Mobile Edge Computing-Enabled Heterogeneous Networks

Chanwon Park and Jemin Lee, *Member, IEEE*

***Abstract*—The mobile edge computing (MEC) has been introduced for providing computing capabilities at the edge of networks to improve the latency performance of wireless networks. In this paper, we provide the novel framework for MEC-enabled heterogeneous networks (HetNets) , composed of the multi-tier networks with access points (APs) (i.e., MEC servers), which have different transmission power and different computing capabilities. In this framework, we also consider multiple-type mobile users with different sizes of computation tasks, and they offload the tasks to a MEC server, and receive the computation resulting data from the server. We derive the successful edge computing probability considering both the computation and communication performance using the queueing theory and stochastic geometry. We then analyze the effects of network parameters and bias factors in MEC server association on the successful edge computing probability. We provide how the optimal bias factors in terms of successful edge computing probability can be changed according to the user type and MEC tier, and how they are different to the conventional ones that did not consider the computing capabilities and task sizes. It is also shown how the optimal bias factors can be changed when minimizing the mean latency instead of successful edge computing probability. This study provides the design insights for the optimal configuration of MEC-enabled HetNets.**

***Index Terms*—Mobile edge computing, heterogeneous network, latency, offloading, queueing theory, stochastic geometry**

I. Introduction

As a wireless communication is getting improved, mobile users are processing a numerous and complex computation tasks. To support the mobile users, the mobile cloud computing has been considered, which enables the centralization of the computing resources in the clouds. On the other hands, in recent years, the computation and battery capabilities of mobile users have been improved, which enables the mobile users to process the complex computation tasks. For that reason, the computation tasks start to be performed in the network edge including mobile users or servers located in small-cell access point (AP) and it is called the mobile edge computing (MEC) [1].

One of the main requirements of future wireless communications is the ultra-low latency. The cloud-radio access network (C-RAN) has been introduced to lower the computation latency of mobile users by making them offload complex computation tasks to a centralized cloud server [2]. However, to utilize the C-RAN, we need to experience inevitable long communication latency to reach to the far located central server. When the MEC is applied, mobile users can compute the large tasks by offloading to the nearby MEC servers, instead of the central server [3]. Although the computing capabilities of MEC servers can be lower than those of the C-RAN servers, offloading tasks to MEC servers can be more benefitial for some latency-critical applications such as autonomous vehicles and sensor networks for health-care services. Hence, the MEC becomes one of the key technologies for future wireless networks.

The performance of MEC in wireless networks has been studied, mostly focusing on the minimization of energy consumption or communication and computing latency. Specifically, the energy minimization problem has been considered for proposing or optimizing the policy of offloading to MEC servers with guaranteeing a certain level of latency [4]–[6]. Users with different computing capabilities are considered in [4], and a single MEC server [4], [5] or multiple MEC servers [6] are used for each user. The energy minimization problem has also been investigated for a energy harvesting user [7] and multicell MIMO systems [8]. The latency minimization problem has been considered by analyzing the computation latency at MEC servers using queueing theory [9]–[12]. The optimal offloading policy was presented for minimizing the mean latency [9] or maximizing the probability of guaranteeing the latency requirements [10]. Recently, the tradeoff between the latency and communication performance (i.e., network coverage) has also been presented in [11].

However, except for [10], most of the prior works are based on the mean (or constant) computation latency, which fails to show the impact of latency distribution on the MEC network performance. Furthermore, there is no work that considers the heterogeneous MEC servers, which have different computing capabilities and transmission power, and various sizes of user tasks, impeding the efficient design of MEC-enabled heterogeneous network (HetNet). In the future, the MEC will be applied not only to APs or base stations (BSs) but also to all computing devices around us such as laptops or mobile devices. Therefore, it is required to investigate how to design the MEC-enabled network that has various types of MEC servers, which is the main objective of this paper.

The HetNet has been studied when APs have different resources such as transmission power [13]–[18], mainly by focusing on the communication performance, not the computing performance. In most of the works, the stochastic geometry has been applied for the spatial model of distributed users and APs

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2017R1C1B2009280). The material in this paper was presented, in part, at the International Conference on Communications, Kansas City, MO, May 2018.

C. Park and J. Lee are with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea, 42988 (e-mail:pcw0311@dgist.ac.kr, jmnlee@dgist.ac.kr).

using Poisson point processes (PPPs) [19]. For example, the baseline model containing the outage probability and average rate for downlink is shown in [13]. The network modeling and coverage analysis are provided in [14] for downlink, in [15] for uplink, and in [16] for decoupling of uplink and downlink. The cell range expansion for load balancing among APs is considered in [14] and [17]. The HetNets with line-of-sight and non-line-of-sight link propagations are also investigated in [18] and [20]. Recently, the stochastic geometry has also been applied for the performance analysis of randomly distributed MEC servers in [11], but the latency distribution, heterogeneous MEC servers, and various sizes of user tasks are not considered for the design of MEC-enabled networks.

In this work, we provide the novel framework of the MEC-enabled HetNets. We consider the multi-tier networks composed of MEC servers having different computing capacities and the multi-type users having different computation task sizes. The MEC servers and users are distributed by PPPs, and users associate to the MEC server based on the bias association rule. The successful edge computing probability is defined to measure the probability that the total latency is less than the target latency, while the communications during offloading to and receiving from the MEC server are reliable. For analyzing the successful edge computing probability, the latency distribution is derived using the queueing theory, and the successful communication probability in HetNet is used for the analysis of the communication reliability. Moreover, we present the mean latency and analyze how differently the bias factor needs to be determined for lower mean latency, compared to the case for successful edge computing probability. The main contribution of this paper can be summarized as below:

- we develop the novel framework of the MEC-enabled HetNets characterized by the *multi-tier* MEC server having different computing capacities and *multi-type* users having different computation task sizes;
- we introduce and derive the successful edge computing probability, which considers both the computing and communication performance, using stochastic geometry and queueing theory; and
- we analyze how the association bias factors for different MEC tiers and different user types affect the successful edge computing probability, and how the optimal bias factors are different to the conventional ones, which do not consider the computing performance of MEC server.

The remainder of this paper is organized as follows. Section II describes the MEC-enabled HetNet model and gives the queueing model for calculating the latency. Section III defines the successful edge computing probability and analyzes the successful computation and communication probability. Section IV presents the effects of association bias and network parameters on the successful edge computing probability and compares the successful edge computing probability and mean latency. Finally, conclusions are given in Section V.

II. MEC-enabled Heterogeneous Network Model

In this section, we present the network model and the latency model of the MEC-enabled HetNet.

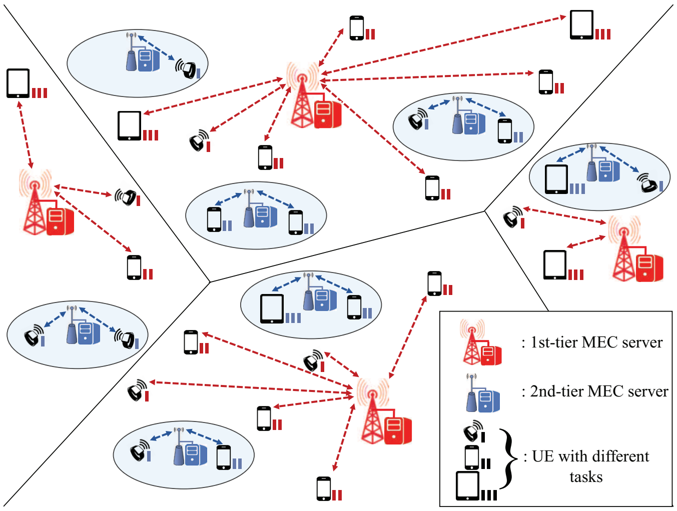

Fig. 1. The MEC-enabled HetNet networks composed of the multi-tier MEC servers and the multi-type users with different size of tasks.

A. Network Model

We consider a MEC-enabled HetNet, which is composed of N_m tiers of MEC servers, located in APs. An example of the network is given in Fig. 1. MEC servers in different tiers have different compution and transmission capabilities. We use $\mathcal{K} = \{1, \cdots, N_m\}$ as the index set of N_m tiers of MEC servers. The MEC servers are distributed according to a homogeneous PPP $\mathbf{\Phi}_{\mathrm{m}}$ with spatial density λ_{m}. The locations of the kth-tier MEC servers are also modeled as a homogeneous PPP $\mathbf{\Phi}_{\mathrm{m},k}$ with spatial density $\lambda_{\mathrm{m},k} = p_{\mathrm{m},k}\lambda_{\mathrm{m}}$ where $p_{\mathrm{m},k}$ is the portion of the kth-tier MEC servers. Each servers are assumed to have one CPU and one queue, which has an infinite waiting space. MEC servers in the kth tier transmit with the power $P_{\mathrm{m},k}$. A channel is assigned to one user only in the cell of MEC server (AP), and the ratio of users using the same uplink channel is κ, which denotes the frequency reuse factor. The density of uplink interfering users offloading to the kth-tier MEC server is then given by $\kappa\lambda_{\mathrm{m},k}$.

Computing the large computation task at a mobile device may not be finished within a required time. Hence, users offload their tasks to the MEC servers, which compute/process the tasks and send the resulting data back to the user. Users are categorized into N_u types according to the size of their offloading tasks. The $\mathcal{I} = \{1, 2, \cdots, N_u\}$ denotes the index set of N_u type users. The ith-type user offloads the computation task with $D_i^{(c)}$ packets to a MEC server. For that, the user transmits the computation request message in $D_i^{(r)}$ packets to the MEC server with the power $P_{\mathrm{u},i}$ (i.e., uplink transmission), and receives the computation resulting data in $D_i^{(d)}$ packets (i.e., downlink transmission). Each transmission is happened in a time slot T_{s}.[1] The locations of users are modeled as a homogeneous PPP $\mathbf{\Phi}_{\mathrm{u}}$ with spatial density λ_{u}. The ith-type users are distributed according to a homogeneous PPP $\mathbf{\Phi}_{\mathrm{u},i}$

[1] Note that the whole uplink and downlink transmission time can consist of different number of time slots. Here, we focus on the transmission in one time slot T_{s} for both uplink and downlink, and the arrival task rate at a MEC server is defined as the number of arriving requests per T_{s}.

