Computation Offloading and Service Caching in Heterogeneous MEC Wireless Networks

Thesis by

Yongqiang Zhang

In Partial Fulfillment of the Requirements

For the Degree of

Masters of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

March, 2021

EXAMINATION COMMITTEE PAGE

The thesis of Yongqiang Zhang is approved by the examination committee

Committee Chairperson: Prof. Mohamed-Slim Alouini
Committee Members: Prof. Basem Shihada, Prof. Xiangliang Zhang, Dr. Abla Kammoun

ABSTRACT

Computation Offloading and Service Caching in Heterogeneous MEC Wireless Networks

Yongqiang Zhang

Abstract—Mobile edge computing (MEC) can dramatically promote the computation capability and prolong the lifetime of mobile users by offloading computation-intensive tasks to edge cloud. In this thesis, a spatial-random two-tier heterogeneous network (HetNet) is modelled to feature random node distribution, where the small-cell base stations (SBSs) and the macro base stations (MBSs) are cascaded with resource-limited servers and resource-unlimited servers, respectively. Only a certain type of application services and finite number of offloaded tasks can be cached and processed in the resource-limited edge server. For that setup, we investigate the performance of two offloading strategies corresponding to integrated access and backhaul (IAB)-enabled MEC networks and traditional cellular MEC networks. By using tools from stochastic geometry and queuing theory, we derive the average delay for the two different strategies, in order to better understand the influence of IAB on MEC networks. Simulations results are provided to verify the derived expressions and to reveal various system-level insights.

ACKNOWLEDGEMENTS

I adore and cherish my two-year master's journey at KAUST very much, filled with various experiences.

I would like to express my heartfelt appreciation to my graduate study supervisor: Prof. Mohamed-Slim Alouini, for approving my M.S/Ph.D. application at KAUST. I have grown a lot in learning under his unreserved guidance in the research and courses. The rewarding study experience at KAUST would not happen without his guidance, encouragement, and unconditional support.

The thesis presented results from the incredible supervision and advice I received from Dr. Mustafa A. Kishk. He taught me all the useful tools and answered all questions I have. I greatly appreciate his direct help and learned a lot from his thinking style and dedication to research. I also would like to give special thanks to my research group members in the Communication Theory Lab for helping both in research and courses.

Finally, I would like to express my profound appreciation to my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF KEY NOTATIONS

Φ_m, $_m$, p_m	PPP of MBSs, the corresponding density, and the corresponding transmission power
Φ_s, $_s$, p_s	PPP of SBSs, the corresponding density, and the corresponding transmission power
Φ_u, $_u$,p_u	PPP of MUs, the corresponding density, and the corresponding transmission power
; b	Path loss exponent; 2=
W	bandwidth
	Power control fraction
T	Set of computation tasks
N	Capacity of the buffer at SBSs
N_s^u	The number of MUs served by the SBS
Λ_i, $_i$	Arrival rate at type-i SBS, corresponding serving rate
A_k	Probability of MUs association with tier k
C_B	Service region of base station located at B
K_X	Serving tier of the MU located at X
H	Small scale fading gain

Chapter 1

Introduction

1.1 Mobile Edge Computing

With the realization of Internet of Things (IoT), it is an irreversible trend that more mobile devices will be connected to the wireless access system. This paradigm has motivated developers to create more interactive applications that require low-latency communication and computation, such as face recognition, natural language processing, and augmented reality (AR). As a growing computing paradigm, computation offloading can break such an obstacle. Computation offloading can effectively reduce the energy consumption of mobile devices, by migrating part or all of the tasks of mobile applications from resource limited mobile devices to the cloud. Mobile edge computing (MEC) is regarded as a promising solution to overcome this limitation. The idea behind MEC is to enhance the computing potentials of mobile devices by placing cloud computing platforms at the edges of the networks [1, 2, 3, 4].

As a key enabled network architecture of MEC, base stations (BSs) cascaded with edge-cloud servers were widely considered in existing works. Moreover, BSs are also able to access the central cloud in case of running out their computing or storage capacity, which results in a hierarchical computation offloading architecture. Different from the consideration of computation offloading architecture, the heterogeneity and diversity of computation services are often overlooked in many recent works. In general, different services require non-identical databases or libraries cached at the BS. For example, the type and datasize of object database for different AR services are

different. While the central cloud server has more resources, the limited computing and storage capacity of edge-cloud server allows only a small set of services to be cached during a typical period. Thus, the type of computation services cached at the BS not only determines the type of tasks that can be offloaded, but also affects the network performance.

Computation offloading has been the central theme of MEC network studies for the past decade. There have been a lot of existing works on designing policies for computation offloading. In [5], Ren *et al.* formulated and solved an optimization problem for minimizing the latency of multi-users partial computation offloading model. The authors in [6] studied an energy-efficient mobile-edge cloud computation offloading optimization problem with constraints on quality of service. In [7], You *et al.* studied the trade-off between energy consumption and computation latency for a multi-user MEC system in cases of infinite and finite cloud computation capacities. Besides, energy harvesting technology has been attracting more attention from researchers recently to prolong the networks lifetime. The authors in [8] incorporated energy harvesting technology into MEC system and proposed an energy-efficient policy which considered latency and offloading failure. Authors in [9] applied microwave power transfer (MPT) technology in low-complexity devices to achieve energy-efficient mobile edge cloud computing. By taking into account the availability of service in edge cloud servers, the joint optimal service caching (or service placement) and task offloading decisions were presented in [10]. Further, with the consideration of cooperative edge-cloud servers, the optimal service placement and computation task routing problem was addressed in [11].

1.2 Integrated Access and Backhaul Network

In recent years, global mobile data traffic has skyrocketed owing to dramatic growth in both end-user populations (e.g., smartphones and tablets) and demand for informa-

tion service (e.g., video streaming and cloud computing) [12]. Moreover, the volume of global mobile traffic reported in 2010 is predicted to increase 670-fold by 2030 [13]. To handle such exponential growth, the density of base stations (BSs) is expected to substantially increase in the future. Bhushan *et al.* discussed the feasibility of denser BSs deployment, referred to as network densification, as well as the anticipated requirement for addressing the increasing traffic growth in [14]. As a promising approach for extending the cell area and meeting the high capacity demand, network densification provides a reliable access channel by reducing the distance between mobile users and BSs and increasing the spectrum reuse [15].

However, in conventional network densification, the capital/operational cost of optical fiber deployment for BSs is a major drawback. For example, one meter of optical fiber deployment is estimated to cost approximately 100–200 USD in a downtown area, of which nearly 85% of the total expense is related to the trenching and installation operations [16]. Thus, the usage of wireless backhaul instead of the conventional wired-optical-fiber backhaul is an attractive solution. Wireless backhaul provides approximately the same transmission rate as that provided by optical fibers but at considerably less cost with extended flexible/timely deployment (e.g., no intrusion) [17].

Wireless backhaul in wireless networks has been extensively studied in the last 20 years [18]. However, no tight integration of access and backhaul has been realized in the studies of Long-Term Evolution (LTE) backhaul networks, since only one single hop with a fixed parent BS is supportable for LTE relay, and strict resource partitioning is specified between access and backhaul [19]. In contrast, the IAB network offers a more flexible deployment solution without the heavy overhead of optical-fiber installation [20]. Based on a wired connection to the core network, the IAB donor can provide communication access to mobile users and wireless backhaul to IAB nodes. IAB nodes are able to wirelessly provide network service access to the mobile user as

well as backhaul the access traffic wirelessly. Therefore, IAB nodes can be regarded as wireless relays for extending the coverage of an IAB donor. This functionality is helpful for networks to achieve robust coverage performance when the line-of-sight (LoS) propagation is blocked by environmental obstacles, such as buildings.

1.3 Stochastic Geometry-based Analysis

Stochastic geometry is applied to describe the random spatial placement of the network entities based on two-dimensional (2D) or three-dimensional (3D) point processes (PPs) in the Euclidean space [21]. It has been established as a strong tool for modeling, analysing, and designing large scale wireless networks [22, 23]. For instance, stochastic geometry has been adopted to the design and analysis of cellular networks, cognitive radio systems, ad-hoc networks, and many other types of wireless communication networks [24].

There have been several studies on stochastic geometry-based analysis of MEC-enabled networks. In particular, single-tier MEC networks were considered in [25, 26, 27], while heterogeneous MEC networks were considered in [28, 29]. Mobile users with identical computation tasks were considered in [26] and [27], with a spatial model characterized by Poisson point processes (PPPs) and Poisson cluster processes (PCPs), respectively. Although the heterogeneity of computation tasks was considered in [28] and [29], the heterogeneity of computation services was not captured. The load migration from the limited computation capacity edge server to the central cloud in a cell-free network was investigated in [29]. However, the detailed process of this load balancing scheme was not further studied.

1.4 Motivation and Main Contributions

In comparison to the vast amount of published work in the IAB and MEC networks, to the best of our knowledge, there is no research work that attempts to investigate the

performance in the IAB-enabled MEC network, which motivates this thesis. In this thesis, making use of stochastic geometry and queueing theory, we develop a tractable analytical framework for a two-tier heterogeneous MEC network, and obtain some useful design guidelines for the deployment studies on IAB-enabled MEC networks.

1.4.1 Tractable model for heterogeneous MEC network

Using tools from stochastic geometry and queuing theory, we develop a tractable and realistic model to analytically characterize the performance of a heterogeneous MEC network in terms of average delays. We assume that only the MBS has access to the fiber backhaul while the SBSs is wirelessly backhauled by the MBS. Due to the limited storage resources, we consider that only finite computation services and a limited set of tasks can be cached and offloaded at a typical SBS. We derive the average delay achieved by a typical mobile user for two strategies:

- Strategy I: The mobile users offload the tasks to the nearest BS. If the mobile users choose to connect with a SBS, and the tagged SBS has not cached the required service, the SBS will offload this task to the nearest MBS.
- Strategy II: The mobile users offload the tasks to the nearest BS that has already cached the required service, whether it is a SBS or a MBS.

1.4.2 System design insights

Using numerical results, we draw multiple useful system-level insights, which are summarized below.

- There exists an optimal association bias factor that minimizes the average delay, which is located between the optimal bias factors for average transmission time and average response time.

- The average delay for the two considered strategies decreases with the density of SBSs or MBSs increases. However, it is noticed that the increase in the MBSs' density has a stronger influence on the system performance than that of the SBSs.

- When the SBS service rate or buffer size exceeds a certain level, the average delay for the two considered strategies will achieve convergence.

- For the low and middle range of BS density and SBS service rate, Strategy I outperforms Strategy II in terms of the average delay. Thus, Strategy I is a more cost-effective solution.

Chapter 2

System Model

We consider a two-tier hierarchical MEC network. There are multiple MBSs and SBSs. Each BS is co-located with a computing server in order to provide computing services for mobile users (MUs). The locations of MBSs, SBSs and MUs are modeled by three independent PPPs, Φ_m with intensity λ_m, Φ_s with intensity λ_s, and Φ_u with intensity λ_u, respectively.

