Functionality and Performance Aspects," *IEEE Communications Magazine*, no.54, no.6, pp.137-143, June 2016.

[3] S. Jha, K. Sivanesan, R. Vannithamby, and A. Koc, "Dual Connectivity in LTE Small Cell Networks," in *Proc. of IEEE GLOBECOM'2014*, Austin, TX, Dec.8-12, 2014.

[4] J. Liu, J. Liu, and H. Sun, "An Enhanced Power Control Scheme for Dual Connectivity," in *Proc. of VTC-Fall'2014*, Vancouver, Canada, Sept.14-17, 2014.

[5] M. Pan, T. Lin, C. Chiu, and C. Wang, "Downlink Traffic Scheduling for LTE-A Small Cell Networks With Dual Connectivity Enhancement," *IEEE Communications Letters*, vol.20, no.4, pp.796-799, Jan. 2016.

[6] H. Wang, C. Rosa, and K. Pedersen, "Dual connectivity for LTE-advanced Heterogeneous networks," *Wireless Networks*, vol. 22, pp.1315-1328, 2016.

[7] Y. Wu, Y. He, L. Qian, and X. Shen, "Traffic Scheduling and Power Allocations for Mobile Data Offloading via Dual-Connectivity," in *Proc. of ICC'2016*, Kuala Lampur, Malaysia, May 23-27, 2016.

[8] P. Wang, W. Song, D. Niyato, and Y. Xiao, "QoS-Aware Cell Association in 5G Heterogeneous Networks with Massive MIMO," *IEEE Network*, vol.29, no.6, pp.76-82, Nov-Dec. 2015.

[9] X. Chen, J. Wu, Y. Cai, H. Zhang, and T. Chan, "Energy-Efficiency Oriented Traffic Offloading in Wireless Networks: A Brief Survey and A Learning Approach for Heterogeneous Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol.33, no.4, pp.627-640, April 2015.

[10] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, "Energy-Aware Traffic Offloading for Green Heterogeneous Networks", *IEEE Journal Selected Areas of Communications*, vol.34, no.5, pp.1116-1129, May 2016.

[11] H. Zhang, C. Jiang, J. Cheng, and V. Leung, "Cooperative Interference Mitigation and Handover Management for Heterogeneous Cloud Small Cell Networks," *IEEE Wireless Communications*, vol.22, no.3, pp.92-99, June 2015.

[12] S. Singh, H. Dhillon, and J. Andrews, "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights," *IEEE Transations on Wireless Communications*, vol.12, no.5, pp.2484-2497, May 2013.

[13] J. Rao, and A. Fapojuwo, "Analysis of Spectrum Efficiency and Energy Efficiency of Heterogeneous Wireless Networks With Intra-Inter-RAT Offloading," *IEEE Transactions on Vehicluar Technology*, vol.64, no.7, pp.3120-3139, July 2015.

[14] Q. Ye, W. Zhuang, L. Li, and P. Vigneron, "Traffic Load Adaptive Medium Access Control for Fully-Connected Mobile Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, DOI:10.1109/TVT.2016.2516910.

[15] Y. Yang, T. Quek, and L. Duan, "Backhaul-Constrained Small Cell Networks: Refunding and QoS Provisioning," *IEEE Transactions on Wireless Communications*, vol.13, no.9, pp.5148-5161, Sep. 2014.

[16] H. Liu, H. Zhang, J. Cheng, and V. Leung, "Energy Efficient Power Allocation and Backhaul Design in Heterogeneous Small Cell Networks," in *Proc. of IEEE ICC'2016*, Kuala Lampur, Malaysia, May 23-27, 2016.

[17] P. Lin, J. Zhang, Q. Zhang, and M. Hamdi, "Enabling the Femtocells: A Cooperation Framework for Mobile and Fixed-line Operators," *IEEE Transactions on Wireless Communications*, vol.12, no.1, pp.158-167, Jan. 2013.

[18] Q. Chen, G. Yu, H. Shan, A. Maaref, G. Li, and A. Huang, "Cellular Meets WiFi: Traffic Offloading or Resource Sharing?" *IEEE Transactions on Wireless Communications*, vol.15, no.5, pp.3354-3367, Jan. 2016.

[19] M. Bennis, M. Simsek, A. Czylwik, W. Saad, S. Valentin, and M. Debbah, "When Cellular Meets WiFi in Wireless Small Cell Networks," *IEEE Communications Magazine*, vol.51, no.6, pp.44-50, June 2013.

[20] S. Andreev, A. Pyattaev, K. Johnsson, O. Galinina, and Y. Koucheryavy, "Cellular Traffic Offloading onto Network-Assisted Device-to-Device Connections," *IEEE Communications Magazine*, vol.52, no.4, pp.20-31, April 2014.