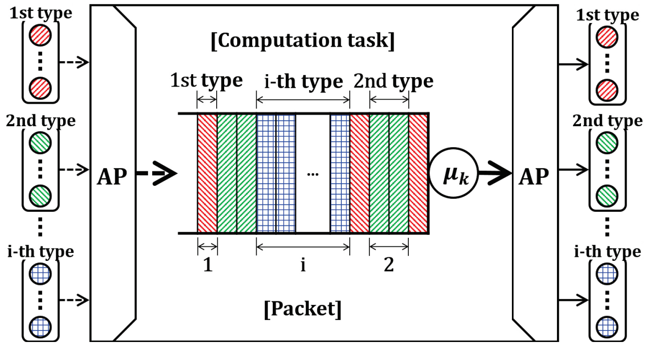

Fig. 2. The kth-tier MEC server when the users offload their computation tasks and receive their resulting data.

with spatial density $\lambda_{\mathrm{u},i} = p_{\mathrm{u},i}\lambda_{\mathrm{u}}$, where $p_{\mathrm{u},i}$ is the portion of the ith-type users.

B. Latency Model

In MEC-enabled HetNet, users can have MEC server computing latency. The MEC server computing latency is caused when the computation tasks are offloaded to the MEC server. When an ith-type user offloads to the kth-tier MEC server, the total latency of the ith-type user, denoted by $T_{\mathrm{total},i}$, can be defined as

$$T_{\mathrm{total},i} = 2T_{\mathrm{s}} + T_{\mathrm{c},i,k} \tag{1}$$

for k≥1 where $T_{\mathrm{c},i,k}$ is the computing latency of the ith-type user at the kth-tier MEC server. In (1), $T_{\mathrm{total},i}$ contains $2T_{\mathrm{s}}$ because the MEC server computing latency includes the uplink and downlink transmission time. Note that the retransmission caused by the communication outage is not considered. In this paper, when the ith-type users offload to the kth-tier MEC server, $T_{\mathrm{c},i,k}$ is given by $T_{\mathrm{c},i,k} = T_{\mathrm{w},k} + T_{\mathrm{sv},i,k}$, where $T_{\mathrm{w},k}$ is the waiting time at the queue and $T_{\mathrm{sv},i,k}$ is the service (computing) time. To analyze the MEC server computing latency, the arrival rate of the computation tasks and the service time distribution at the MEC server need to be determined.

Since the users are distributed as a PPP, the arrivals of computing tasks at the server follow a Poisson process with a certain arrival rate. Here, the arrival rate for the ith-type user task to the kth-tier MEC server, denoted by $\nu_{i,k}$, is determined as

$$\nu_{i,k} = \frac{\lambda_{\mathrm{u}} p_{\mathrm{u},i} p_{\mathrm{o},i,k}}{\lambda_{\mathrm{m}} p_{\mathrm{m},k}} \tag{2}$$

where $p_{\mathrm{o},i,k}$ is the probability that the ith-type user offloads to the kth-tier MEC server. When users offload the computation tasks, we assume users select the MEC servers using the association rule, which is based on the biased average receive power, defined as [14]

$$k = \arg\max_{j\in\mathcal{K}} \left\{ \max_{\mathbf{x}_j \in \mathbf{\Phi}_{\mathrm{m},j}} \mathcal{W}_{i,j} Z^{-\alpha}_{\mathbf{x}_{\mathrm{o}},\mathbf{x}_j} \right\} \tag{3}$$

where $Z_{\mathbf{x},\mathbf{y}}$ is the distance between $\mathbf{x}$ and $\mathbf{y}$, α is the pathloss exponent, $\mathcal{W}_{i,j} = P_{\mathrm{m},j}B_{i,j}$ is the weighting factor for the ith-type user offloading to the jth-tier MEC server, and $B_{i,j}$ is the bias factor. This shows that the ith-type user located at $\mathbf{x}_{\mathrm{o}}$ is associated to the kth-tier MEC server. Based on the association rule in (3), $p_{\mathrm{o},i,k}$ in (2) is given by [14]

TABLE I
NOTATIONS USED THROUGHOUT THE PAPER.

Notation	Definition
$\mathbf{\Phi}_{\mathrm{m},k}$	PPP for the kth-tier MEC servers distribution
$\mathbf{\Phi}_{\mathrm{u},i}$	PPP for the ith-type users distribution
$\lambda_{\mathrm{m},k}$	Spatial density of the kth-tier MEC servers
$\lambda_{\mathrm{u},i}$	Spatial density of the ith-type users
$\nu_{i,k}$	Arrival rate of the ith-type users to the kth-tier MEC servers
μ_k	Service rate of the kth-tier server
$p_{\mathrm{u},i}$	Portion of the ith-type users
$p_{\mathrm{m},k}$	Portion of the kth-tier MEC servers
$D_i^{(r)}$	Computation request message size
$D_i^{(c)}$	Computation task size
$D_i^{(d)}$	Computation resulting data size
$P_{\mathrm{u},i}$	Transmission power of the ith-type user
$P_{\mathrm{m},k}$	Transmission power of the kth-tier MEC server
T_{s}	Time slot
T_{t}	Target latency
$T_{\mathrm{total},i}$	Total latency of the ith-type user
$T_{\mathrm{c},i,k}$	Computing latency of the ith-type user offloading to the kth-tier MEC server
$T_{\mathrm{w},k}$	Waiting time at the kth-tier MEC server
$T_{\mathrm{sv},i,k}$	Service time of the ith-type user offloading to the kth-tier MEC server
$B_{i,k}$	Bias factor of the ith-type user offloading to kth-tier MEC server
$\mathcal{W}_{i,k}$	Weighting factor of the ith-type user offloading to kth-tier MEC server
$p_{\mathrm{o},i,k}$	The probability that the ith-type user offloads to the kth-tier MEC server
$p_{\mathrm{s},i,k}$	Successful edge computing probability that the ith-type user offloads to the kth-tier MEC server
$p_{\mathrm{cp},i,k}$	Successful computing probability that the ith-type user offloads to the kth-tier MEC server
$p_{\mathrm{cm},i,k}$	Successful communication probability that the ith-type user offloads to the kth-tier MEC server

$$p_{\mathrm{o},i,k} = 2\pi\lambda_{\mathrm{m},k} \int_0^\infty x \exp\left\{ -\pi x^2 \sum_{j\in\mathcal{K}} \lambda_{\mathrm{m},j} \hat{\mathcal{W}}_{i,j}^{2/\alpha} \right\} dx \tag{4}$$

where $\hat{\mathcal{W}}_{i,j}$ is $\mathcal{W}_{i,j}/\mathcal{W}_{i,k}$.

The distribution of the service time for one packet in the kth-tier MEC server is modeled as the exponential distribution with $1/\mu_k$, where μ_k is the service rate of the kth-tier server. Since the ith-type user offloads $D_i^{(c)}$ packets of the task, $T_{\mathrm{sv},i,k}$

follows the Erlang distribution, and the probability density function (pdf) of $T_{\mathrm{sv},i,k}$, $f_{T_{\mathrm{sv},i,k}}(t)$, is given by

$$f_{T_{\mathrm{sv},i,k}}(t)=\frac{\mu_k^{D_i^{(c)}}t^{D_i^{(c)}-1}\exp\left(-\mu_k t\right)}{\left(D_i^{(c)}-1\right)!}. \tag{5}$$

The service time in the kth-tier MEC server, denoted by $T_{\mathrm{sv},k}$, is the weighted sum of $T_{\mathrm{sv},i,k}$ given by $T_{\mathrm{sv},k}=\sum_{i\in\mathcal{I}}\left(\nu_{i,k}/\nu_k\right)T_{\mathrm{sv},i,k}$ where $\nu_k=\sum_{i\in\mathcal{I}}\nu_{i,k}$ is the arrival rate of users offloading to kth-tier MEC server. Hence, the pdf of $T_{\mathrm{sv},k}$ is given by

$$f_{T_{\mathrm{sv},k}}(t)=\sum_{i\in\mathcal{I}}\left(\frac{\nu_{i,k}}{\nu_k}\right)f_{T_{\mathrm{sv},i,k}}(t). \tag{6}$$

III. Successful Edge Computing Probability Analysis in HetNet

In this section, we derive the successful edge computing probability as the performance metric for MEC-enabled HetNet. The successful edge computing probability for the ith-type user offloading to the kth-tier MEC server is defined as

$$p_{\mathrm{s},i,k}=p_{\mathrm{cm},i,k}p_{\mathrm{cp},i,k} \tag{7}$$

where $p_{\mathrm{cp},i,k}$ and $p_{\mathrm{cm},i,k}$ are the successful computation and communication probability, respectively. The $p_{\mathrm{cp},i,k}$ is the probability that the computation is finished within a target latency, T_{t}. The $p_{\mathrm{cm},i,k}$ is the probability that the data rate is bigger than the target rate during the uplink and downlink transmission. Using the law of total probability, the successful edge computing probability p_{s} is given by

$$p_{\mathrm{s}}=\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}}p_{\mathrm{u},i}p_{\mathrm{o},i,k}p_{\mathrm{s},i,k}. \tag{8}$$

A. Successful Computation Probability

When the ith-type user offloads to the kth-tier MEC server, $p_{\mathrm{cp},i,k}$ is defined by

$$p_{\mathrm{cp},i,k}=\mathbb{P}\left\{T_{\mathrm{w},k}\leqq T_{\mathrm{t}}-T_{\mathrm{sv},i,k}-2T_{\mathrm{s}}\right\}. \tag{9}$$

Using (9) and [21], $p_{\mathrm{cp},i,k}$ is derived in the following theorem.