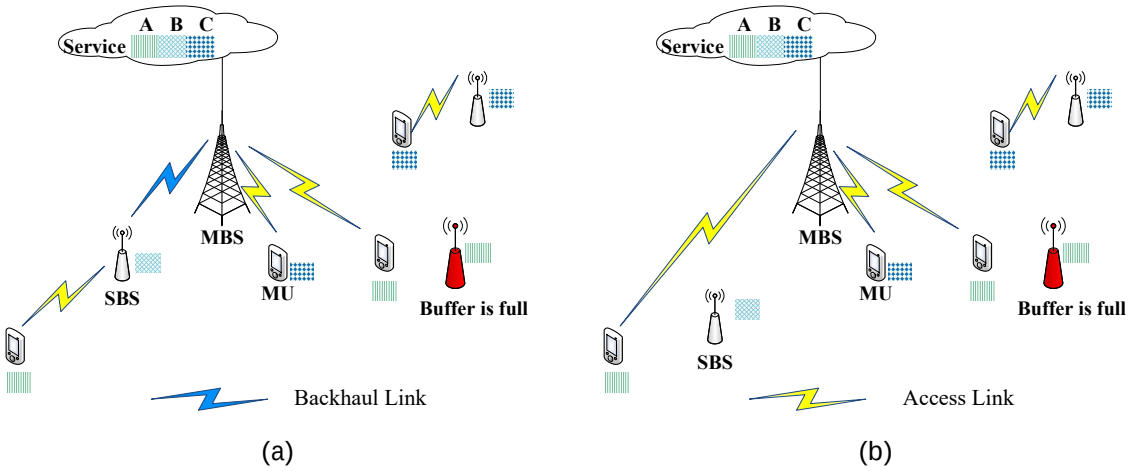

Figure 2.1: System model of : (a) Strategy I, (b) Strategy II.

MUs are capable of offloading their time-consuming and computation-intensive tasks to the MBSs or the SBSs, and then receiving the output of the task through the wireless network. The time consumption of computation offloading in the considered system consists of three sequential phases: transmitting, computing, and receiving phases.

As shown in Fig. 2.1, we consider two different computation offloading strategies

for the considered system. For strategy II, a typical MU is only able to offload its task to a MBS or an *available* SBS. A SBS is *available* if it has already cached the required service for the computing task generated at the MU, and it has not run out of the limited storage space at its buffer. Compared with Strategy II, Strategy I is more flexible, since the operation inside the dash-lined box in Fig. 2.2 is considered for Strategy I. For instance, MU can offload its task to a SBS even though this SBS has not cached the requested service. With the consideration of IAB, a MBS provides cloud computing service for MUs and SBSs in its voronoi cell through access link and backhaul link, respectively. In other words, SBSs in Strategy I can act as relays to assist MUs in offloading their computation tasks to the MBS.

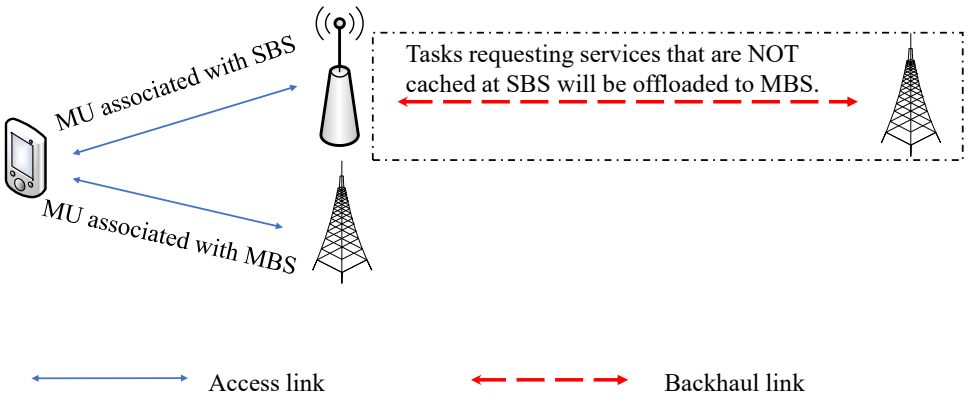

Figure 2.2: Illustration of Strategy I.

2.1 Communication model

We assume MBSs and SBSs have different transmitting powers (which are denoted by p_m and p_s), and the signals experience path loss with the same path loss exponent . The received power at a receiver located at Y from a transmitter located at X is $p_X HkX \quad Yk$, where H denotes the channel power gain, p_X is the transmitting power. The random channel gains are followed the Rayleigh distribution with unitary average power, i.e., $H \quad \exp(1)$. Besides, we assume that the MUs in the same small

cell use different orthogonal resource blocks, as well as the SBSs in the same macro cell.

We consider using the maximum biased average power-based cell association rule. Let D_m and D_s denote the distance of a typical MU from the nearest MBS and the nearest SBS, respectively. The typical MU located at X will choose to connect to tier K if

$$K_X = \arg \max_{k \in \{s,m\}} p_k B_k D_k \ ; \tag{2.1}$$

where B_k denotes the association bias factor. The association cell of a BS of type $k \in \{s,m\}$ located at X is given by

$$C_X = \{ Z \in \mathbb{R}^2 : p_k B_k \| Z - X_k \| \leq p_j B_j D_j \ ; \forall j \in \{s,m\} \}. \tag{2.2}$$

Lemma 1 (Mobile user association probability)**.** *The probability that a typical MU connects with tier k is given by*

$$A_k = \frac{\lambda_k}{\lambda_k + \sum_j \lambda_j [(p_j B_j)/(p_k B_k)]^{2/\alpha}}. \tag{2.3}$$

Proof. See Appendix A. □

As for the backhaul link in the Strategy I, the typical SBS associates with its nearest MBS.

2.2 Service and Computation Model

There is a set of computation tasks $\{T_i | i = 1, 2, \ldots, M\}$, each T_i requires a specific service database S_i. The data size of any given task is D. A typical user generates a task T_i with probability q_i. Moreover, due to the limited storage capacity at the

SBS, we assume that each SBS only cached one specific S_i, with probability q_i. In the following, we refer to the SBS which has cached S_i as the type-i SBS.

Since both MUs and SBSs are distributed as PPP, the arrivals of offloading tasks at a computing server can be regarded as a Poisson process with a specific arrival rate. Similar to [28], the arrival rate for at a type-i SBS in two considered strategies are given respectively by

$$\Lambda^1_{s;i} = \frac{q_i A_s \ u}{s}; \qquad (2.4a) \qquad\qquad \Lambda^2_{s;i} = \frac{A_{s;i} \ u}{s}; \qquad (2.4b)$$

where $A_{s;i}$ denotes the probability that a typical MU with type-i task associates with a SBS in Strategy II.

Moreover, due to the limited computation and communication capacity at the SBSs, we assume that the computing buffer at each SBS is bounded by N. The SBS can not provide service for additional MUs when its computing buffer is full. The probability that the SBS has a full buffer is provided next.

According to [30], the probability that the buffer is not full for the type-i SBS is defined as

$$P^i_{\text{o d}} = 1 \quad \Pr(\text{The SBS's buffer is full}) = 1 \quad \frac{(1 \quad _i) \ _i{}^N}{1 \quad _i{}^{N+1}}; \qquad (2.5)$$

where $_i = \Lambda_{s;i} = \ _s$, and $\Lambda_{s;i}$ is the arrival rate given by (2.4a) and (2.4b) for Strategy I and Strategy II, respectively.

For Strategy I, an SBS needs to offload a task to the MBS if the MU requests a task that is not cached at the SBS. The fraction of SBSs that are offloading tasks to MBSs using wireless backhaul can be computed using the below probability.

Definition 1. *The typical SBS is backhaul-active (BH-actv) if it receives a task from at least one MU that is not cashed in this SBS. The probability for a typical SBS to*

be active is

$$P_{\text{actv}} = \Pr(\text{SBS received } T_j \mid \text{SBS cached service } i; i \neq j)$$
$$= \mathsf{E}_{N_s^u}\Big[\sum_{i=1}^{M}(1 - q_i)\sum_{k=1}^{N_s^u} q_i^k(1 - q_i)^{N_s^u - k}\Big]; \tag{2.6}$$

where N_s^u is the number of MUs served by typical SBS.

According to [31], the probability mass function (PMF) of N_s^u is given by

$$\Pr(N_s^u = n) = \frac{\Gamma(n + 3.5)}{\Gamma(3.5)(n-1)!}\frac{3.5^{3.5}}{\left(\frac{\lambda_u A_s}{s}\right)^{(n-1)}}\left(3.5 + \frac{\lambda_u A_s}{s}\right)^{(n+3.5)}. \tag{2.7}$$

Similar to (2.4a) and (2.4b), the arrival rates at MBS for Strategy I and Strategy II respectively are

$$\Lambda_M^1 = \frac{A_m \lambda_u + P_{\text{actv}} \lambda_s}{m}; \tag{2.8a}$$

$$\Lambda_M^2 = \sum_{i=1}^{M} \frac{A_{m;i} \lambda_u}{m}; \tag{2.8b}$$

where $A_{m;i}$ is the probability that a MU with type-i task associates a MBS in Strategy II.

Unlike SBSs, the MBSs are supposed to be equipped with an infinite computing buffer. Computing servers at both MBSs and SBSs are assumed to operate in first-come-first-serve manner. The computation time for the tasks at edge server are exponential RVs with mean $t_k; k \in \{s; m\}$. Therefore, the service rates at SBSs and MBSs are $\mu_s = 1/t_s$ and $\mu_m = 1/t_m$. In this context, we are able to adopt M/M/1 and M/M/1/N queueing models to analyze the computing server cascaded with MBSs and SBSs, respectively.

Chapter 3

Analytical Framework

3.1 Strategy I

We consider using the Orthogonal Resource Allocation (ORA) as the bandwidth partitioning scheme for Strategy I [32]. In ORA, there exists a fraction $_a$ of bandwidth reserved for access links, and the rest is used for backhaul links. Therefore, the backhaul link will not experience interference from the access link and vice versa.

3.1.1 Uplink Coverage Analysis

Let $\Phi_{s^1_{o\ d}}$ denote the point process that models the locations of offloadable SBSs in Strategy I, which can be regarded as an independent thinning of Φ_s. By using (2.5) and the law of total probability, the thinning probability is defined as $P_{o\ d} = \mathsf{E}[P^i_{o\ d}] = \sum_{i=1}^{M} q_i P^i_{o\ d}$. According to [33], the intensity of $\Phi_{s^1_{o\ d}}$ is $P_{o\ d\ \ s}$.

If the typical MU is connected to the nearest SBS/MBS located at $\mathbf{B_a}$, assuming that the distance between the user and $\mathbf{B_a}$ is $R_{a;1}$, the signal-to-interference-plus-noise ratio (SINR) at the tagged SBS/MBS for the uplink access link is given by

$$\mathrm{SINR}^u_{a;1} = \frac{p_u H_u R_{a;1}^{(\ \ 1)}}{^2 + p_u I_{u;1}}; \tag{3.1}$$

with

$$I_{u;1} = \sum_{\mathbf{x}2\ \ ^a_{u;1}} (R_x)\ H_x k\mathbf{x} \quad \mathbf{B_a} k \quad ; \tag{3.2}$$

where R_x denotes the serving link distance for interfering MU, $\Phi^a_{u;1}$ denotes the point process for MUs that use the same resource block as the typical user in Strategy I.