[21] G. Yu, Y. Jiang, L. Xu, and G. Li, "Multi-Objective Energy-Efficient Resource Allocation for Multi-RAT Heterogeneous Networks," *IEEE Journal on Selected Areas in Communications*, vol.33, no.10, pp.2118-2127, Oct. 2015.

[22] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A Survey on Wireless Security: Technical Challenges, Recent Advances and Future Trends," avaiable online at http://arxiv.org/abs/1505.07919.

[23] H. Wang, X. Zhou, and M. Reed, "Physical Layer Security in Cellular Networks: A Stochastic Geometry Approach," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2776-2787, June 2013.

[24] J. Zhu, R. Schober, and V. Bhargava, "Secure Transmission in Multicell Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4766-4781, Sep. 2014.

[25] H. Wu, X. Tao, N. Li, and J. Xu, "Secrecy Outage Probability in Multi-RAT Heterogeneous Networks," *IEEE Communications Letters*, vol. 20, no. 1, pp. 53-56, Jan. 2016.

[26] J. Yue, C. Ma, H. Yu, and W. Zhou, "Secrecy-based Access Control for Device-to-Device Communication Underlaying Cellular Networks," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2068-2071, Nov. 2013.

[27] M. Waler, "On the Security of 3GPP Networks," 3GPP SA3-Security, avaiable online at ftp://www.3gpp.org/tsg_sa/WG3_Security/TSGS3_13_Yokohama/Docs/PDF/S3-000326.pdf.

[28] Y. Qian, and N. Moayeri, "Design of Secure and Application-Oriented VANETs," in *Proc. of IEEE VTC'2008-Spring*, pp.2794-2799, Singapore, May 11-14, 2008.

[29] J. Choi, and S. Jung, "A Security Framework with Strong Non-Repudiation and Privacy in VANETs," in *Proc. of IEEE CCNC'2009*, pp.1-5, Las Vegas, NV, Jan.10-13, 2009.

[30] Y. Qian, K. Lu, and N. Moayeri, "A Secure VANET MAC Protocol for DSRC Applications", in *Proc. of IEEE GLOBECOM'2008*, New Orleans, LA, Nov.30-Dec.4, 2008.

[31] P. Papadimitratos, and J. Hubaux, Report on the "Secure Vehicular Communications: Results and Challenges Ahead" workshop, *ACM SIGMOBILE Mobile Computing and Communications Review*, vol.12 no.2, pp.53-64, April 2008.

[32] G. Samara, W. Al-Salihy, and R. Sures, "Security Analysis of Vehicular Ad Hoc Nerworks (VANET)," in *Proc. of International Conference on Network Applications, Protocols and Services 2010*, pp.55-60, Malaysia, Sept.22-23, 2010.

[33] C. Lai, K. Zhang, N. Cheng, H. Li, and X. Shen, "SIRC: A Secure Incentive Scheme for Reliable Cooperative Downloading in Highway VANETs", *IEEE Transactions on Intelligent Transportation Systems*, vol.PP, no.99, pp.1-16, DOI:10.1109/TITS.2016.2612233, Oct. 2016.

[34] Y. Jing, J. Zhu, and Y. Zou, "Secrecy Outage Analysis of Multi-User Cellular Networks in the Face of Cochannel Interference," in *Proc. of IEEE International Conf. on Cognitive Informatics & Cognitive Computing*, Beijing, China, July 6-7, 2015.

[35] J. Barros, and M. Rodrigues, "Secrecy Capacity of Wireless Channels," in *Proc. of ISIT'2006*, Seattle, WA, July 7-14, 2016.

[36] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[37] Y. Zhang, L. Qian, and J. Huang, "Monotonic Optimization in Communication and Networking Systems," *Foundation and Trends in Networking*, Now Publisher, October 2013.

[38] H. Tuy, "Monotonic Optimization: Problems and Solution Approaches," *SIAM Journal of Optimization*, vol.11, no.2, pp.464-494, 2000.

[39] National Instruments, "Introduction to UMTS Device Testing Transmitter and Receiver Measurements for WCDMA Devices," available online at http://download.ni.com/evaluation/rf/Introduction_to_UMTS_Device_Testing.pdf.

[40] O. Bejarano, E. Knightly, "IEEE 802.11ac: From Channelization to Multi-User MIMO," *IEEE Communications Magazine*, vol.51, no.10, pp.84-90, Oct. 2013.