Theorem 1: The successful computation probability of the ith-type user offloading to the kth-tier MEC server, $p_{\mathrm{cp},i,k}$, is given by

$$\begin{aligned}p_{\mathrm{cp},i,k}=&\int_0^\infty\int_0^{T_{\mathrm{t}}-2T_{\mathrm{s}}-r}\mathcal{L}^{-1}_{T_{\mathrm{w},k}}\left[\frac{\left(1-\rho_k\right)s}{s-\nu_k+\nu_k\mathcal{L}_{T_{\mathrm{sv},k}}(s)}\right]\\&\times\frac{\mu_k^{D_i^{(c)}}r^{D_i^{(c)}-1}\exp\left(-\mu_k r\right)}{\left(D_i^{(c)}-1\right)!}dtdr\end{aligned} \tag{10}$$

where $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}(s)$ is the inverse Laplace transform of the waiting time distribution and ρ_k is the utilization factor of the kth-tier MEC server given by

$$\rho_k=\sum_{i\in\mathcal{I}}\nu_{i,k}\frac{D_i^{(c)}}{\mu_k} \tag{11}$$

for $0\leqq\rho_k<1$. In (10), $\mathcal{L}_{T_{\mathrm{sv},k}}(s)$ is the Laplace transform of the service time distribution in kth-tier MEC server given by

$$\mathcal{L}_{T_{\mathrm{sv},k}}(s)=\sum_{i\in\mathcal{I}}\frac{\nu_{i,k}}{\nu_k}\left(\frac{\mu_k}{s+\mu_k}\right)^{D_i^{(c)}}. \tag{12}$$

Proof: The Laplace transform of the waiting time distribution is refered to as the Pollaczek-Khinchin (P-K) transform equation of M/G/1 queue in [21]. Using the equation, $\mathcal{L}_{T_{\mathrm{w},k}}(s)$ is given by

$$\mathcal{L}_{T_{\mathrm{w},k}}(s)=\frac{\left(1-\rho_k\right)s}{s-\nu_k+\nu_k\mathcal{L}_{T_{\mathrm{sv},k}}(s)}. \tag{13}$$

The $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}(s)$ is obtained using (13), which shows the pdf of the waiting time distribution. Since $T_{\mathrm{sv},i,k}$ is a random variable with the pdf in (5), $p_{\mathrm{cp},i,k}$ is given by

$$p_{\mathrm{cp},i,k}=\int_0^\infty\int_0^{T_{\mathrm{t}}-2T_{\mathrm{s}}-r}\mathcal{L}^{-1}_{T_{\mathrm{w},k}}\left[\frac{\left(1-\rho_k\right)s}{s-\nu_k+\nu_k\mathcal{L}_{T_{\mathrm{sv},k}}(s)}\right]f_{T_{\mathrm{sv},i,k}}(r)\,dtdr. \tag{14}$$

According to [21], ρ_k and $\mathcal{L}_{T_{\mathrm{sv},k}}(s)$ are given by, respectively

$$\rho_k=\nu_k\mathbb{E}\left\{T_{\mathrm{sv},k}\right\}=\nu_k\sum_{i\in\mathcal{I}}\frac{\nu_{i,k}}{\nu_k}\frac{D_i^{(c)}}{\mu_k}=\sum_{i\in\mathcal{I}}\nu_{i,k}\frac{D_i^{(c)}}{\mu_k} \tag{15}$$

$$\begin{aligned}\mathcal{L}_{T_{\mathrm{sv},k}}(s)&=\sum_{i\in\mathcal{I}}\frac{\nu_{i,k}}{\nu_k}\mathcal{L}\left[\frac{\mu_k^{D_i^{(c)}}t^{D_i^{(c)}-1}\exp\left(-\mu_k t\right)}{\left(D_i^{(c)}-1\right)!}\right]\\&=\sum_{i\in\mathcal{I}}\frac{\nu_{i,k}}{\nu_k}\left(\frac{\mu_k}{s+\mu_k}\right)^{D_i^{(c)}}.\end{aligned} \tag{16}$$

Substituting (15), (16) and (5) into (13), (14) becomes (10). ■

The $p_{\mathrm{cp},i,k}$ is hard to be presented in a closed form because of the inverse Laplace transform. However, $p_{\mathrm{cp},k}$ can be given in a closed form for some cases as the following corollaries.

Corollary 1: For the user type set $\mathcal{I}=\{1\}$ and $D_i^{(c)}=1$, $p_{\mathrm{cp},i,k}$ is given by

$$p_{\mathrm{cp},i,k}=1-\exp\{\left(-\mu_k+\nu_k\right)\left(T_{\mathrm{t}}-2T_{\mathrm{s}}\right)\}. \tag{17}$$

Proof: For the user type set $\mathcal{I}=\{1\}$ and $D_i^{(c)}=1$, $\mathcal{L}_{T_{\mathrm{sv},k}}(s)=\frac{\mu_k}{s+\mu_k}$, so we have

$$\mathcal{L}_{T_{\mathrm{w},k}}(s)=\frac{\left(1-\rho_k\right)s}{s-\nu_k+\frac{\nu_k\mu_k}{s+\mu_k}}=\left(1-\rho_k\right)\left\{1+\frac{\nu_k}{s+\mu_k-\nu_k}\right\}. \tag{18}$$

Since $\mathcal{L}^{-1}\{\frac{1}{s+\mu_k-\nu_k}\}=\exp\{\left(-\mu_k+\nu_k\right)t\}$, $f_{T_{\mathrm{w},k}}(t)$, i.e., $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}(s)$, is presented by

$$f_{T_{\mathrm{w},k}}(t)=(1-\rho_k)\delta\left(t\right)+\nu_k(1-\rho_k)\exp\{(-\mu_k+\nu_k)t\} \tag{19}$$

where $\delta\left(t\right)$ is the delta function which means that a user has zero wait with probability $\left(1-\rho_k\right)$. The t in (19) is substituted into the threshold in (9). Since $T_{\mathrm{sv},i,k}$ in threshold is a random variable, by substituting $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}\left[\frac{(1-\rho_k)s}{s-\nu_k+\nu_k\mathcal{L}_{T_{\mathrm{sv},k}}(s)}\right]$ in (10) into (19) and applying the pdf of $T_{\mathrm{sv},i,k}$, $p_{\mathrm{cp},i,k}$ is given by

$$p_{\mathrm{cp},i,k}=1-\left(\frac{\mu_k}{\nu_k}\right)^{D_i^{(c)}-1}\exp\{(-\mu_k+\nu_k)(T_{\mathrm{t}}-2T_{\mathrm{s}})\}. \tag{20}$$

Substituting $D_i^{(c)}$ into 1, (20) becomes (17). ■

Corollary 2: For the user type set $\mathcal{I} = \{1, 2\}$ and $D_i^{(c)} \in \{1, 2\}$, $p_{\mathrm{cp},i,k}$ is given by

$$p_{\mathrm{cp},i,k} = 1 - \rho_k + \frac{1-\rho_k}{\zeta_1 - \zeta_2} \times \left[\frac{(\zeta_1+\mu_k)^2}{\zeta_1} \left\{ \left(\frac{\mu_k}{\mu_k+\zeta_1} \right)^{D_i^{(c)}} \exp\{\zeta_1(T_\mathrm{t} - 2T_\mathrm{s})\} - 1 \right\} - \frac{(\zeta_2+\mu_k)^2}{\zeta_2} \left\{ \left(\frac{\mu_k}{\mu_k+\zeta_2} \right)^{D_i^{(c)}} \exp\{\zeta_2(T_\mathrm{t} - 2T_\mathrm{s})\} - 1 \right\} \right] \tag{21}$$

where ζ_1 and ζ_2 are

$$\zeta_1 = -\mu_k + \frac{\nu_k}{2} + \sqrt{\frac{\nu_k^2}{4} + \mu_k\left(\nu_k - \nu_{1,k}\right)} \tag{22}$$

$$\zeta_2 = -\mu_k + \frac{\nu_k}{2} - \sqrt{\frac{\nu_k^2}{4} + \mu_k\left(\nu_k - \nu_{1,k}\right)}. \tag{23}$$

Proof: For the user type set $\mathcal{I} = \{1, 2\}$ and $D_i^{(c)} \in \{1, 2\}$, $\mathcal{L}_{T_{\mathrm{sv},k}}(s) = \frac{\nu_{1,k}}{\nu_k}\frac{\mu_k}{s+\mu_k} + \frac{\nu_{2,k}}{\nu_k}\left(\frac{\mu_k}{s+\mu_k}\right)^2$, so $\mathcal{L}_{T_{\mathrm{w},k}}(s)$ is given by

$$\begin{aligned}\mathcal{L}_{T_{\mathrm{w},k}}(s) &= \frac{(1-\rho_k)\,s}{s - \nu_k + \nu_k\left(\frac{\nu_{1,k}}{\nu_k}\frac{\mu_k}{s+\mu_k} + \frac{\nu_{2,k}}{\nu_k}\left(\frac{\mu_k}{s+\mu_k}\right)^2\right)} \\ &= \frac{1-\rho_k}{\zeta_1-\zeta_2}\left\{\frac{(\zeta_1+\mu_k)^2}{s-\zeta_1} - \frac{(\zeta_2+\mu_k)^2}{s-\zeta_2} + \zeta_1 - \zeta_2\right\}\end{aligned} \tag{24}$$

where ζ_1 and ζ_2 are (22) and (23), and $\nu_k = \nu_{1,k} + \nu_{2,k}$. Since $\mathcal{L}^{-1}\{\frac{1}{s-\zeta_1}\} = \exp\{\zeta_1 t\}$, $f_{T_{\mathrm{w},k}}(t)$, i.e., $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}(s)$, is given by

$$f_{T_{\mathrm{w},k}}(t) = (1-\rho_k)\,\delta(t) + \frac{1-\rho_k}{\zeta_1-\zeta_2} \times \left[(\zeta_1+\mu_k)^2\exp\{\zeta_1 t\} - (\zeta_2+\mu_k)^2\exp\{\zeta_2 t\}\right]. \tag{25}$$

Substituting $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}\left[\frac{(1-\rho_k)s}{s-\nu_k+\nu_k\mathcal{L}_{T_{\mathrm{sv},k}}(s)}\right]$ in (10) into (25), $p_{\mathrm{cp},i,k}$ is given by

$$p_{\mathrm{cp},i,k} = \int_0^\infty \left[(1-\rho_k) + \frac{1-\rho_k}{\zeta_1-\zeta_2}\left\{\frac{(\zeta_1+\mu_k)^2}{\zeta_1}\left(\exp\{\zeta_1 t\} - 1\right) - \frac{(\zeta_2+\mu_k)^2}{\zeta_2}\left(\exp\{\zeta_2 t\} - 1\right)\right\}\right] \frac{\mu_k^{D_i^{(c)}} t^{D_i^{(c)}-1}\exp\left(-\mu_k r\right)}{\left(D_i^{(c)}-1\right)!} dr \tag{26}$$

which is the same as (21). ■

For other cases, if approximation is applied, it is available to describe the $p_{\mathrm{cp},i,k}$ in closed forms. There are some method to approximate the waiting time distributions such as [22], [23] and [24] or Laplace transform such as [25]. One of the simple approximation method is the Gamma approximation in [24], can be used to approximate for the waiting time distribution of the M/G/1 queue. The Gamma approximation can be applied to get $p_{\mathrm{cp},i,k}$ for some complicate cases such as $\mathcal{I} = \{1, 2, 3\}$, $D_i^{(c)} \in \{1, 2, 3\}$ in the following corollary.