If the tagged SBS has not already cached the service, the SBS will transmit the offloaded task to the nearest MBS (located at $\mathbf{B}_b$), the SINR at the tagged MBS for the uplink backhaul link is given by

$$\mathrm{SINR}^u_b = \frac{p_s H_s R_b}{^2 + p_s I_s}; \tag{3.3}$$

with

$$I_s = \sum_{\mathbf{x}2\ ^b_{S_{\mathrm{actv}}}} H_j k\mathbf{x} \quad \mathbf{B_b}k \quad ; \tag{3.4}$$

where $\Phi^b_{S_{\mathrm{actv}}}$ and R_b denote the point process that models the locations of the active SBSs that are using the same resource block as the typical SBS and the serving distance from typical SBS to its nearest MBS, respectively.

According to the law of total probability, the uplink coverage probability in Strategy I is given by

$$\begin{aligned} P^u_1(\ ) &= \sum_{i=1} q_i \Pr(\mathrm{SINR}^{\mathrm{u}}_{\mathrm{a};1} > \ j\text{Typical user has task } T_i) \\ &= \sum_{i=1} q_i \Big[\Pr(K = s)\ q_i P^u_{s;1}(\ ) + (1 \quad q_i) P^u_{s;1}(\ ) P^u_b(\ ) \ + \Pr(K = m) P^u_{m;1}(\ ) \Big]; \end{aligned} \tag{3.5}$$

in which

$$P^u_{k;1}(\ ) = \Pr(\mathrm{SINR}^u_{a;1} > \ \ jK = k; k \ 2 \ fs; mg)$$
$$= \mathsf{E}\left[\exp\left(\frac{(\ ^2 + p_u I_{u;1})}{p_u R_{a;1}^{(\ \ 1)}} \right) \ \middle| \ jK = k \right]$$
$$= \mathsf{E}\left[L_{u;1jK=k}\left(R_{a;1}^{(1\ \ )} \right) \exp\left(\frac{R_{a;1}^{(1\ \ )\ \ 2}}{p_u} \right) \ \middle| \ jK = k \right]; \qquad (3.6)$$

and

$$P^u_b(\ ) = \Pr(\mathrm{SINR}^u_b > \ \ )$$
$$= \mathsf{E}\left[\exp\left(\frac{(\ ^2 + p_s I_s)}{p_s R_b} \right) \right]$$
$$= \mathsf{E}\left[L_s(R_2\ \ ) \exp\left(\frac{R_2\ \ ^2}{p_s} \right) \right]; \qquad (3.7)$$

where L_s and $L_{u;1jK=k}$ are the Laplace transform of I_s and $I_{u;1}$ conditional on tier k being the serving tier, respectively. In the following, we use $k = 1$ or $k = 2$ instead of $k = m$ or $k = s_{0\ \ d}^{1}$ for simplification.

To derive the uplink coverage probability, we first need to characterize the path loss distribution for the desired link of the typical MU and the Laplace transform of the interference. The distribution of the path loss between a typical MU and its serving BS is given in the Lemma below.

Lemma 2. *The probability density function (PDF) of path loss at MU to its serving BS (i.e., $L_{a;1} = R_{a;1}$) is given by*

$$f_{L_{a;1}jK=k}(l) = \frac{ba_k}{A_k} l^{b\ \ 1} \exp(\ \ G_k l^b); \qquad (3.8)$$

where $b = \frac{2}{}$, $a_k = \ \ _k\ \ $, $G_k = \sum_{j=1}^{\mathbb{P}2} a_j (p_j B_j = p_k B_k)^b$, and A_k is de ned in Lemma 1.

Proof. See Appendix B. □

Similarly, we can obtain the PDF of the path loss between the SBS and its serving MBS, i.e., $L_b = R_b$ as

$$f_{L_b}(l) = ba_1 l^{b\ 1} \exp(\ \ a_1 l^b): \tag{3.9}$$

The path loss distribution of the interfering MU at a typical BS is the conditional distribution given that the interfering MU is not associated with the tagged BS. Therefore, the distribution of the path loss between an interfering MU and the tagged BS is not identical to the distribution given in Lemma 2. This correlation is formalized in the following.

The PDF of the path loss of an interfering MU located at U served by a tier j BS, conditioned on it not lying in the associated cell of the tagged tier k BS located at Y, is given by

$$\begin{aligned} &f_{L_U}(l|K_U = j; Y\ 2 \text{ tier } k; U\ 6\ C_Y; kU\ \ \ Yk\ \ = y) \\ &= \frac{bG_j}{1\ \ \exp(\ \ G_k y^b)} l^{b\ 1} \exp(\ \ G_j l^b); \quad 0\ \ \ l\ \ \ \frac{p_j B_j}{p_k B_k} y: \end{aligned} \tag{3.10}$$

Since different MUs associated with the same BS transmit on different resource blocks, there is only one MU from each cell can act as interferer to other cellular cells. Thereby $\Phi^a_{u;1}$ is not a PPP but a Poisson-Voronoi perturbed lattice. For tractability purpose, we adopt the method to approximate $\Phi^a_{u;1}$ as an inhomogeneous PPP [34]. In the following, we will characterize the Laplace transform of the interference.

Remark 1. *Suppose a typical MU is located at X, and a tier k BS is located at Y, the probability that this MU is associated with the base station is* $\exp(\ \ G_k kX\ \ \ Yk\ )$. *Then, the intensity measure function for interference from MUs associated with tier j BS ($\Phi^a_{u_j;1}$) can be written as*

$$^{j}_{u;1jk} = ba_j x^{b\ 1}[1\ \ \ \exp(\ \ G_k kX\ \ \ Yk\ )]\,(\mathrm{d}x): \tag{3.11}$$

.

Lemma 3. *We assume that the point processes of interfering MUs from each tier are independent, and the path loss distributions of interfering MUs are independent. In that case, the Laplace transform of MU interference is given by*

$$L_{u;1jK=k}(s) = \exp\left(\frac{bs}{1 \quad b} \sum_{j=1}^{\infty} \left(\frac{p_j B_j}{p_k B_k} \right)^{1 \quad b} a_j \quad \mathsf{E}_{L_{ajK=j}} \quad L_a^{b \quad (1 \quad)} C_b \left(\frac{s p_j B_j}{p_k B_k L_a^1} \right) \right); \tag{3.12}$$

where $C_b(x) = \ {}_2F_1(1; 1 \quad b; 2 \quad b; \quad x)$ *and* ${}_2F_1(\ ; \ ; \ ; \ )$ *is the hypergeometric function.*

Proof. See Appendix C. □

Lemma 4. *Laplace transform of SBSs interference is*

$$L_s(s) = \exp\left(\quad 2 \quad {}_{sI} \int_0^{\infty} \frac{[1 \quad \exp(\quad a_m x^2)]\, x}{1 + x \ = s} \mathrm{d}x \right); \tag{3.13}$$

where ${}_{sI} = \min(\ {}_m; \ {}_s P_{\text{actv}})$ *is density of active SBSs that might potentially interfere in the uplink.*

Proof. The proof is similar to that of [35, Theorem 1] and hence omitted here. □

By using Lemma. 4, (3.6) and (3.7) can be recast in the following reduced forms

$$P^u_{s;1}(\ ) = \frac{b a_2}{A_2} \int_0^{\infty} x^{b \ 1} \exp\left(\frac{b \ x^{(1 \ )}}{1 \quad b} \sum_{j=1}^{\infty} \left(\frac{p_j B_j}{p_s B_s} \right)^{1 \quad b} a_j \mathsf{E}_{L_{ajK=j}} \quad L_a^{b \quad (1 \quad)} C_b \left(\frac{x^{(1 \quad)} p_j B_j}{p_s B_s L_a^1} \right) \right)$$
$$\frac{x^{(1 \quad) \ 2}}{p_u} \quad G_2 x^b \quad \mathrm{d}x; \tag{3.14}$$

$$P^{u}_{m;1}(\ ) = \frac{ba_1}{A_1}\int_0^{Z^1} x^{b\ 1}\exp\left(-\frac{b\ x^{(1\ )}}{1\quad b}\sum_{j=1}^{2}\left(\frac{p_j B_j}{p_m B_m}\right)^{1\ b}\left(\frac{x^{(1\ )\ 2}}{p_u}\right)^{\quad}G_1 x^b\right.$$
$$\left.a_j \mathsf{E}_{L_a jK=j}\left[L_a^{b\ (1\ )}\right]C_b\left(\frac{x^{(1\ )} p_j B_j}{p_m B_m L_a^1}\right)^{\quad}\right)\mathrm{d}x; \tag{3.15}$$

$$P^{u}_{b}(\ ) = ba_1\int_0^{Z^1} x^{b\ 1}\exp\left(-\frac{x^{\ 2}}{p_s}\right)^{\quad}a_1 x^b\quad 2\quad {}_{SI}\int_0^{Z^1}\frac{[1\quad \exp(\ \ a_1 y^2)]\,y}{1+x\ y}\mathrm{d}y\Bigg)^{!}\mathrm{d}x: \tag{3.16}$$

By substituting (3.14), (3.15), and (3.16) into (3.5), we can derive the analytical expression of the uplink coverage probability.

3.1.2 Downlink Coverage Analysis

For the downlink transmission, the SINR at the typical MU associated with BS located at X for the access link is given by

$$\mathrm{SINR}^{d}_{a;1} = \frac{p_X H_d R_{a;1}}{^2 + I_{d;1}}; \tag{3.17}$$

with

$$I_{d;1} = \sum_{\mathbf{x}2\ \ {}^{a}_{d;1}} p_x H_x k\mathbf{x}k^{\quad}\ ; \tag{3.18}$$

where $\Phi^{a}_{d;1}$ denotes the point process of SBSs and MBSs using the same resource block as the associated BS.

The SINR at the typical SBS for the downlink backhaul is given by

$$\mathrm{SINR}_b^d = \frac{p_m H_m R_b}{^2 + I_{s^0}}; \tag{3.19}$$

with

$$I_{s^0} = \sum_{x 2 \ {}_b^{Sm}} p_m H_x k x k \quad ; \tag{3.20}$$

where Φ_b^m denotes the MBSs use the same resource block as the typical SBS serving MBS.