[41] O. Kundakcioglu, and S. Alizamir, "Generalized Assignment Problem," in *Encyclopedia of Optimization*, pp.1153-1162, Springer-Verlag Press, 2009.

[42] M. Garey, and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman & Co., New York, 1990.

[43] N. Karmarkar, "A New Polynomial-Time Algorithm for Linear Programming", *Combinatorica*, vol.4, no.4, pp.373-396, 1984.

[44] L. Schrage, *Optimization Modeling with LINGO*, The 5th Edition, Lindo System, Jan. 1999.

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/TVT.2017.2689504, IEEE Transactions on Vehicular Technology

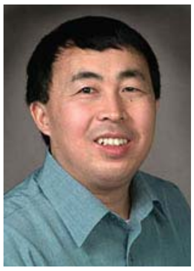
Yuan Wu (S'08-M'10-SM'16) received the Ph.D degree in Electronic and Computer Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2010. He is currently an Associate Professor in the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. During 2010-2011, he was the Postdoctoral Research Associate at the Hong Kong University of Science and Technology. During 2016-2017, he is the visiting scholar at the Broadband Communications Research (BBCR) group, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include optimal radio resource management for wireless communications and networks, green communications and networking, and smart grid. He is the recipient of the Best Paper Award in IEEE International Conference on Communications (ICC) in 2016 and the recipient of the Best Paper Award in International Conference on Wireless Communications and Signal Processing (WCSP) in 2016. He served/serves as the guest editor of IEEE Network with the special issue on "Vehicle-to-Grid Networks" in 2016, IEEE Transactions on Industrial Informatics with the special issue on "Energy Informatics for Green Cities" in 2017, and IET Communications with the special issue on "Advances in Enabling Technologies for Green Communications and Networking" in 2017.

Jiaochao Zheng(M'16) received his B.S. degree in Communications Engineering, and Ph.D. degree in Communications and Information Systems from College of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2010 and 2016 respectively. He is currently an assistant professor in College of Communications Engineering, PLA University of Science and Technology. During 2015-2016, he is the visiting scholar at the Broadband Communications Research (BBCR) group, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests focus on Interference mitigation techniques, learning theory, game theory, and optimization techniques. He has published several papers in international conferences and reputed journals in his research area.

Kuanyang Guo is currently pursuing his M.S. degree in College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His research interest focuses on radio resource allocations for wireless communications and networks, and secrecy-based communications.

Li Ping Qian (S'08-M'10-SM'16) received the Ph.D. degree in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2010. From 2010 to 2011, she was a Post- Doctoral Research Assistant with the Department of Information Engineering, The Chinese University of Hong Kong. She was a Visiting Student with Princeton University in 2009. She was with Broadband Communications Research Laboratory, University of Waterloo, from 2016 to 2017. She is currently an Associate Professor with the College of Information Engineering, Zhejiang University of Technology, China. Her research interests lie in the areas of wireless communication and networking, cognitive networks, and smart grids, including power control, adaptive resource allocation, cooperative communications, interference cancellation for wireless networks, and demand response management for smart grids.

Dr. Qian was a co-recipient of the IEEE Marconi Prize Paper Award in wireless communications (the Annual Best Paper Award of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS) in 2011, the Second-class Outstanding Research Award for Zhejiang Provincial Universities, China, in 2012, the Second-class Award of Science and Technology given by Zhejiang Provincial Government in 2015, and the Best Paper Award from ICC2016. She was a finalist of the Hong Kong Young Scientist Award in 2011, and a recipient of the scholarship under the State Scholarship Fund of China Scholarship Council for Visiting Scholars in 2015, and the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars in 2015. She serves as an Associate Editor of the IET Communications.



Xuemin (Sherman) Shen (IEEE M97-SM02-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, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is also the Associate Chair for Graduate Studies. Dr. Shen's research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He is an elected member of IEEE ComSoc Board of Governor, and the Chair of Distinguished Lecturers Selection Committee. 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 QShine'06, and the Chair for IEEE Communications Society Technical Committee on Wireless Communications. He also serves/served as the Editor-in-Chief for IEEE IoT Journal, IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; an Associate Editor-in-Chief for IEEE Internet of Things Journal, 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, IEEE Communications Magazine, and ACM Mobile Networks and Applications, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo, and the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. Dr. Shen is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, 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.

Yueming Cai (M'05-SM'12) received the B.S. degree in Physics from Xiamen University, Xiamen, China in 1982, the M.S. degree in Micro-electronics Engineering and the Ph.D. degree in Communications and Information Systems both from Southeast University, Nanjing, China in 1988 and 1996 respectively. His current research interest includes cooperative communications, signal processing in communications, wireless sensor networks, and physical layer security.