Corollary 3: Considering the case of the user type set $\mathcal{I} = \{1, 2, 3\}$, $D_i^{(c)} \in \{1, 2, 3\}$, $\hat{p}_{\mathrm{cp},i,k}$ is described by

$$\hat{p}_{\mathrm{cp},i,k} = 1 - \rho_k \sum_{\substack{n=0 \\ n\in\mathbb{Z}}}^{\beta_{k,1}} \frac{\mu_k^{D_i^{(c)}} \left(T_\mathrm{t} - 2T_\mathrm{s}\right)^{n+D_i^{(c)}} \exp\left\{\frac{-T_\mathrm{t}+2T_\mathrm{s}}{\beta_{k,2}}\right\}}{\left(D_i^{(c)}-1\right)!\beta_{k,2}^n} \times \left[\frac{\Gamma\left(D_i^{(c)}\right)}{\Gamma\left(n+D_i^{(c)}+1\right)} {}_1F_1\left(D_i^{(c)}; n+D_i^{(c)}+1; \tau_k\right) + (-1)^n \exp\{\tau_k\}\, U\left(n+1, n+D_i^{(c)}+1, -\tau_k\right)\right] \tag{27}$$

where ${}_1F_1(\cdot;\cdot;\cdot)$ is the hypergeometric function, $U(\cdot,\cdot,\cdot)$ is the hypergeometric U function, and the τ_k is given by

$$\tau_k = \left(\frac{1}{\beta_{k,2}} - \mu_k\right)\left(T_\mathrm{t} - 2T_\mathrm{s}\right). \tag{28}$$

In (27), $\beta_{k,1}$ and $\beta_{k,2}$ are defined, respectively, as

$$\beta_{k,1} = \frac{\mathbb{E}\left\{T_{\mathrm{w},k}\right\}^2}{\mathbb{E}\left\{T_{\mathrm{w},k}^2\right\} - \mathbb{E}\left\{T_{\mathrm{w},k}\right\}^2} \tag{29}$$

$$\beta_{k,2} = \frac{\mathbb{E}\left\{T_{\mathrm{w},k}\right\}}{\beta_{k,1}} \tag{30}$$

where $\mathbb{E}\left\{T_{\mathrm{w},k}\right\}$ and $\mathbb{E}\left\{T_{\mathrm{w},k}^2\right\}$ are given, respectively, by

$$\mathbb{E}\left\{T_{\mathrm{w},k}\right\} = \frac{\nu_k\sum_{i\in\mathcal{I}}\left(\nu_{i,k}/\nu_k\right) D_i^{(c)}\left(D_i^{(c)}+1\right)}{2\mu_k^2\left(1-\nu_k\sum_i\left(\nu_{i,k}/\nu_k\right) D_i^{(c)}/\mu_k\right)} \tag{31}$$

$$\mathbb{E}\left\{T_{\mathrm{w},k}^2\right\} = 2\mathbb{E}\left\{T_{\mathrm{w},k}\right\}^2 + \frac{\nu_k\sum_{i\in\mathcal{I}}\left(\nu_{i,k}/\nu_k\right) D_i^{(c)}\left(D_i^{(c)}+1\right)\left(D_i^{(c)}+2\right)}{3\mu_k^3\left(1-\nu_k\sum_i\left(\nu_{i,k}/\nu_k\right) D_i^{(c)}/\mu_k\right)}. \tag{32}$$

Proof: See Appendix A. ■

Fig. 3 shows $p_{\mathrm{cp},i,k}$ of corollaries as a function of T_t for fixed $p_{\mathrm{o},i,k}$ considering the 2 tier MEC networks. For this figure, $\lambda_\mathrm{u} = 40$ nodes/km^2, $p_{\mathrm{m},k} = \{0.25, 0.75\}$, and other parameters in Table II are used. From Fig. 3, we can see the good match between $p_{\mathrm{cp},i,k}$ for corollary 3 and $p_{\mathrm{cp},i,k}$ obtained by the simulation for corollary 3. Hence, the results of corollary 3 can be used to get the numerical results.

B. Successful Communication Probability

In this subsection, we analyze the successful communication probability when the ith-type user communicates with the kth-tier MEC server, denoted by $p_{\mathrm{cm},i,k}$. We define $p_{\mathrm{cm},i,k}$ as

$$p_{\mathrm{cm},i,k} = \int_{y>0} S_{i,k}(y)\, f_{Y_{i,k}}(y)\, dy \tag{33}$$

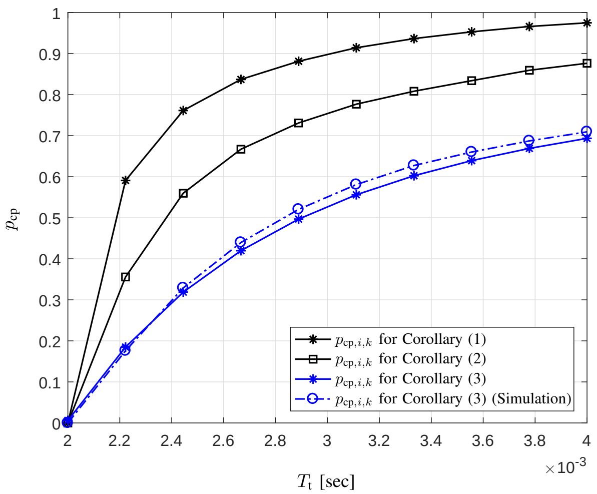

Fig. 3. The successful computation probabilities in 2 tier MEC network for fixed offloading probability.

where $f_{Y_{i,k}}(y)$ is the pdf of the distance between the ith-type user and the associated kth-tier MEC server, denoted by $Y_{i,k}$, given by [14, Lemma 4]

$$f_{Y_{i,k}}(y) = \frac{2\pi\lambda_{{\rm m},k}}{p_{{\rm o},i,k}} y \exp\left\{-\pi \sum_{j\in\mathcal{K}} \lambda_{{\rm m},j} \hat{\mathcal{W}}_{i,j}^{2/\alpha} y^2\right\}. \quad (34)$$

In (33), $S_{i,k}(y)$ is the successful transmission probability that uplink and downlink signal-to-interference-plus-noise ratios (SINRs) are larger than the thresholds when the ith-type user offloads to and receives from the kth-tier MEC server located at y. For the analytical tractability, some assumptions are used here to derive $S_{i,k}(y)$.

1) *Assumption 1*: The distribution of uplink interfering users follows the PPP.

The uplink interfering user distribution is not a PPP because locations of the users using same uplink channel is from the dependent thinning of the AP locations. However, according to [15], such effect can be weak. Hence, we use this assumption, as in other papers [15], [26], for analysis tractability.

2) *Assumption 2:* Uplink and downlink interference are independent.

Uplink and downlink interference are not independent because the locations of MEC servers and interfering users are dependent. Although some papers like [27] considered this dependence by a simplified method, it is still complicate to analyze the dependence. As the uplink analysis is not main objective of this work, we apply this assumption.[2]

Using those assumptions, $S_{i,k}(y)$ is presented by

$$S_{i,k}(y) = \mathbb{P}\left\{\text{SINR}_{i,k}^{(u)}(y) > \epsilon_i^{(u)}\right\} \mathbb{P}\left\{\text{SINR}_{i,k}^{(d)}(y) > \epsilon_i^{(d)}\right\} \quad (35)$$

TABLE II
PARAMETER VALUES IF NOT OTHERWISE SPECIFIED

Parameters	Values	Parameters	Values
$\lambda_{\rm m}$ [nodes/m^2]	$2*10^{-5}$	κ	0.7
$W^{(u)}$ [Hz]	$5*10^{6}$	$W^{(d)}$ [Hz]	$10*10^{6}$
$P_{{\rm u},i}$ [dBm]	23	$P_{{\rm m},1}$ [dBm]	45
$P_{{\rm m},2}$ [dBm]	30	α	4
N_0 [dBm]	-104	$U_{\rm s}$ [bit]	512
$T_{\rm t}$ [sec]	$3.5*10^{-3}$	$T_{\rm s}$ [sec]	10^{-3}
μ_1 [packet/slot]	18	μ_2 [packet/slot]	4

where $\text{SINR}_{i,k}^{(u)}(y)$ and $\text{SINR}_{i,k}^{(d)}(y)$ are the received SINRs in uplink and downlink, respectively, and $\epsilon_i^{(u)}$ and $\epsilon_i^{(d)}$ are SINR thresholds, given by

$$\epsilon_i^{(u)} = 2^{\frac{D_i^{(r)} U_{\rm s}}{W^{(u)} T_{\rm s}}} - 1, \quad \epsilon_i^{(d)} = 2^{\frac{D_i^{(d)} U_{\rm s}}{W^{(d)} T_{\rm s}}} - 1. \quad (36)$$