The downlink coverage probability is given by

$$\begin{aligned} P_1^d(\ ) &= \sum_{i=1}^{M} q_i \Pr(\mathrm{SINR}_1^d > \ \ j\text{Typical user has task } T_i) \\ &= \sum_{i=1}^{M} q_i \quad \Pr(K=m) P_{m;1}^d(\ ) + \Pr(K=s)\ \ q_i P_{s;1}^d(\ ) + (1 \quad q_i) P_{s;1}^d(\ ) P_b^d(\ ) \quad ; \end{aligned} \tag{3.21}$$

in which

$$\begin{aligned} P_{k;1}^d(\ ) &= \Pr(\mathrm{SINR}_{a;1}^d > \ \ jK=k; k\ 2\ fs; mg) = \mathsf{E}\ \ \exp \quad \frac{(\ ^2 + I_{d;1})}{p_k R_a} \ \ jK=k \\ &= \mathsf{E}\ \Big[L_{d;1jK=k} \ \ \frac{R_a}{p_k} \quad \exp \quad \frac{R_{a;1}\ ^2}{p_k} \ \ jK=k \Big] \ ; \end{aligned} \tag{3.22}$$

and

$$\begin{aligned} P_b^d(\ ) &= \Pr(\mathrm{SINR}_b^d > \ \ ) = \mathsf{E}\ \ \exp \quad \frac{(\ ^2 + I_s^{\ 0})}{p_m R_b} \\ &= \mathsf{E}\ \ L_{s^0} \ \ \frac{R_b}{p_m} \quad \exp \quad \frac{R_b\ ^2}{p_m} \quad : \end{aligned} \tag{3.23}$$

Lemma 5. *The Laplace transform of downlink access link interference ($I_{d;1}$) when*

the serving BS belongs to tier k and the corresponding path loss is l can be expressed as

$$L_{d;1|L_{a=l}}(s) = \exp\left(\frac{bs}{1-b}\sum_{j=1}^{2} p_j \left(\frac{p_j B_j l}{p_k B_k}\right)^{b-1} a_j C_b \left(\frac{sp_k B_k}{l B_j}\right)^{b}\right); \tag{3.24}$$

The Laplace transform of downlink backhaul link interference (I_{s^0}) when the corresponding path loss is l can be expressed as

$$L_{s^0|L_b=l}(s) = \exp\left(\frac{2s}{b-2} p_m a_m l^{b-1} C_b \left(\frac{sp_m}{l}\right)^{b}\right); \tag{3.25}$$

Proof. The proof is similar to that of [36, Theorem 1] and hence omitted here. □

By using Lemma 5, (3.22) and (3.23) can be written in the following reduced forms

$$P^d_{s;1}(\ ) = \frac{ba_2}{A_2}\int_0^{1} x^{b-1}\exp\left(\frac{bx^b}{(1-b)}\sum_{j=1}^{2}\left(\frac{B_j}{B_s}\right)^{b-1}\left(\frac{p_j}{p_s}\right)^{b} a_j C_b \left(\frac{B_s}{B_j}\right)^{b}\left(\frac{x^2}{p_s}\right)^{b} G_2 x^b\right) dx; \tag{3.26}$$

$$P^d_{m;1}(\ ) = \frac{ba_1}{A_1}\int_0^{1} x^{b-1}\exp\left(\frac{bx^b}{(1-b)}\sum_{j=1}^{2}\left(\frac{B_j}{B_m}\right)^{b-1}\left(\frac{p_j}{p_m}\right)^{b} a_j C_b \left(\frac{B_m}{B_j}\right)^{b}\left(\frac{x^2}{p_m}\right)^{b} G_1 x^b\right) dx; \tag{3.27}$$

$$P^d_b(\ ) = ba_1\int_0^{1} x^{b-1}\exp\left(\frac{2a_1 x^b}{2} C_b(\ )\left(\frac{x^2}{p_m}\right)^{b} a_1 x^b\right) dx: \tag{3.28}$$

By substituting (3.26), (3.27), and (3.28) into (3.21), we can derive the analytical expression of the downlink coverage probability.

3.1.3 Average Delay

For simplicity of analysis, we assume the load (i.e., the number of MUs associated with the tagged base station) is equal to its mean.

Recalling that $_a$ is the fraction of bandwidth reserved for access links, then for a given $_a$, the bandwidth obtained by a mobile user in the access link W^a or by the SBS in the backhaul link W^b is

$$\begin{cases} W^a_{m;1} = \dfrac{_aW}{N^u_m}; & \text{MU connected to MBS;} \\ W^a_{s;1} = \dfrac{_aW}{N^u_s}; & \text{MU connected to SBS;} \\ W^b = \dfrac{(1 \quad _a)W}{N^s_m}; & \text{MU connected to SBS:} \end{cases} \tag{3.29}$$

Definition 2 (Response Time)**.** *The response time is the time that a computation task spends in server, i.e., the sum of waiting time and service time.* $T^r_{m;1}$ *and* $T^r_{s_i;1}$ *denote the response time of MBS server and type-*i *SBS server, respectively.*

For the $M/M/1$ queuing model applied at MBS server, the response time is exponentially distributed with parameter $_m \quad \Lambda^1_m$. Therefore the average response time is $T^r_{m;1} = \frac{1}{_m \quad \frac{1}{m}}$, where $\Lambda^1_m = \frac{_uA_mP^u_{m;1}(\ )}{_m} + \frac{_sP_{\text{actv}}P^u_b(\ )}{_m}$.

According to [30], the expected response time at type-i SBS is given by

$$T^r_{s_i;1} = \frac{\bar{Q}_i}{\Lambda^1_{s;i}(1 \quad _i)}; \tag{3.30}$$

where $\bar{Q}_i = \frac{_i\left(1 \quad (N+1)\ ^N_i + N\ ^{N+1}_i\right)}{(1 \quad)\left(1 \quad ^{N+1}_i\right)}$, $_i = \frac{^N_i}{\sum_{n=0} \quad ^n_i}$, $_i = \Lambda^1_{s;i}/ \ _s$, and $\Lambda^1_{s;i} = q_iA_s \ _uP^u_{s;1}(\ )/ \ _{s^1_0 \ d}$.

Let denote the SINR threshold for successful offloading, the average uplink

transmission rate can be expressed as

$$\begin{cases} R^u_{m;1} = \mathsf{E}[W^a_{m;1}\log_2(1+\ )\mathsf{I}(\mathrm{SINR}^u_{m;1}\ \ )] \\ \qquad\ = P^u_{m;1}(\ )W^a_{m;1}\log_2(1+\ ); \quad \text{MU-MBS;} \\ R^u_{s;1} = \mathsf{E}[W^a_{s;1}\log_2(1+\ )\mathsf{I}(\mathrm{SINR}^a_{s;1}\ \ ] \\ \qquad\ = P^u_{s;1}(\ )W^a_{s;1}\log_2(1+\ ); \quad \text{MU-SBS;} \\ R^u_b = \mathsf{E}[W^b\log_2(1+\ )\mathsf{I}(\mathrm{SINR}^u_b\ \ )] \\ \qquad\ = P^u_b(\ )W_b\log_2(1+\ ); \quad \text{SBS-MBS;} \end{cases}$$

where $\mathsf{I}(\ )$ denotes the indicator function.

Therefore, for a given task data size D, the average uplink transmission time is

$$\begin{cases} T^{t;u}_{m;1} = \dfrac{D}{R^u_{m;1}}; & \text{From MU to MBS;} \\ T^{t;u}_{s;1} = \dfrac{D}{R^u_{s;1}}; & \text{From MU to SBS;} \\ T^{t;u}_b = \dfrac{D}{\min(R^u_{s;1}; R^u_b)}; & \text{From MU to MBS with the aid of SBS:} \end{cases} \tag{3.31}$$

Similarly, the average downlink transmission rate can be expressed as

$$\begin{cases} R^d_{m;1} = \mathsf{E}[W^a_{m;1}\log_2(1+\ )\mathsf{I}(\mathrm{SINR}^d_{m;1}\ \ )] \\ \qquad\ = P^d_{m;1}(\ )W^a_{m;1}\log_2(1+\ ); \quad \text{MBS-MU;} \\ R^d_{s;1} = \mathsf{E}[W^a_{s;1}\log_2(1+\ )\mathsf{I}(\mathrm{SINR}^d_{s;1}\ \ )] \\ \qquad\ = P^d_{s;1}(\ )W^a_{s;1}\log_2(1+\ ); \quad \text{MU-SBS;} \\ R^d_b = \mathsf{E}[W^b\log_2(1+\ )\mathsf{I}(\mathrm{SINR}^d_b\ \ )] \\ \qquad\ = P^d_b(\ )W_b\log_2(1+\ ); \quad \text{SBS-MBS:} \end{cases} \tag{3.32}$$

The downlink transmission time is given by,

$$\begin{cases} T_{m;1}^{t;d} = \dfrac{D}{R_{m;1}^{d}}; & \text{From MBS to MU;} \\ T_{s;1}^{t;d} = \dfrac{D}{R_{s;1}^{d}}; & \text{From SBS to MU;} \\ T_{b}^{t;d} = \dfrac{D}{\min(R_{s;1}^{d}; R_{b}^{d})}; & \text{From MBS to MU with the aid of SBS:} \end{cases} \tag{3.33}$$

The average delay for a MU with T_i is given by

$$\begin{aligned} T_{\text{avg};1}^{i}(\ ) =& \mathsf{E}\left[T_{\text{respon:}} + T_{\text{trans:}}\right] \\ =& A_s \left(q_i \left[T_{s_i;1}^{r} + T_{s;1}^{t;u}(\ ) + T_{s;1}^{t;d}(\ ) \right] + (1 - q_i) \left[T_{m;1}^{r} + T_{b}^{t;u}(\ ) + T_{b}^{t;d}(\ ) \right] \right) \\ &+ A_m \left[T_{m;1}^{r} + T_{m;1}^{t;d}(\ ) + T_{m;1}^{t;u}(\ ) \right]: \end{aligned} \tag{3.34}$$

3.2 Strategy II

For Strategy II, the typical MU only offloads its task to the MBS or SBS via the access link directly. Hence, there are no wireless backhaul links between MBSs and SBSs. In this section, we use $_{s_i}$ to denote the density of the SBSs that have cached the service database S_i and $_{u_i}$ to denote the density of the MUs with task T_i, where $\sum_{i=1}^{\infty} {}_{s_i} = {}_{s}$ and $\sum_{i=1}^{\infty} {}_{u_i} = {}_{u}$.

Accordin to [34], the average number of MUs connected to a SBS with S_i is

$$\mathsf{E}[N_{s_i}^{u}] = 1 + \frac{1{:}28 \ {}_{u_i} A_{s_i}}{{}_{s_i}}; \tag{3.35}$$

in which

$$A_{s_i} = \frac{{}_{s_i}}{{}_{s_i} k + \ {}_{m} (p_m B_m / p_s B_s)^{2/}}: \tag{3.36}$$

By using (2.5), we can obtain $P^{i}_{\mathrm{o\ d}}$ in Strategy II. The sets of BSs can be regraded as $M+1$ tiers, where tier 0 denotes the MBSs and tier i ($1 \quad i \quad M$) denotes the SBSs which have cached S_i. In the following, we use $_0 = \ _m$ and $_i = \ _{s_i} \quad P^{i}_{\mathrm{o\ d}}$ to represent the intensity of point processes for MBSs and offloadable type-i SBSs, respectively.

3.2.1 Uplink Coverage Analysis

If the typical MU with T_i is connected to the nearest offload-able type-i SBS or MBS located at $\mathbf{B}$, denoting the distance between the user and $\mathbf{B}$ as R, the SINR at the tagged SBS/MBS is given by

$$\mathrm{SINR}^{u}_{a;2} = \frac{p_u H_u R^{\ (\ \ 1)}}{^2 + p_u I_{u;2}}; \tag{3.37}$$

with

$$I_{u;2} = \sum_{\mathbf{x}_i 2 \ \ ^{u;2}_{a}} (R_{x_i})\ H_i k \mathbf{x}_i \quad \mathbf{B} k \quad : \tag{3.38}$$

The uplink coverage probability is given by

$$\begin{aligned} P^{u}_{2}(\ \ ) &= \sum_{i=1}^{M} q_i \Pr(\mathrm{SINR}^{u}_{2} > \ \ j\text{Typical MU with task } T_i) \\ &= \sum_{i=1}^{M} q_i \ \ \Pr(K = i) P^{u}_{s_i}(\ \ ) + \Pr(K = 0) P^{u}_{0_i}(\ \ ) \ \ ; \end{aligned} \tag{3.39}$$

in which

$$
\begin{aligned}
P^u_{s_i}(\ ) =& \Pr(\mathrm{SINR}^u_{a;2} > \ \ | K = i) \\
=& \mathrm{E} \ \exp \ \frac{(\ ^2 p_u + p_u I_{u;2})}{p_u R^{\ (\ \ 1)}} \ \ | K = i \\
& " \qquad\qquad\qquad\qquad\qquad\qquad\qquad \# \\
=& \mathrm{E} \ \ L_{u;2|K=i} \ \ R^{\ (1\ \ )} \qquad \exp \ \ \frac{R^{\ (1\ \ )\ \ \ 2}}{p_u} \ \ | K = i \ \ ;
\end{aligned}
\tag{3.40}
$$

and

$$
\begin{aligned}
P^u_{0_i}(\ ) =& \Pr(\mathrm{SINR}^u_{a;2} > \ \ | K = 0; \text{MU with } T_i) \\
=& \mathrm{E} \ \exp \ \frac{(\ ^2 + p_u I_{u;2})}{p_u R^{\ (\ \ 1)}} \ \ | K = 0; \text{MU with } T_i \\
& " \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \# \\
=& \mathrm{E} \ \ L_{u;2|K=0} \ \ R^{\ (1\ \ )} \qquad \exp \ \ \frac{R^{\ (1\ \ )\ \ \ 2}}{p_u} \ \ \ K = 0; \text{ MU with } T_i \ \ :
\end{aligned}
\tag{3.41}
$$

The PDF of $L_i = \{R \ \ | K = i; i > 0\}$ is

$$
f_{L_i}(l) = \frac{b a_i}{A_{s_i}} l^{b\ \ 1} \exp(\ \ G_i l^b); \tag{3.42}
$$

where 8 $1 \ \ i \ \ M$, $a_i = P^i_{0\ \mathrm{d}\ \ i}$, and $G_i = a_i + a_0 \ \ \frac{p_m B_m}{p_s B_s} \ ^{b}$.

The PDF of path loss of MU with task T_i to MBS is

$$
f_{L_{0_i}}(l) = b G_{0_i} l^{b\ \ 1} \exp(\ \ G_{0i} l^b); \tag{3.43}
$$

where $G_{0_j} = a_0 + a_j \ \ \frac{p_s B_s}{p_m B_m} \ ^{b}$.

The PDF of the path loss of a MU with task T_k located at $\mathbf{X}$ associated with tier j, conditioned on it not lying in the serving cell of the tagged tier i BS located at B

and the corresponding path loss $k\mathbf{X} \quad \mathbf{B}k \ = y$, is

$$
\begin{aligned}
& f_{L_X}(l|jK_X = j; X \ 6 \ C_B; B \ 2 \text{ tier } i; k\mathbf{X} \quad \mathbf{B}k \ = y) \\
& = \begin{cases}
\dfrac{bG_j}{1 \quad \exp(\ \ G_{0j}y^b)} l^{b \ \ 1} \exp(\ \ G_j l^b); & \\
\qquad\qquad 0 \quad l \quad \dfrac{p_s B_s}{p_m B_m} y; i = 0; j \ 6= 0; & \\
\dfrac{bG_{0k}}{1 \quad \exp(\ \ G_{0k}y^b)} l^{b \ \ 1} \exp(\ \ G_{0k} l^b); & \\
\qquad\qquad 0 \quad l \quad y; i = j = 0; & \\
\dfrac{bG_j}{1 \quad \exp(\ \ G_j y^b)} l^{b \ \ 1} \exp(\ \ G_j l^b); & \\
\qquad\qquad 0 \quad l \quad y; \ i \ 6= 0; j = i; & \\
bG_j l^{b \ \ 1} \exp(\ \ G_j l^b); \qquad\qquad i \ 6= 0; \ i \ 6= j; j \ 6= 0; & \\
\dfrac{bG_{0i}}{1 \quad \exp(\ \ G_i y^b)} l^{b \ \ 1} \exp(\ \ G_{0i} l^b); & \\
\qquad\qquad 0 \quad l \quad \dfrac{p_m B_m}{p_s B_s} y; i \ 6= 0; \ j = 0; k = i; & \\
bG_{0k} l^{b \ \ 1} \exp(\ \ G_{0k} l^b); \qquad\qquad i \ 6= 0; \ j = 0; k \ 6= i: &
\end{cases}
\end{aligned}
\tag{3.44}
$$

Proof. 1) For $i = 0$ and $j \ 6= 0$, given the association policy, L_X should be bounded by $\frac{P_s}{P_m} k\mathbf{X} \quad \mathbf{B}k$. With $G_{0j} \quad \frac{P_s B_s}{P_m B_m}^{b} = G_j$, we can get the result for the first case in the above distribution. 2) Similarly, for $i = j = 0$, $8k$, we can obtain the result for the second case. 3) If $i \ 6= 0$ and $j = i$, with the fact that the path loss of an interfering user to its serving BS has to be smaller than its distance to the tagged BS, the distribution is given for the third case. 4) As for $i \ 6= 0$, $i \ 6= j$, as well as $j \ 6= 0$, the user associated with tier j will always be interfering at tagged BS in tier i. Thus, there is no bound for L_X. 5) When $i \ 6= 0$, $j = 0$, and $k = i$, L_X has an upper bound. 6) However, if $k \ 6= j$, the user will always be the interference at the tagged BS. Thereby, there is no upper bound in the sixth case. □

Remark 2. *According to our association policy, conditioned on a BS of tier i located at V, a MU with task T_j at U associates with V with probability*

$$\begin{cases} \exp(-G_{0i}kU - Vk) & ; i = 0; \\ \exp(-G_ikU - Vk) & ; i = j; \\ 0 & ; i \neq j: \end{cases} \tag{3.45}$$

Conditioned on the tagged BS located at B and of tier i, the propagation process of interfering MUs from tier j ($j \geq 1$; i.e., type-j SBS) to B, $I_{u;j} := f k\mathbf{X} - \mathbf{B}k\ g_{X2\ u;j}$ with intensity measure function

$$\Lambda_{u;j}(dx) = \begin{cases} ba_jx^{b-1} - 1 - \exp(-G_{0j}x^b)(dx); & i = 0; j \neq 0; \\ ba_jx^{b-1} - 1 - \exp(-G_jx^b)(dx); & i \neq 0; j = i; \\ ba_jx^{b-1}(dx); & i \neq 0; i \neq j: \end{cases} \tag{3.46}$$

If $i = j = 0$, the propagation process of interference comes from the MUs with T_k, $I_0^{u_k} := f k\mathbf{X} - \mathbf{B}k\ g_{X2\ u_k;0}$ with intensity measure function

$$\Lambda_0^{u_k}(dx) = \frac{q_kA_{0k}}{\sum_{n=1}^{M} q_nA_{0n}} ba_0x^{b-1} - 1 - \exp(-G_{0k}x^b) : \tag{3.47}$$

If $i \neq 0$ and $j = 0$, the propagation process of interference comes from the MUs with T_k, $I_i^{u_k} := f k\mathbf{X} - \mathbf{B}k\ g_{X2\ u_k;0}$ with intensity measure function

$$\Lambda_i^{u_k}(dx) = \begin{cases} \hat{q}_kba_0x^{b-1} - 1 - \exp(-G_kx^b)(dx); & k = i; \\ \hat{q}_kba_0x^{b-1}; & k \neq i: \end{cases} \tag{3.48}$$

In a typical MBS cell, MUs use different resource blocks. If we randomly choose one MU from a MBS cell, the probability that the chosen MU has task T_k is $\hat{q}_k =$

$\frac{q_k \ u A_{0k}}{\sum_{n=1}^{M} u q_n A_{0n}}$, where A_{0k} denotes the probability that the typical MU with task T_k associates with MBS.

Lemma 6. *Laplace transform of MUs interference at the tagged MBS is* $L_{ujK=0}(s)$ *is*

$$L_{u;2jK=0}(s) = \prod_{j=0}^{N} L_{u;2j0;j}(s); \tag{3.49}$$

where $L_{u;2j0;j}(s)$ *denotes the Laplace transform of interference from tier* j *MU:*

$$L_{u;2j0;j}(s) \tag{3.50}$$

$$= \begin{cases} \exp\left(\frac{bs}{1-b} \sum_{k=1}^{K} \hat{q}_k a_0 \left(\mathsf{E}_{L_{0k}} \left[L_{0k}^{b}(1-) \right] C_b \left[\frac{s}{L_{0k}^{1}} \right] \right) \right), & j = 0; \\ \exp\left(\frac{bs}{1-b} a_j \left(\frac{p_s B_s}{p_m B_m} \right)^{1-b} \left(\mathsf{E}_{L_j} \left[L_j^{b}(1-) \right] C_b \left[\frac{s p_s}{L_j^{1} p_m} \right] \right) \right), & j \neq 0. \end{cases} \tag{3.51}$$

Proof. See Appendix D. □

By substituting (3.50) into (3.41), we can have

$$P_{0i}^{u}(\ ) = b G_{0_i} \int_{0}^{Z^1} x^{b-1} \exp\left(\frac{x^1}{p_u} \right)^{2} \left(G_{0i} x^b + \hat{q}_j a_0 \left(\mathsf{E}_{L_{0j}} \left[L_{0j}^{b}(1-) \right] C_b \left[\frac{x^1}{L_{0j}^{1}} \right] \right) \right) \left(\frac{b x^1}{1-b} \sum_{j=1}^{K} a_j \left(\frac{p_s B_s}{p_m B_m} \right)^{1-b} \left(\mathsf{E}_{L_j} \left[L_j^{b}(1-) \right] C_b \left[\frac{x^1 \ p_s B_s}{L_j^{1} \ p_m B_m} \right] \right) \right) \mathrm{d}x: \tag{3.52}$$

Lemma 7. *Laplace transform of MUs interference at the tagged type* i *SBS is given*

by

$$L_{ujK=i}(s) = \exp\left(\frac{bs}{1-b} \hat{q}_i \left(\frac{p_m B_m}{p_s B_s}\right)^{1-b} a_0 \mathsf{E}_{L_{0i}}\left[L_{0i}^{b(1-)}\right] C_b \left(\frac{sp_m B_m}{L_{0i}^{1-} p_s B_s}\right)^{"} \right.$$
$$\left. \sum_{k=1;k\in i} \hat{q}_k a_0 \mathsf{E}_{L_{0k}}\left[L_{0k}^{b(1-)}\right] C_b \left(\frac{s}{L_{0k}^{1-}}\right) \sum_{j=1} a_j \mathsf{E}_{L_j}\left[L_j^{b(1-)}\right] C_b \left(\frac{s}{L_j^{1-}}\right)^{!\#} \right) : \tag{3.53}$$

Proof. See Appendix E. □

Making use of (3.53), the analytical expression of $P_{0_i}^{d}(\ )$ is given by

$$P_{s_i}^{u}(\ ) = bG_i \int_0^{Z^1} x^{b-1} \exp\left(x^{1-} {}^{2}=p_u \left[G_i x^b \left(\frac{bx^{1-}}{1-b}\right) \sum_{j=1} a_j \mathsf{E}_{L_j}\left[L_j^{b(1-)}\right] C_b \left(\frac{x^{1-}}{L_j^{1-}}\right)^{"} \right]^{!\#} \right.$$
$$\left(\frac{bx^{1-}}{1-b}\right) \hat{q}_i \left(\frac{p_m B_s}{p_s B_s}\right)^{1-b} a_0 \mathsf{E}_{L_{0i}}\left[L_{0i}^{b(1-)}\right] C_b \left(\frac{x^{1-} p_m B_m}{L_{0i}^{1-} p_s B_s}\right)^{!}$$
$$\left. \left(\frac{bx^{1-}}{1-b}\right) \sum_{k=1;k\in i} \hat{q}_k a_0 \mathsf{E}_{L_{0k}}\left[L_{0k}^{b(1-)}\right] C_b \left(\frac{x^{1-}}{L_{0k}^{1-}}\right) \right) \mathrm{d}x: \tag{3.54}$$

3.2.2 Downlink Coverage Analysis

For the downlink transmission, the SINR at the typical MU associated with BS located at X for the access link is given by

$$\mathrm{SINR}_2^d = \frac{p_X H_d R_X}{{}^{2} + I_{d;2}}; \tag{3.55}$$

with

$$I_{d;2} = \sum_{\mathbf{x}2\ \Phi_{d;2}^{a}} p_x H_x k\mathbf{x}k \quad ; \tag{3.56}$$

where $\Phi_{d;2}^{a}$ denotes the point process of SBSs and MBSs using the same resource block as the serving BS of the typical MU.