Here, U_s is the unit packet size, and $W^{(u)}$ and $W^{(d)}$ are the bandwidths for uplink and downlink, respectively. Using Theorem 1 in [15], uplink successful transmission probability is given by

$$\begin{aligned} &\mathbb{P}\left\{\text{SINR}_{i,k}^{(u)}(y) > \epsilon_i^{(u)}\right\} \\ &= \exp\left\{-\frac{\epsilon_i^{(u)}}{\frac{P_{{\rm u},i} y^{-\alpha}}{N_0}}\right\} \mathcal{L}_{I_{ik}^{(u)}}\left(y^\alpha P_{{\rm u},i}^{-1} \epsilon_i^{(u)}\right) \end{aligned} \quad (37)$$

where N_0 is the noise power and $I_{i,k}^{(u)}$ is the interference when the ith-type user offloads to the kth-tier MEC server. The $\mathcal{L}_{I_{i,k}^{(u)}}(s)$ in (37) is presented as [17]

$$\mathcal{L}_{I_{i,k}^{(u)}}(s) = \exp\left\{-2\pi\lambda_{{\rm m},k} \int_{z_{i,k}}^{\infty} \frac{x}{1 + (sP_{{\rm u},i})^{-1} x^\alpha} dx\right\} \quad (38)$$

where $z_{i,k}$ is the distance to the nearest kth-tier MEC server unassociated with the ith-type user. Since the nearest interfering user can be closer than the associated user, $z_{i,k}$ becomes zero. By replacing $s = y^\alpha \epsilon_i^{(u)} P_{{\rm u},i}^{-1}$ in (38) and $z_{i,k} = 0$, (37) is calculated by

$$\begin{aligned} &\mathbb{P}\left\{\text{SINR}_{i,k}^{(u)}(y) > \epsilon_i^{(u)}\right\} \\ &= \exp\left\{-\frac{\epsilon_i^{(u)}}{\frac{P_{{\rm u},i} y^{-\alpha}}{N_0}} - \pi\kappa\lambda_{{\rm m},k} y^2 Z\left(\epsilon_i^{(u)}, \alpha, 0\right)\right\} \\ &= \exp\left\{-\frac{\epsilon_i^{(u)}}{\frac{P_{{\rm u},i} y^{-\alpha}}{N_0}} - 2\pi\lambda_{{\rm m},k} \int_0^\infty \frac{x}{1 + \left(y^{-\alpha}/\epsilon_i^{(u)}\right) x^\alpha} dx\right\} \end{aligned} \quad (39)$$

where $Z(a,b,c) = (a)^{2/b} \int_{(c/a)^{2/b}}^{\infty} \frac{1}{1+u^{b/2}} du$. From the analysis in [14] and [17], the downlink successful transmission probability is given by

$$\begin{aligned} &\mathbb{P}\left\{\text{SINR}_{i,k}^{(d)}(y) > \epsilon_i^{(d)}\right\} \\ &= \exp\left\{-\frac{\epsilon_i^{(u)}}{\frac{P_{{\rm u},i} y^{-\alpha}}{N_0}}\right\} \prod_{j\in\mathcal{K}} \mathcal{L}_{I_{i,j}^{(d)}}\left(y^\alpha P_{{\rm m},k}^{-1} \epsilon_i^{(d)}\right) \end{aligned} \quad (40)$$

[2] There is a recent results in [28], which provides the uplink performance with less assumptions. However, we note that applying more accurate analysis does not change our framework.

where $I_{i,j}^{(d)}$ is the interference when the ith-type user receives from the jth-tier MEC servers. In (40), $\mathcal{L}_{I_{i,j}^{(d)}}(s)$ is given by substituting $P_{\mathrm{u},i}$ and $z_{i,k}$ in (38) into $P_{\mathrm{m},j}$ and $\hat{\mathcal{W}}_{i,j}^{1/\alpha}y$, respectively. By replacing $s = y^{\alpha}\epsilon_i^{(d)}P_{\mathrm{m},k}^{-1}$, (40) is represented by

$$\begin{aligned}&\mathbb{P}\left\{\mathrm{SINR}_{i,k}^{(d)}(y) > \epsilon_i^{(d)}\right\}\\ &= \exp\left\{-\frac{\epsilon_i^{(d)}}{\frac{P_{\mathrm{m},k}y^{-\alpha}}{N_0}} - \pi\sum_{j\in\mathcal{K}} y^2 \hat{P}_{\mathrm{m},j}^{2/\alpha}\lambda_{\mathrm{m},j} Z\left(\epsilon_i^{(d)}, \alpha, \hat{B}_{i,j}\right)\right\}\\ &= \exp\left\{-\frac{\epsilon_i^{(d)}}{\frac{P_{\mathrm{m},k}y^{-\alpha}}{N_0}} - \sum_{j\in\mathcal{K}} 2\pi\lambda_{\mathrm{m},j}\int_{z_{i,j}}^{\infty}\frac{x}{1+\left(\hat{P}_{\mathrm{m},j}^{-1}y^{-\alpha}/\epsilon_i^{(d)}\right)x^{\alpha}}dx\right\}\end{aligned} \tag{41}$$

where $\hat{P}_{\mathrm{m},j}$ is $P_{\mathrm{m},j}/P_{\mathrm{m},k}$ and $\hat{B}_{i,j}$ is $B_{i,j}/B_{i,k}$. Substituting (39), (41), and (34) into (33), we can obtain $p_{\mathrm{cm},i,k}$. For the path loss factor $\alpha = 4$ is given by

$$p_{\mathrm{cm},i,k} \overset{(\mathrm{a})}{=} \frac{\pi\lambda_{\mathrm{m},k}\sqrt{\pi}\exp\left\{\frac{\pi^2\eta_2^2}{4\eta_1}\right\}\mathsf{Erfc}\left(\frac{\pi\eta_2}{2\sqrt{\eta_1}}\right)}{2\sqrt{\eta_1}p_{\mathrm{o},i,k}} \tag{42}$$

where η_1 and η_2 are given, respectively, by

$$\begin{aligned}\eta_1 &= \frac{\epsilon_i^{(d)}N_0}{P_{\mathrm{m},k}} + \frac{\epsilon_i^{(u)}N_0}{P_{\mathrm{u},i}}\\ \eta_2 &= \lambda_{\mathrm{m},k}Z\left(\epsilon_i^{(u)}, \alpha, 0\right)\\ &\quad + \sum_{j\in\mathcal{K}}\lambda_{\mathrm{m},j}\left\{\hat{P}_{\mathrm{m},j}^{2/\alpha}Z\left(\epsilon_i^{(d)}, \alpha, \hat{B}_{i,j}\right) + \hat{\mathcal{W}}_{i,j}^{2/\alpha}\right\}.\end{aligned} \tag{43}$$

Here, (a) is from the [29, eq. (2.33)]. Substituting (4), (33), and (10) into (8), p_{s} can be derived. From (8), we can see that deriving the optimal bias factors in terms of p_{s} is hard due to the complex structure. However, we show the existence of the optimal bias factor using numerical results in Section IV.

IV. Numerical Results

In this section, we provide numerical results on the successful edge computing probability and the mean latency for the 2 tier MEC networks. The MEC servers in 1st tier have higher computing capabilities than those in 2nd tier. The $D_i^{(r)}$ and $D_i^{(c)}$ are set to be equal for simplicity, while $D_i^{(d)}$ is 2 times bigger than $D_i^{(c)}$ because the downlink data rate is generally bigger than uplink data rate. The values of network parameters used for simulations are presented in Table II.

A. Successful Edge Computing Probability

In this subsection, we show how the bias factors in MEC server association affect p_{s} via numerical results.

Fig. 4 shows p_{s}, p_{cp} and p_{cm} as a function of the bias factor $B_{1,2}$. For this figure, $D_i^{(c)} = 1$, $p_{\mathrm{m},k} = \{0.125, 0.875\}$, $\lambda_{\mathrm{u}} = 6.2 * 10^{-5}$ nodes/m^2, and $B_{1,1} = 10$dB are used, and increasing $B_{1,2}$ (i.e., x-axis in Fig. 4) means more users offload their required computation to 2nd-tier MEC servers than 1st-tier MEC servers. From Fig. 4, we first see that the simulation result of p_{cp} matches well with our analysis, but that of p_{cm} does not due to the assumptions used for analytical tractability. Consequently, simulation results of p_{s} are different to the analysis, but we see that the trends according to the bias factor are the same. From Fig. 4, we also see that when $B_{1,2}$ is small, both p_{cp} and p_{cm} are small because the 1st-tier MEC servers are heavy-loaded (which lowers p_{cp}), and the communication links between a user and the MEC server are long (which lowers p_{cm}) due to the lower MEC server density in 1st-tier $\lambda_{\mathrm{m},1}$ than that of 2nd-tier servers $\lambda_{\mathrm{m},2}$. As $B_{1,2}$ increases, all probabilities increase first because the computation tasks are starting to be offloaded to the 2nd-tier MEC servers, which can be located closer to the users and less loaded. However, after certain points of $B_{1,2}$, p_{cp} decreases since the 2nd-tier MEC servers becomes heavy-loaded, and p_{cm} also decreases due to longer communication link. Correspondingly, p_{s} also increases

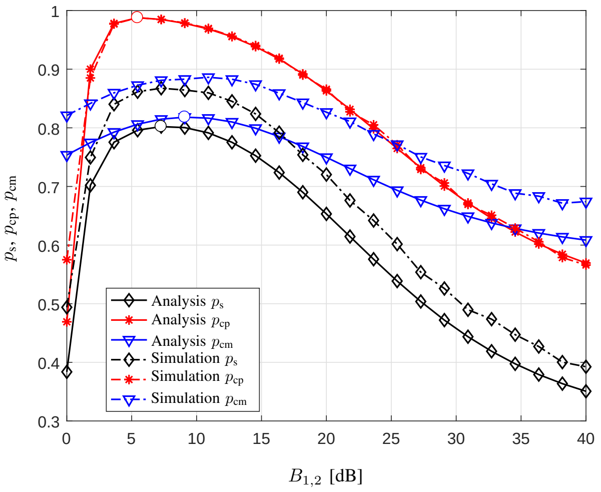

Fig. 4. Successful edge computing probability p_{s}, successful communication probability p_{cm}, and successful computation probability p_{cp} of 2 tier MEC-enabled HetNet with 1 type users as a function of bias factor for offloading to 2nd-tier MEC servers $B_{1,2}$.