The downlink coverage probability in Strategy II is given by

$$P_2^d(\ ) = \sum_{i=1}^{N} q_i \Pr(\mathrm{SINR}_2^d > \ j\text{Typical MU with task } T_i)$$
$$= \sum_{i=1}^{N} q_i \ \Pr(K=i)P_{s_i}^d(\ ) + \Pr(K=0)P_{0_i}^d(\ )\ ; \tag{3.57}$$

in which

$$P_{0_i}^d(\ ) = \Pr(\mathrm{SINR}_2^d > \ jK=0; \text{Typical MU with task } T_i)$$
$$= \mathsf{E}_{L_{0i}} \ L_{d;2jK=0} \ \frac{L_{0i}}{p_m} \ \exp(\ \frac{L_{0i}}{p_m}^2)\ ; \tag{3.58}$$

and

$$P_{s_i}^d(\ ) = \Pr(\mathrm{SINR}_2^u > \ jK=i; i>0)$$
$$= \mathsf{E}_{L_i}\left[L_{d;2jK=i}(L_i = P_s) \ \exp(\ L_i\ ^2 = p_s)\right]. \tag{3.59}$$

Lemma 8. *The Laplace transform of downlink interference $I_{d;2}$ at the type-k MU when the serving BS belongs to MBS and type-k SBS are given respectively by*

$$L_{d;2jL_{0k}=I}(s)$$
$$= \exp\left(\frac{bsI^{b\ 1}}{1\ \ b} \ p_m a_0 + p_s \ \frac{B_s p_s}{B_m p_m}^{b\ 1} a_k \ C_b \ \frac{sp_m}{I} \ \ 2 \sum_{j=1;j\in k}^{\ } \int_0^{1} \frac{x}{1+\frac{x}{sp_s}} \mathrm{d}x \right)\ ; \tag{3.60}$$

and

$$L_{d;2j\,L_k=I}(s) = \exp\left(-\frac{bsI^{b}}{1-b}\left(p_s a_k + p_m\left(\frac{B_m p_m}{B_s p_s}\right)^{b}\right)^{-1} a_0\, C_b\left(\frac{sp_s}{I}\right) \cdot 2\pi \sum_{j=1;j\neq k}^{M} \int_0^{\infty} \frac{x}{1+\frac{x}{sp_s}}\mathrm{d}x\right). \tag{3.61}$$

Proof. Since all of the tier j SBS ($j \neq k$) contribute to interference at the type-k MU, the Laplace transform of interference results from tier j SBS is easily to be obtained by PGFL. The final result is the product of Laplace transform of SBS-tier and MBS-tier. □

Making use of Lemma 8, we can deuce (3.58) and (3.59) into (3.62) and (3.63) as follows

$$P_{0i}^{d}(\ ) = bG_{0_i}\int_0^{\infty} x^{b-1}\exp\left(-\frac{x^{2}}{p_m} - G_{0i}x^{b}\cdot 2\pi\sum_{j=1;j\neq k}^{M}\int_0^{\infty}\frac{y}{1+\frac{p_m y}{x^{b} p_s}}\mathrm{d}y - \frac{bx^{b}}{1-b}\left(a_0 C_b(\ ) + \left(\frac{B_s}{B_m}\right)^{b}\left(\frac{p_s}{p_m}\right)^{b} a_i C_b\left(\frac{B_m}{B_s}\right)\right)\right)\mathrm{d}x; \tag{3.62}$$

$$P_{s_i}^{d}(\ ) = bG_i\int_0^{\infty} x^{b-1}\exp\left(-\frac{bx^{b}}{1-b}\left(a_i C_b(\ ) + \left(\frac{B_m}{B_s}\right)^{b}\left(\frac{p_m}{p_s}\right)^{b} a_0 C_b\left(\frac{B_s}{B_m}\right)\right) - \frac{x^{2}}{p_s} - G_i x^{b}\cdot 2\pi\sum_{j=1;j\neq k}^{M}\int_0^{\infty}\frac{y}{1+\frac{y}{x}}\mathrm{d}y\right)\mathrm{d}x. \tag{3.63}$$

The downlink coverage probability can be derived by substituting (3.62) and (3.63) into (3.57).

3.2.3 Average Delay

Without the consideration of backhaul link, the total bandwidth W is allocated to the typical MU based on the number of MUs associated its serving BS. The bandwidth obtained by a typical MU is given by

$$\begin{cases} W^a_{m;2} = \dfrac{W}{N^u_m}; & \text{MU connected to MBS;} \\ W^a_{s_i} = \dfrac{W}{N^u_{s_i}}; & \text{MU connected to type-}i\text{ SBS:} \end{cases} \tag{3.64}$$

The average uplink and downlink transmission time are

$$\begin{cases} T^{t;u}_{0i} = \dfrac{D}{P^u_{0i}(\ )W^a_{m;2}\log_2(1+\ \ )}; & \text{Type-}i\text{ MU-MBS;} \\ T^{t;u}_{s_i} = \dfrac{D}{P^u_{s_i}(\ )W^a_{s_i}\log_2(1+\ \ )}; & \text{Type-}i\text{ MU-SBS;} \end{cases} \tag{3.65}$$

and

$$\begin{cases} T^{t;d}_{0i} = \dfrac{D}{P^d_{0i}(\ )W^a_{m;2}\log_2(1+\ \ )}; & \text{Type-}i\text{ MU-MBS;} \\ T^{t;d}_{s_i} = \dfrac{D}{P^d_{s_i}(\ )W^a_{s_i}\log_2(1+\ \ )}; & \text{Type-}i\text{ MU-SBS:} \end{cases} \tag{3.66}$$

The arrival rate at MBS and type-i SBS are given by

$$\Lambda^2_m = \sum_{i=1}^{\ } \frac{A_{0i}q_i\ \ {}_uP^u_{0i}(\ )}{m}; \tag{3.67}$$

and

$$\Lambda^2_{s;i} = \frac{A_{s_i}\ \ {}_uP^u_{s_i}(\ )}{i}. \tag{3.68}$$

The average response time at MBS in Strategy II is $T^r_{m;2} = \frac{1}{m\ \ \frac{2}{m}}$. By replacing $\Lambda^1_{s;i}$ in (3.30) with $\Lambda^2_{s;i}$, we can derive the average response time at type-i SBS in Strategy II, namely $T^r_{s_i;2}$.

Thus, the average delay for the typical MU with T_i in Strategy II is given by

$$T^i_{\text{avg};2}(\ ) = A_{0i}\left[T^r_{m;2} + T^{t;u}_{0i}(\ ) + T^{t;d}_{0i}(\ )\right] + A_{s_i}\left[T^r_{s_i;2} + T^{t;u}_{s_i}(\ ) + T^{t;d}_{s_i}(\ )\right] : \qquad (3.69)$$

Chapter 4

Simulation Results

In this section, we present numerical results for the MEC-enabled 2-tier HetNet in the area of 1 km^2. The simulation parameters are: $_u = 100/km^2$, $B_m = 1$, $W = 60$Mhz, $= 0.6$, $p_m = 43$ dBm, $P_s = 33$ dBm, and $p_u = 23$ dBm. The service rate at MBS is 12 s^{-1}. We consider that there is three types of services (i.e., $fT_i ji = 1; 2; 3g$), and corresponding popularity probabilities are 0.2, 0.3, and 0.5 (i.e., $q_1 = 0.2; q_2 = 0.3; q_3 = 0.5$), respectively (except for Fig. 4.1 that considers two types of services). The buffer size at SBS is 5, and the service rate at SBSs is 4 s^{-1}. The other parameters are introduced when the corresponding figures are discussed.

4.1 Verification of Accuracy

In this section, we provide numerical simulation to validate the analytical expressions for coverage probability. In each case, we perform at least 10^5 times of the realizations of the positions of communication nodes and the channel. We consider that there are two types of services with popularity probability 0.3 and 0.7. The density of MBSs and SBSs are $10/km^2$ and $20/km^2$, respectively. In Fig. 4.1, we plot the uplink and the downlink coverage probability for Strategy I and Strategy II. The results show the accuracy of the derived analytical expressions for the coverage probabilities, which are perfectly matching with simulations.

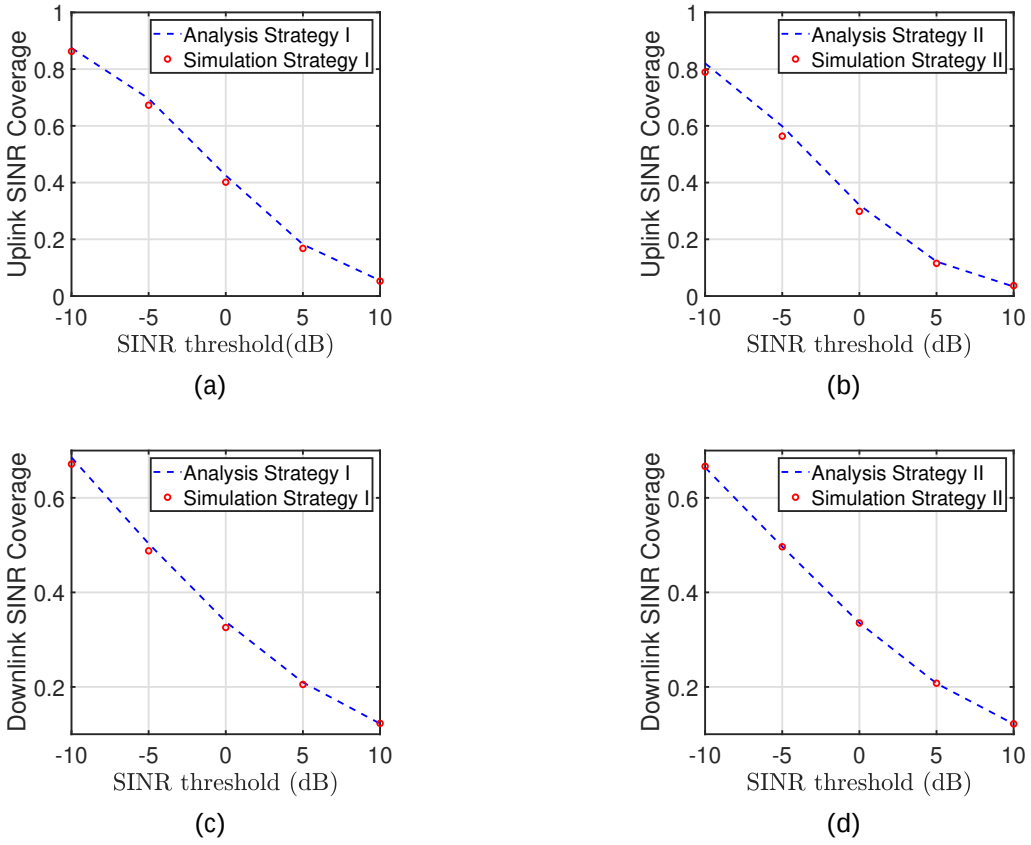

Figure 4.1: Validation of uplink and downlink coverage probability.