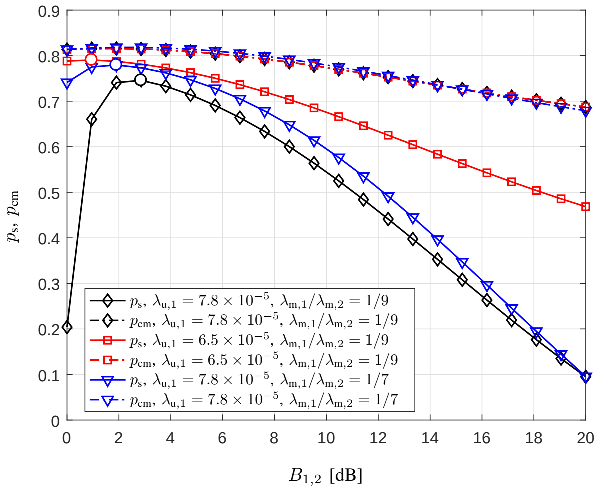

Fig. 5. Successful edge computing probability p_{s} and successful communication probability p_{cm} of 2 tier MEC-enabled HetNet with 1 type users as a function of bias factor for offloading to 2nd-tier MEC servers $B_{1,2}$ for different values of user density $\lambda_{\mathrm{u},1}$ and MEC server density ratio $\lambda_{\mathrm{m},1}/\lambda_{\mathrm{m},2}$.

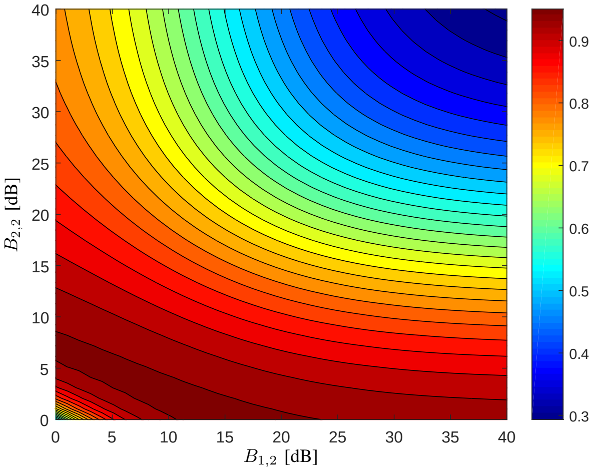

Fig. 6. Successful communication probability $p_{\rm cm}$ of 2 tier MEC-enabled HetNet with 2 type users as functions of bias factors for offloading to 2nd-tier MEC servers $B_{1,2}$ and $B_{2,2}$.

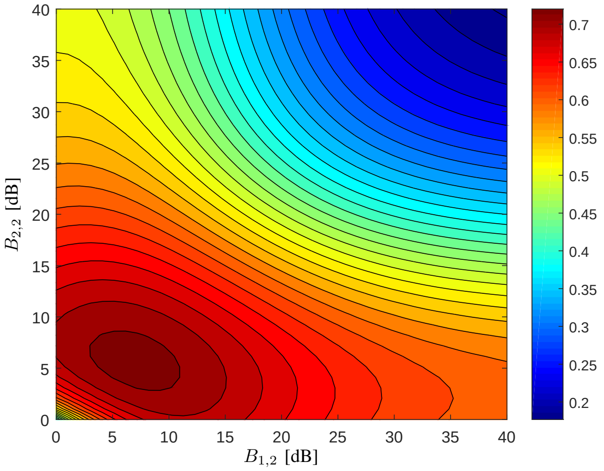

Fig. 7. Successful edge computing probability $p_{\rm s}$ of 2 tier MEC-enabled HetNet with 2 type users as functions of bias factors for offloading to 2nd-tier MEC servers $B_{1,2}$ and $B_{2,2}$.

up to a certain point of $B_{1,2}$, and decrease. Therefore, there exists the optimal bias factors $B^*_{1,2}$ (marked by circles in Fig. 4) in terms of $p_{\rm cp}$, $p_{\rm cm}$, and $p_{\rm s}$, which are all different.

In general HetNets, the optimal bias factor is determined in terms of communication performance like the successful communication probability $p_{\rm cm}$. However, when we consider the computation capability and the amount of computation tasks, the optimal bias factor becomes different as shown in Fig. 4. In this figure, $B^*_{1,2}$ for $p_{\rm s}$ is smaller than $B^*_{1,2}$ for $p_{\rm cm}$ since the computation performance changes more sensitively according to the amount of offloaded tasks (i.e., the bias factor) than the communication performance.

Fig. 5 shows $p_{\rm s}$ and $p_{\rm cm}$ as a function of $B_{1,2}$ for different $\lambda_{{\rm u},1}$, the MEC server density ratio $\lambda_{{\rm m},1}/\lambda_{{\rm m},2}$, when $B_{1,1} =$ 3dB under the same environment of Fig 4. From Fig. 5, we can see that $B^*_{1,2}$ decreases as $\lambda_{\rm u}$ decreases. This is because the offloaded tasks to the 1st-tier MEC servers are not large even for small $B_{1,2}$ due to the low $\lambda_{\rm u}$. Hence, offloading the users' tasks to the 1st-tier servers improves $p_{\rm s}$ and results in decreasing $B^*_{1,2}$. When $\lambda_{{\rm m},1}/\lambda_{{\rm m},2}$ increases, we have more 1st-tier MEC servers in the network, so the users who offload to the 1st-tier MEC servers can have higher probabilities for both $p_{\rm cm}$ and $p_{\rm s}$. Hence, for large $\lambda_{{\rm m},1}/\lambda_{{\rm m},2}$, it is better to associate more to the 1st-tier, so $B^*_{1,2}$ decreases as shown in Fig 4.

Fig. 6 and Fig. 7 show the contour of $p_{\rm s}$ and $p_{\rm cp}$, respectively, having 2 types of users $\mathcal{I} = \{1, 2\}$ with $p_{{\rm m},k} = \{0.125, 0.875\}$, $\lambda_{\rm u} = 4.4 * 10^{-5}$nodes/m^2, $p_{{\rm u},i} = \{0.5, 0.5\}$, and $D_i^{(c)} = \{1, 2\}$ as a function of $B_{1,2}$ and $B_{2,2}$. Both $B_{1,1}$ and $B_{2,1}$ are set to be 10dB. First, from Fig. 6, we can see that $p_{\rm cp}$ decreases more by $B_{2,2}$ than by $B_{1,2}$. This is because the variation of service time for large size tasks (i.e., $D_2^{(c)}$) is bigger than that for small size tasks (i.e., $D_1^{(c)}$), so $p_{\rm cp}$ becomes more sensitive by the arrival of $D_2^{(c)}$ size tasks.

By comparing Figs. 6 and 7, we can see that the optimal bias factors for $p_{\rm s}$ and $p_{\rm cm}$ are different. First, from Fig. 6, we can see that even when we offload all tasks of type 1 user or type 2 user to the 1st-tier MEC servers and no task to the 2nd-tier MEC servers, i.e., $B^*_{1,2} = 0$ or $B^*_{2,2} = 0$, we can achieve the best performance in terms of successful computation probability $p_{\rm cp}$. However, it becomes different when we consider the successful edge computation probability as a performance metric. From Fig. 7, we can see that offloading certain amount of tasks to 1st-tier MEC servers and 2nd-tier MEC servers, i.e., $B^*_{1,2} > 0$ or $B^*_{2,2} > 0$, can achieve the best performance in terms of $p_{\rm s}$. This is because in $p_{\rm s}$, the communication performance is also considered, which can achieve low performance due to the longer link distance when all users associate to certain tier MEC servers. Therefore, the communication and computation performance needs to be considered together when we determine the bias factors for MEC server association, which can be also seen for the case with three types of users in Figs. 8 and 9.

Fig. 8 and Fig. 9 show the contour of $p_{\rm s}$ and $p_{\rm cp}$, respectively, having 3 types of users $\mathcal{I} = \{1, 2, 3\}$ with $p_{{\rm m},k} = \{0.125, 0.875\}$, $\lambda_{\rm u} = 3.5 * 10^{-5}$nodes/m^2, $p_{{\rm u},i} = \{0.4, 0.2, 0.4\}$, and $D_i^{(c)} = \{1, 2, 3\}$ as a function of the ratio of the bias $B_{1,2}/B_{2,2}$ and $B_{3,2}/B_{2,2}$. Both $B_{1,2}$ and $B_{2,2}$ and both $B_{1,1}$ and $B_{3,1}$ are set to be 10dB. From Fig. 8, we can also see that $p_{\rm cp}$ decreases faster by $B_{3,2}$ than $B_{1,2}$ due to the large task size of 3rd-type users, and by comparing Figs. 8 and 9, the optimal bias factors for $p_{\rm s}$ becomes different to the one for $p_{\rm cp}$.

B. Mean Latency

In this subsection, we present how the bias factors in MEC server association affect the mean latency in the MEC-enabled HetNet, and compare the optimal bias factors in terms of the mean latency to the ones in terms of successful edge computing probability. For this, we first analyze the mean latency as follows. Assuming that the communication for transmitting and receiving the computation request and result

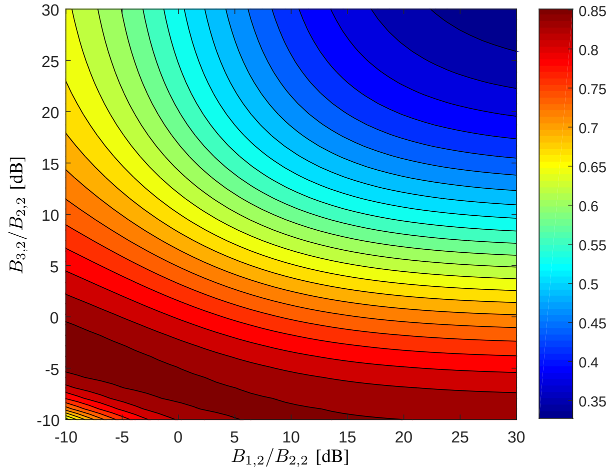

Fig. 8. Successful computation probability p_{cp} of 2 tier MEC-enabled HetNet with 3 type users as functions of the ratios of the bias $B_{1,2}/B_{2,2}$ and $B_{3,2}/B_{2,2}$.

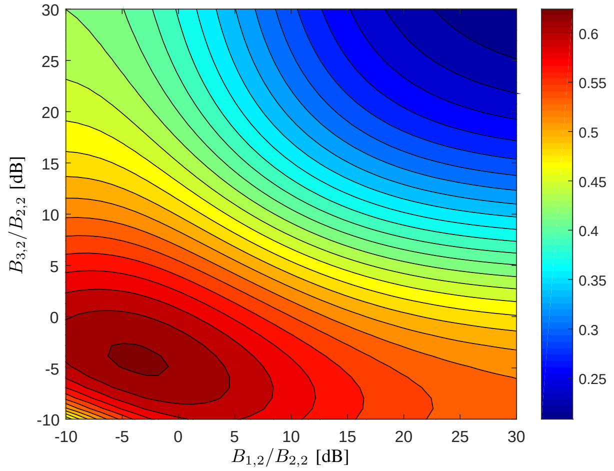

Fig. 9. Successful edge computing probability p_{s} of 2 tier MEC-enabled HetNet with 3 type users as functions of the ratios of the bias $B_{1,2}/B_{2,2}$ and the $B_{3,2}/B_{2,2}$.

is always successful (i.e., no retransmission), the mean latency for the ith-type user denoted by $\mathbb{E}\left\{T_{\text{total},i}\right\}$, is given by

$$\mathbb{E}\left\{T_{\text{total},i}\right\} = \sum_{k\in\mathcal{K}} p_{\text{o},i,k}\left[\mathbb{E}\left\{T_{\text{c},i,k}\right\}+2T_{\text{s}}\right] \tag{44}$$

where $\mathbb{E}\left\{T_{\text{c},i,k}\right\}$ is the mean computing time. For the arrival rate $\nu_{i,k}$, the service rate μ_k, and the computation task size $D_i^{(c)}$, $\mathbb{E}\left\{T_{\text{c},i,k}\right\}$ is obtained by applying the mean waiting time and mean service time for M/G/1 queue [21], given by

$$\begin{aligned}\mathbb{E}\left\{T_{\text{c},i,k}\right\} &= \mathbb{E}\left\{T_{\text{w},k}\right\}+\mathbb{E}\left\{T_{\text{sv},i,k}\right\}\\ &= \frac{\sum_{i\in\mathcal{I}}\nu_{i,k}\frac{D_i^{(c)}\left(D_i^{(c)}+1\right)}{\mu_k^2}}{2\left(1-\sum_{i\in\mathcal{I}}\nu_{i,k}\frac{D_i^{(c)}}{\mu_k}\right)}+\frac{D_i^{(c)}}{\mu_k}.\end{aligned} \tag{45}$$

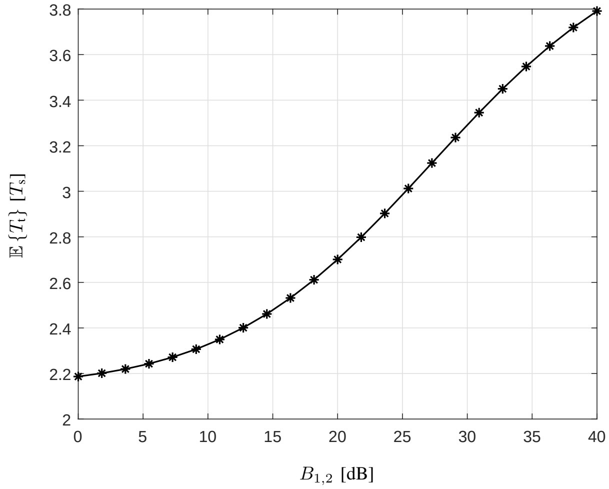

Fig. 10. Mean latency $\mathbb{E}\left\{T_{\text{total}}\right\}$ of 2 tier MEC-enabled HetNet with 1 type users as a function of bias factor for offloading to 2nd-tier MEC servers $B_{1,2}$.

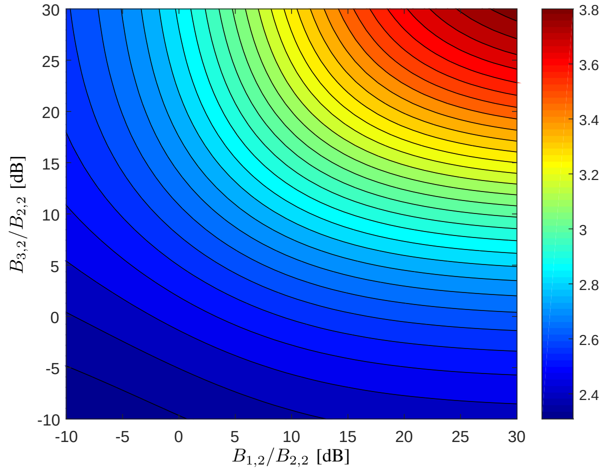

Fig. 11. Mean latency $\mathbb{E}\left\{T_{\text{total}}\right\}$ of 2 tier MEC-enabled HetNet with 3 type users as functions of the ratio of the bias $B_{1,2}/B_{2,2}$ and the $B_{3,2}/B_{2,2}$.

Using the law of total probability, (44), and (45), the mean latency $\mathbb{E}\left\{T_{\text{total}}\right\}$ is given by

$$\mathbb{E}\left\{T_{\text{total}}\right\} = \sum_{i\in\mathcal{I}} p_{\text{u},i}\mathbb{E}\left\{T_{\text{total},i}\right\}. \tag{46}$$

From (46), we show the effect of offloading to the MEC server and how much the latency is reduced in following figures. Since the derivation of the optimal bias factors in terms of $\mathbb{E}\left\{T_{\text{total}}\right\}$ is also hard, we show the existence of the optimal bias factor in Fig. 10.

Fig. 10 displays the mean latency $\mathbb{E}\left\{T_{\text{total}}\right\}$ as a function of $B_{1,2}$. For this figure, the same network parameters are used as the ones used in Fig. 4. From Fig. 10, it can be seen that as $B_{1,2}$ increases, $\mathbb{E}\left\{T_{\text{total}}\right\}$ also increases because more computation tasks are offloaded to the 2nd-tier MEC server having lower computing capabilities than 1st-tier server. Hence, the optimal bias factor is $B_{1,2}^{*}=0$, which means in terms of the mean latency, offloading more users to the 1st-tier MEC server shows better performance. When we compare the results in Fig. 10 to the results of p_{s} in Fig. 4, we can see that optimal bias factors are different. Specifically, the optimal bias factor in terms of successful edge computing probability in Fig. 4 is bigger than the one in terms of mean latency in Fig. 10. This is because the mean latency is less affected by the heavy-loaded cases at 1st-tier MEC server. Therefore, we can see that analyzing and designing the MEC-enabled HetNet using the successful edge computing probability can be more effective to see the system reliability than using the mean latency. We can also see the effectiveness of the successful edge computing probability in Fig. 11.

Fig. 11 presents the contour of $\mathbb{E}\left\{T_{\text{total}}\right\}$ as a function of the ratio of the bias $B_{1,2}/B_{2,2}$ and $B_{3,2}/B_{2,2}$. For this figure, the same parameters of Fig. 9 are used. From Fig. 11, we can see that as $B_{1,2}/B_{2,2}$ and $B_{3,2}/B_{2,2}$ increase, $\mathbb{E}\left\{T_{\text{total}}\right\}$ also increases due to more offloading to the 2nd-tier MEC servers. When comparing Figs. 9 and 11, we can see that the bias factors minimizing the $\mathbb{E}\left\{T_{\text{total}}\right\}$ are located at zero, i.e., $B_{1,2}^{*} = 0$, while the bias factors maximizing p_{s} are located in a specific point.

V. Conclusions

In this paper, we propose the MEC-enabled HetNet, composed of the users with different computation task sizes and the MEC servers with different computing capacities. We derive the successful edge computing probability by analyzing the latency distribution and the communication reliability using stochastic geometry and queueing theory. We then evaluate the effects of bias factors in association and network parameters on the successful edge computing probability. Our results show that the MEC-enabled HetNet has different optimal association bias factors from conventional ones, which do not consider the computing performance. Specifically, the optimal bias factor for successful edge computing probability is generally bigger than that for the conventional successful communication probability, and the optimal bias factor for offloading to the high-capable MEC servers (i.e., 2nd tier) decreases with the density of mobile users. In addition, we also analyze the mean computing and communication latency for MEC-enabled HetNet and shows that the optimal bias factors for the mean latency are different to the ones for successful edge computing probability. The outcomes of this work provide the design insights for the optimal configuration of the MEC-enabled wireless networks.

Appendix

A. Proof of Corollary 3

The Gamma Approximation used in [24] is derived from the relationship between the Erlang distribution $\text{Erlang}\left(\beta_1, \beta_2\right)$ and the Gamma distribution $\Gamma\left[\beta_1, \beta_2\right]$. The pdf of Gamma distribution, denoted by $f_{\Gamma}\left(t\right)$, is given by

$$f_{\Gamma}\left(t\right) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{{\beta_2}^{-\beta_1} t^{\beta_1 - 1} \exp\{-t/\beta_2\}}{\Gamma(\beta_1)} & \text{if } t > 0 \end{cases} \tag{47}$$

where β_1 and β_2 are the shape parameter and scale parameter of the Gamma distribution, respectively. According to the properties of Gamma distribution, if β_1 is a positive integer, the pdf of Erlang distribution is same as the pdf of Gamma distribution (47). It means that the cdf of the Gamma distribution, denoted by $F_{\Gamma}\left(t\right)$, is also same as the cdf of the Erlang distribution $\text{Erlang}\left(\beta_1, \beta_2\right)$ under the condition that β_1 is a positive integer. The cdf of the Erlang distribution $F_{\Gamma}\left(t\right)$ is described by

$$F_{\Gamma}\left(t\right) = \begin{cases} 0 & \text{if } t < 0 \\ 1 - \exp\left\{-t/\beta_{k,2}\right\} \sum_{\substack{0 \le c < \beta_{k,1} \\ c \in \mathbb{Z}}} \frac{\left(t/\beta_{k,2}\right)^c}{c!} & \text{else.} \end{cases} \tag{48}$$

The mean and the variance of $\Gamma\left[\beta_1, \beta_2\right]$ are given by

$$\mathbb{E}\left\{T\right\} = \beta_1 \beta_2 \tag{49}$$

$$\mathbb{V}\left\{T\right\} = \beta_1 \beta_2^2 \tag{50}$$

where $\mathbb{E}\left\{T\right\}$ and $\mathbb{V}\left\{T\right\}$ are the mean and the variance of T, respectively. From the (49) and (50), we can obtain the β_1 and β_2 as

$$\beta_1 = \frac{\mathbb{E}\left\{T\right\}^2}{\mathbb{V}\left\{T\right\}} = \frac{\mathbb{E}\left\{T\right\}^2}{\mathbb{E}\left\{T^2\right\} - \mathbb{E}\left\{T\right\}^2} \tag{51}$$

$$\beta_2 = \frac{\mathbb{E}\left\{T\right\}}{\beta_1}. \tag{52}$$

Therefore, if a mean and variance of distribution are given, the Gamma approximation can be used to approximate a distribution.