4.2 Effect of Bias Factor

Fig. 4.2 depicts how the bias factors in cell association affect the performance of the considered system via numerical results. The average delay first decreases as the B_s increases, then increases with the increase of B_s. For a given type of task, that the average delay in Strategy I is always shorter than Strategy II. This is reasonable since the IAB setting in Strategy I enables computation load migration from SBSs to MBSs, which is beneficial to reduce the average response time. Recalling that the popularity probability of task 3 is larger than task 1, Fig. 4.2 (a) indicates that MUs with a smaller popularity probability task can always achieve a smaller average delay in Strategy I. On the other hand, for Strategy II, this trend only holds when the bias factor is above a specific value, while the opposite trend takes place when the bias

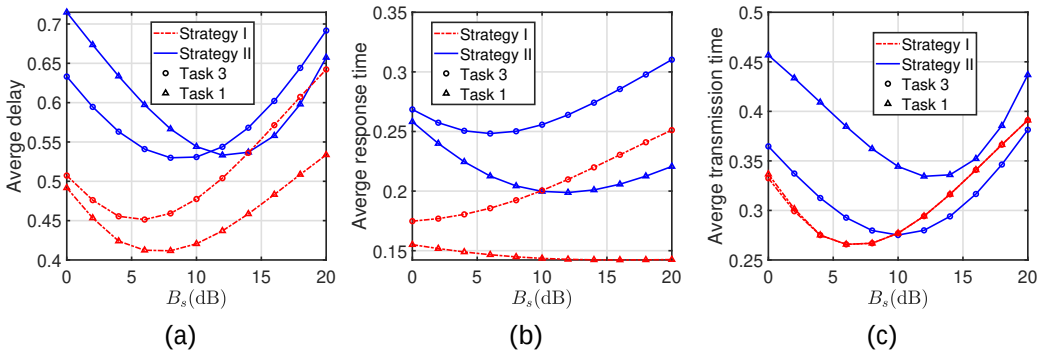

Figure 4.2: Effect of Bias Factor.

factor is below this value.

4.3 Effect of BS Density

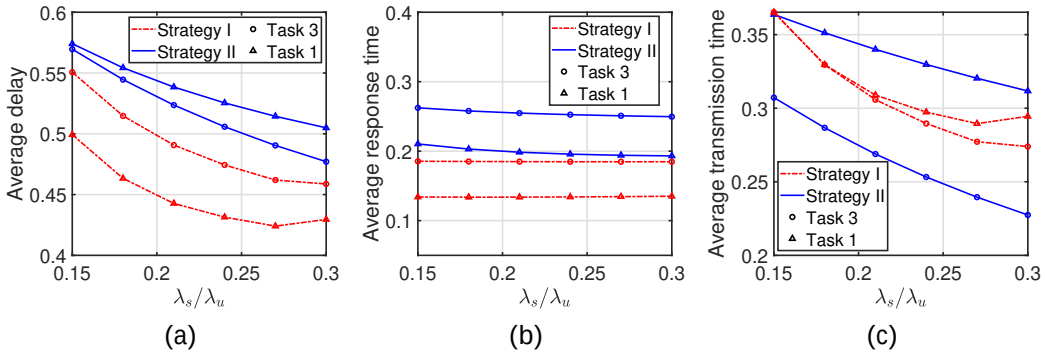

Figure 4.3: Effect of SBS Density.

Fig. 4.3 shows the impacts of the SBS density on the system performance. In Strategy I, the MU with task-1 can achieve a lower average delay than the MU with task-3, while the reverse is true in Strategy II. For Strategy I, recalling the average delay expression given in (3.34), the MU with task 1 has a higher probability of utilizing the more powerful computation capacity at the MBS via the backhaul link, which is able to further reduce the average response time of its task. However, as shown in Fig. 4.3 (a) and Fig. 4.3 (c), when $_s= _u$ exceeds a certain level, the heavy-loaded backhaul link will degrade the average transmission time as well as the

average delay for the MU with task 1. As for Strategy II, since there is no backhaul link bottleneck, the average delay is a monotonically decreasing function of $_s = _u$. Moreover, due to the higher density of type-3 SBSs, the transmission distance for the access link of the type-3 MU is shorter than the type-1 MU, which makes the type-3 MU achieve a much shorter average transmission time. As a result, in Strategy II, type-3 MUs outperform type-1 MUs.

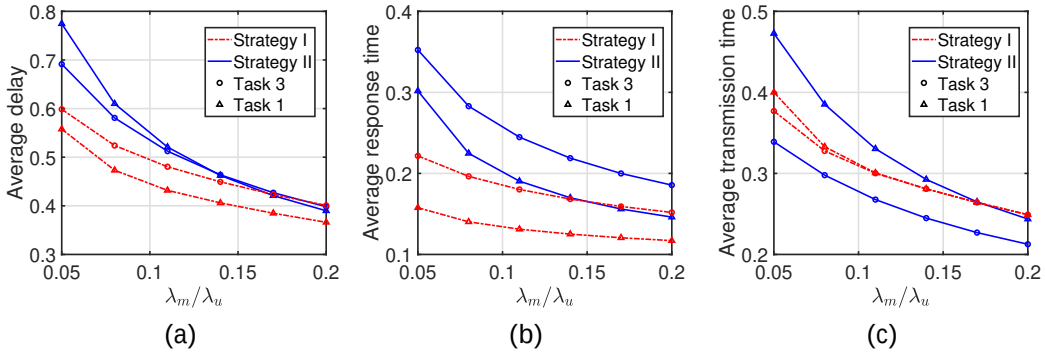

Figure 4.4: Effect of MBS Density.

Compared with the SBS density, from Fig. 4.4, we can observe that increasing MBS density can achieve a more remarkable performance improvement. Fig. 4.4 (a) shows that the average delay decreases as $_m = _u$ increases in all considered cases. Moreover, the average delay of a MU with task 1 becomes smaller than that of a MU with task 3 only after $_m = _u$ exceeds a certain level.

4.4 Effect of SBS Service Rate

Fig. 4.5 demonstrates the system performance versus the SBS service rate. The average delay can be regarded as a monotonically decreasing function of $_s$ in all considered cases. When $_s$ is small, the average response time for the type-3 MU is larger than that of the type-1 MU in both considered strategies, since the computation load at the type-1 SBS is smaller than the load at the type-3 SBS and $P_{o\ d}$ is smaller than 1. After certain values of $_s$, the average delay of the type-3 MU starts to be

smaller than type-1 MU. This is reasonable since a large $_s$ will lead to $P_{o\ d} = 1$, which causes the decrease of computation load at both SBSs and MBSs.

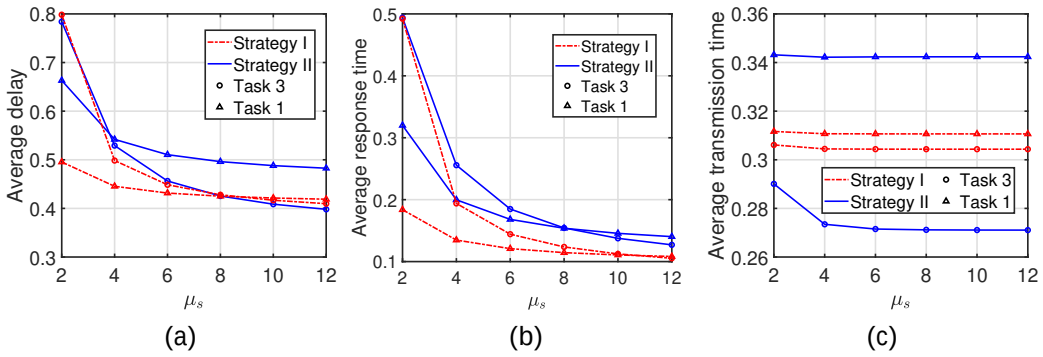

Figure 4.5: Effect of SBS Service Rate.

4.5 Effect of SBS Buffer Size

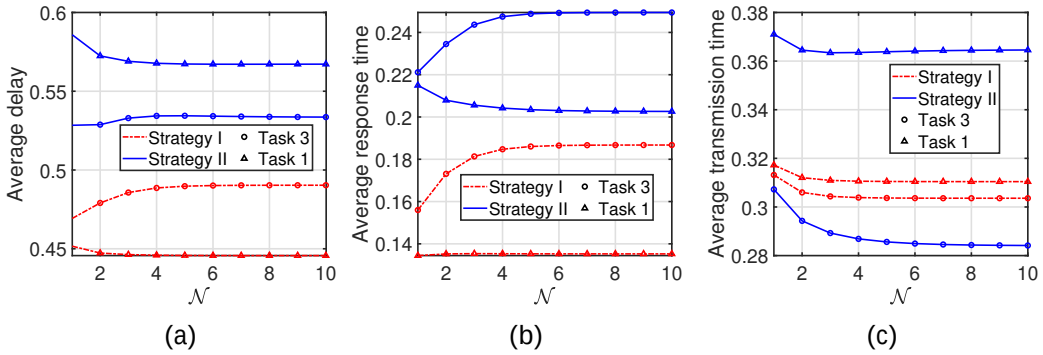

Figure 4.6: Effect of SBS Buffer Size.

Fig. 4.6 shows the changes of average delay versus the buffer size at SBSs. In all considered strategies, before the average delay getting converged, the average delay for MUs with task 1 and task 3 is decreasing and increasing with the increase of N, respectively. Moreover, the average delay in Strategy I is smaller than that of Strategy II for both types of MUs. With increasing N, the typical MU will experience a longer waiting time at its serving SBS. Meanwhile, $P_{o\ d}$ increases as N increases, which can alleviate the computing load at both MBS-tier and SBS-tier. Since the type-1

MU has a larger probability of associating with the MBS-tier and the response time at MBS-tier is smaller than type-1 SBS, the average response time decreases as N increases. On the contrary, the average response time for the type-3 MU increases with the increase of N. Since the computing load at the type-1 SBS in Strategy I is much lighter, N has a negligible impact on average response time for type-1 MUs in Strategy I.