In the corollary 3, the mean waiting time and the variance of the waiting time are obtained by using the Takacs Recursion Formula in [21]. Thus, the mean waiting time and the mean square of waiting time for kth tier MEC server is obtained by

$$\mathbb{E}\left\{T_{\mathrm{w},k}\right\} = \frac{\sum_{i \in \mathcal{I}} \nu_k \mathbb{E}\left\{T_{\mathrm{sv},i,k}^2\right\}}{2\left(1 - \rho_k\right)} \tag{53}$$

$$\mathbb{E}\left\{T_{\mathrm{w},k}^2\right\} = 2\mathbb{E}\left\{T_{\mathrm{w},k}\right\}^2 + \frac{\sum_{i \in \mathcal{I}} \nu_k \mathbb{E}\left\{T_{\mathrm{sv},i,k}^3\right\}}{3\left(1 - \rho_k\right)} \tag{54}$$

where $\mathbb{E}\left\{T_{\mathrm{sv},i,k}^2\right\}$ and $\mathbb{E}\left\{T_{\mathrm{sv},i,k}^3\right\}$ is defined by

$$\mathbb{E}\left\{T_{\mathrm{sv},i,k}^2\right\} = \int_0^\infty t^2 f_{T_{\mathrm{sv},i,k}}\left(t\right) dt = \frac{D_i^{(c)}\left(D_i^{(c)} + 1\right)}{\mu_k^2} \tag{55}$$

$$\begin{aligned} \mathbb{E}\left\{T_{\mathrm{sv},i,k}^3\right\} &= \int_0^\infty t^3 f_{T_{\mathrm{sv},i,k}}\left(t\right) dt \\ &= \frac{D_i^{(c)}\left(D_i^{(c)} + 1\right)\left(D_i^{(c)} + 2\right)}{\mu_k^3}. \end{aligned} \tag{56}$$

Using (53), (54), (55), and (56), we can present the (29) and (30). By substituting the (29) and (30) into (48), the approximated distribution of waiting time for tasks, which are not immediately served upon arrival, is obtained. The approximated waiting time distribution for overall tasks is presented by

$$F\left(t\right) = 1 - \rho_k + \rho_k F_{\Gamma}\left(t\right). \tag{57}$$

Substituting $\mathcal{L}^{-1}_{T_{\mathrm{w},k}}\left[\frac{(1-\rho_k)s}{s-\nu_k+\nu_k\mathcal{L}_{T_{\mathrm{sv},k}}(s)}\right]$ in (10) into (57), $p_{\mathrm{cp},i,k}$ is given by

$$\begin{aligned} p_{\mathrm{cp},i,k} =& \int_0^\infty \left[1 - \rho_k + \rho_k F\left(T_{\mathrm{t}} - 2T_{\mathrm{s}} - r\right)\right] f_{T_{\mathrm{sv},i,k}}\left(r\right) dr \\ =& 1 - \rho_k \sum_{\substack{0 \le n < \beta_{k,1} \\ n \in \mathbb{Z}}} \frac{\mu_k^{D_i^{(c)}} \exp\{\frac{-T_{\mathrm{t}}+2T_{\mathrm{s}}}{\beta_{k,2}}\}}{c!\left(D_i^{(c)} - 1\right)!\beta_{k,2}^n} \\ &\times\int_0^\infty \exp\left\{\left(\frac{1}{\beta_{k,2}} - \mu_k\right) r\right\} r^{D_i^{(c)}-1} \left(T_{\mathrm{t}} - 2T_{\mathrm{s}} - r\right)^n dr. \end{aligned} \tag{58}$$

By substituting the $\frac{1}{T_{\mathrm{t}}-2T_{\mathrm{s}}}$ for γ_1 and dividing the interval of integrals into subintervals, (58) is presented by

$$\begin{aligned} p_{\mathrm{cp},i,k} =& 1 - \rho_k \sum_{\substack{0 \le n < \beta_{k,1} \\ n \in \mathbb{Z}}} \frac{\mu_k^{D_i^{(c)}} \left(T_{\mathrm{t}} - 2T_{\mathrm{s}}\right)^{n+D_i^{(c)}} \exp\{\frac{-T_{\mathrm{t}}+2T_{\mathrm{s}}}{\beta_{k,2}}\}}{n!\left(D_i^{(c)} - 1\right)!\beta_{k,2}^n} \\ &\times \left[\int_0^1 F_{\gamma_1} d\gamma_1 + \int_1^\infty F_{\gamma_1} d\gamma_1\right] \end{aligned} \tag{59}$$

where F_{γ_1} is given by

$$\begin{aligned} F_{\gamma_1} =& \exp\left\{\left(\frac{1}{\beta_{k,2}} - \mu_k\right)\left(T_{\mathrm{t}} - 2T_{\mathrm{s}}\right)\gamma_1\right\} \\ &\times \gamma_1^{D_i^{(c)}-1}\left(1 - \gamma_1\right)^n. \end{aligned} \tag{60}$$

The former integration of F_{γ_1} is transformed by the hypergeometric function ${}_1F_1(\cdot;\cdot;\cdot)$. Substituting the $\gamma_1 + 1$ for γ_2, the latter integration of F_{γ_2} is changed and (61) is presented by

$$\begin{aligned} p_{\mathrm{cp},i,k} =& 1 - \rho_k \sum_{\substack{0 \le n < \beta_{k,1} \\ n \in \mathbb{Z}}} \frac{\mu_k^{D_i^{(c)}} \left(T_{\mathrm{t}} - 2T_{\mathrm{s}}\right)^{n+D_i^{(c)}} \exp\{\frac{-T_{\mathrm{t}}+2T_{\mathrm{s}}}{\beta_{k,2}}\}}{n!\left(D_i^{(c)} - 1\right)!\beta_{k,2}^n} \\ &\times \left[\int_0^1 F_{\gamma_1} d\gamma_1 + (-1)^n \int_0^\infty F_{\gamma_2} d\gamma_2\right] \end{aligned} \tag{61}$$

where F_{γ_2} is given by

$$\begin{aligned} F_{\gamma_2} =& \exp\left\{\left(\frac{1}{\beta_{k,2}} - \mu_k\right)\left(T_{\mathrm{t}} - 2T_{\mathrm{s}}\right)\left(\gamma_2 + 1\right)\right\} \\ &\times \left(\gamma_2 + 1\right)^{D_i^{(c)}-1}\gamma_2^n. \end{aligned} \tag{62}$$

Since the integration of F_{γ_2} is transformed by the hypergeometric U function by modifying the exponential term, (61) becomes (27).

References

[1] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, Mar. 2017.

[2] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges," *IEEE Commun. Lett.*, vol. 55, no. 4, pp. 54–61, Apr. 2017.

[3] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Aug. 2017.

[4] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

[5] X. Tao, K. Ota, M. Dong, H. Qi, and K. Li, "Performance guaranteed computation offloading for mobile-edge cloud computing," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 774–777, Dec. 2017.

[6] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. Quek, "Offloading in mobile edge computing: Task allocation and computational frequency scaling," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3571–3584, Apr. 2017.

[7] Y. Mao, J. Zhang, S. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *Proc. IEEE Global Telecomm. Conf.*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[8] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.

[9] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," in *Proc. IEEE Int. Symp. on Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 1451–1455.

[10] T. Zhao, S. Zhou, X. Guo, and Z. Niu, "Tasks scheduling and resource allocation in heterogeneous cloud for delay-bounded mobile edge computing," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–7.

[11] S.-W. Ko, K. Han, and K. Huang, "Wireless networks for mobile edge computing: Spatial modeling and latency analysis (extended version)." [Online]. Available: http://arxiv.org/abs/1709.01702

[12] Y.-H. Kao, B. Krishnamachari, M.-R. Ra, and F. Bai, "Hermes: Latency optimal task assignment for resource-constrained mobile computing," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3056–3069, Mar. 2017.

[13] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Mar. 2012.

[14] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[15] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2669–2679, Jun. 2013.

[16] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in hetnets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, Oct. 2015.

[17] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[18] Q. Zhang, H. H. Yang, T. Q. Quek, and J. Lee, "Heterogeneous cellular networks with LoS and NLoS transmissions–the role of massive MIMO and small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7996–8010, Dec. 2017.

[19] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks, Volume I — Theory*, ser. Foundations and Trends in Networking. NoW Publishers, 2009.

[20] H. Cho, C. Liu, J. Lee, T. Noh, and T. Q. Quek, "Impact of elevated base stations on the ultra-dense networks," *IEEE Commun. Lett.*, to appear.

[21] L. Kleinrock, *Queueing Systems, Volume 1: Theory*. New York: Wiley-Interscience, 1975.

[22] Y. Jiang, C.-K. Tham, and C.-C. Ko, "An approximation for waiting time tail probabilities in multiclass systems," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 175–177, Apr. 2001.

[23] M. Ackroyd, "Computing the waiting time distribution for the G/G/1 queue by signal processing methods," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 52–58, Jan. 1980.

[24] M. Menth, R. Henjes, C. Zepfel, and P. Tran-Gia, "Gamma-approximation for the waiting time distribution function of the M/G/1 queue," in *2nd Conference on Next Generation Internet Networks Traffic Engineering (NGI)*, Valencia, Spain, Jan. 2006.

[25] H. Stehfest, "Algorithm 368: Numerical inversion of laplace transforms [D5]," *Communications of ACM*, vol. 13, no. 1, pp. 47–49, Jan. 1970.

[26] H. ElSawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4454–4469, Aug. 2014.

[27] J. Lee and T. Q. Quek, “Hybrid full-/half-duplex system analysis in heterogeneous wireless networks,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2883–2895, May 2015.

[28] W. U. Mondal and G. Das, “Uplink user process in poisson cellular network,” *IEEE Wireless Commun. Lett.*, vol. 21, no. 9, pp. 2013–2016, Sep. 2017.

[29] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA: Academic Press, Inc., 2007.