Chapter 5

Conclusion & Future Work

In this thesis, we formulated a two-tier MEC HetNet spatial model, which consisted of the multi-type MUs with the request of different service types and the two-tier MEC servers with different computing capacities. Due to the limited resource, SBS-tier MEC servers can only cache a specific type of service, and their computing buffer is finite. We studied two strategies corresponding to two different settings: 1) Strategy I for an IAB-enabled MEC HetNet, and 2) Strategy II for traditional MEC HetNet. By using tools from stochastic geometry and queueing theory, we first derived the analytical expressions for coverage probability, and then analyzed the average delay for all considered strategies. Finally, we discussed and compared the performance of the two proposed strategies by extensive simulations.

The current work and results presented in this thesis can be extended in several future research directions. In this thesis, we considered that each SBS only can cache one specific service and the storage size at SBSs are identical. However, considering multiple services caching makes the average delay analysis more challenging but is of practical relevance. In general, the BS with a larger storage size implies that more services can be cached. The effect of the heterogeneity and diversity of storage sizes is not researched yet. Moreover, the optimal caching and offloading strategies design in the heterogeneous MEC networks is still an open issue, especially in the IAB-enabled networks.

REFERENCES

[1] U. Drolia, R. Martins, J. Tan, A. Chheda, M. Sanghavi, R. Gandhi, and P. Narasimhan, "The case for mobile edge-clouds," in *Proc., IEEE International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, 2013, pp. 209–215.

[2] 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, 2015.

[3] H. Wu, Q. Wang, and K. Wolter, "Tradeoff between performance improvement and energy saving in mobile cloud offloading systems," in *Proc. IEEE ICC workshops*, 2013, pp. 728–732.

[4] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268–4282, 2016.

[5] J. Ren, G. Yu, Y. Cai, Y. He, and F. Qu, "Partial offloading for latency minimization in mobile-edge computing," in *Proc., IEEE Global Communications Conference(GLOBECOM)*, 2017, pp. 1–6.

[6] 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, 2017.

[7] C. You and K. Huang, "Multiuser resource allocation for mobile-edge computation offloading," in *Proc., IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[8] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590–3605, 2016.

[9] C. You, K. Huang, and H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1757–1771, 2016.

[10] J. Xu, L. Chen, and P. Zhou, "Joint service caching and task offloading for mobile edge computing in dense networks," in *Proc., IEEE INFOCOM*, 2018, pp. 207–215.

[11] K. Poularakis, J. Llorca, A. M. Tulino, I. Taylor, and L. Tassiulas, "Joint service placement and request routing in multi-cell mobile edge computing networks," in *Proc., IEEE INFOCOM*, 2019, pp. 10–18.

[12] D. Soldani and A. Manzalini, "Horizon 2020 and beyond: On the 5G operating system for a true digital society," *IEEE Veh. Technol. Mag.*, vol. 10, no. 1, pp. 32–42, Mar. 2015.

[13] I. Union, "IMT traffic estimates for the years 2020 to 2030," *Report ITU*, pp. 2370–0, 2015.

[14] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[15] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 2016.

[16] H. A. Willebrand and B. S. Ghuman, "Fiber optics without fiber," *IEEE Spectr.*, vol. 38, no. 8, pp. 40–45, 2001.

[17] O. Teyeb, A. Muhammad, G. Mildh, E. Dahlman, F. Barac, and B. Makki, "Integrated access backhauled networks," in *Proc., IEEE 90th Vehicular Technology Conference (VTC-Fall)*, Sep. 2019.

[18] H. Raza, "A brief survey of radio access network backhaul evolution: part I," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 164–171, Jun. 2011.

[19] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmwaves," in *Proc., IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–7.

[20] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*, 2nd ed. San Diego: Academic Press, 2020.

[21] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, 2011.

[22] 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, 2012.

[23] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," vol. 19, no. 1, pp. 167–203, 2017.

[24] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, 2013.

[25] M. Liu, F. R. Yu, Y. Teng, V. C. Leung, and M. Song, "Computation offloading and content caching in wireless blockchain networks with mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 008–11 021, 2018.

[26] S.-W. Ko, K. Han, and K. Huang, "Wireless networks for mobile edge computing: Spatial modeling and latency analysis," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5225–5240, 2018.

[27] Y. Gu, C. Li, B. Xia, D. Xu, and Z. Chen, "Modeling and performance analysis of stochastic mobile edge computing wireless networks," in *Proc., IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–5.

[28] C. Park and J. Lee, "Mobile edge computing-enabled heterogeneous networks," available online: https://arxiv.org/abs/1804.07756.

[29] S. Mukherjee and J. Lee, "Edge computing-enabled cell-free massive mimo systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2884–2899, 2020.

[30] J. Sztrik, "Basic queueing theory," *University of Debrecen, Faculty of Informatics*, vol. 193, pp. 35–35, 2012.

[31] E. Gilbert, "Random subdivisions of space into crystals," *The Annals of mathematical statistics*, vol. 33, no. 3, pp. 958–972, 1962.

[32] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5G: Performance analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2669–2684, 2019.

[33] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," available online: https://arxiv.org/abs/1604.03183.

[34] 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.

[35] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," vol. 59, no. 11, pp. 3122–3134, nov 2011.

[36] 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.

APPENDICES

A Proof of Lemma 1

$$
\begin{aligned}
A_k &= \Pr\left[\, p_k B_k D_k \;>\; p_j B_j D_j \;\; ; k \neq j \right] \\
&= \int_0^{\infty} \left[1 - F_{D_j}\left(\frac{p_j B_j}{p_k B_k} x \right) \right] f_{D_k}(x)\,\mathrm{d}x \\
&= \int_0^{\infty} 2 \left(\lambda_k x \exp\left[-x^2 \left(\lambda_k + \lambda_j \left(\frac{p_j B_j}{p_k B_k} \right)^2 \right) \right] \right) \mathrm{d}x \\
&= \frac{\lambda_k}{\lambda_k + \lambda_j (p_j B_j / p_k B_k)^2}. \qquad \text{(A.1)}
\end{aligned}
$$

B Proof of Lemma 2

The complementary cumulative distribution function (CCDF) of $L_{a;1}$ conditioned on serving tier being k is

$$\begin{aligned}\Pr(L_{a;1} > l | K = k) &= \frac{\Pr(D_k > l; p_k B_k D_k^{-} > p_j B_j D_j^{-})}{A_k} \\ &= \frac{\mathsf{E}_{D_k > l}\left[\exp\left(-\sum_j \frac{p_j B_j}{p_k B_k}^{2/\alpha} D_k^2\right)\right]}{A_k} \\ &= \frac{b a_k}{A_k} \int_l^{\infty} x^{b-1} \exp\left(-x^b G_k\right) \mathrm{d}x. \end{aligned} \tag{B.1}$$

Taking the derivative of the CCDF leads to the final expression in Lemma 2.

C Proof of Lemma 3

We know that $L_{u;1jK=k}(s) = \prod_{j=1}^{Q} L^{j}_{u;1jK=k}(s)$. The Laplace transform of the interference at the typical tier k BS located at $\mathbf{B}$ from MU associated with tier j BSs is given by

$$L^{j}_{u;1jK=k}(s) = \mathrm{E}\left[\sum_{X \in \Phi_j \setminus a_{u_j;1}} (R_X)^{-\alpha} H_X \|X - \mathbf{B}\|^{-\alpha}\right]. \tag{C.1}$$

Since $\{H_X\}$ are i.i.d exponential random variables, (C.1) can be rewritten as

$$L^{j}_{u;1jK=k}(s) = \mathrm{E}\left[\prod_{X \in \Phi_j \setminus a_{u_j;1}} \frac{1}{1 + sL_X\|X - \mathbf{B}\|^{-\alpha}}\right]. \tag{C.2}$$

By using Remark 1 and probability generating functional (PGFL), we can deuce (C.2) into

$$\begin{aligned} L^{j}_{u;1jK=k}(s) &= \mathrm{E}\left[\prod_{X \in \Phi_j \setminus a_{u_j;1}} \mathrm{E}_{L_X}\left[\frac{1}{1 + sL_X X^{-\alpha}}\right]\right] \\ &= \exp\left(-\int_{x>0} \mathrm{E}_{L_a}\left[\frac{1}{1 + (sL)^{-1} x^{\alpha}}\right] j\{K_X = j; L < \frac{p_j B_j}{p_k B_k} x^{\alpha}\} \Lambda^{j}_{u;k}(\mathrm{d}x)\right). \end{aligned} \tag{C.3}$$

By employing a change of variable $t = \left(\frac{xP_j B_j}{P_k B_k L}\right)^{\frac{2}{\alpha}}$ and the definition of hypergeometric

function, we can rewrite (C.1) as

$$
\begin{aligned}
& L^{j}_{u;1jK=k}(s) \\
& = \exp\left(\mathsf{E}_{L_a jK=j}\left[a_j \left(\frac{p_k B_k L_a}{p_j B_j} \right)^{b} \int_1^{Z^1} \frac{\mathrm{d}t}{1 + L_a^{(1)} t^{1=b} p_k B_k / (p_j B_j s)} \right] \right) \\
& = \exp\left(\frac{bs}{1-b} \left(\frac{p_j B_j}{p_k B_k} \right)^{1-b} a_j \, \mathsf{E}_{L_a jK=j}\left[L_a^{b} \, {}^{(1)}C_b \left(\frac{s p_j B_j}{p_k B_k L_a^{1}} \right) \right] \right) : \qquad \text{(C.4)}
\end{aligned}
$$

D Proof of Lemma 6

Let $\mathcal{L}_{I_0^{u_k}}(s)$ denote the interference from type-k MUs associated with MBSs, we can have

$$\mathcal{L}_{u;2j0;0}(s) = \prod_{k=1}^{\infty} \mathcal{L}_{I_0^{u_k}}(s). \tag{D.1}$$

By replacing the intensity measure function and PDF of path loss in the derivation of (3.12) with results given in Assumption 3.2.1 and (3.43), the expression of $\mathcal{L}_{I_0^{u_k}}(s)$ can be written as

$$\exp\left(\frac{bs}{1-b}\hat{q}_k a_0 \mathsf{E}_{L_{0k}}\left[L_{0k}^{b}{}^{(1-)}\right] C_b \left[\frac{s}{L_{0k}^{1}}\right]^{-}\right). \tag{D.2}$$

This proof is concluded by substituting (D.2) into (D.1).

E Proof of Lemma 7

Let $\mathcal{L}_{u;2ji;j}(s)$ denote the interference from the MU associated with tier j BS at the typical type-i SBS, we can have

$$\mathcal{L}_{u;2jK=i}(s) = \prod_{j=0}^{K} \mathcal{L}_{u;2ji;j}(s). \tag{E.1}$$

Let $\mathcal{L}_{I_i^{u_k}}(s)$ denote the interference from the type-k MU associated with MBS at the type-i SBS, then

$$\mathcal{L}_{u;2ji;0}(s) = \prod_{k=1}^{K} \mathcal{L}_{I_i^{u_k}}(s). \tag{E.2}$$

By using Assumption 3.2.1 and following by the derivation of (3.12), the expression of $\mathcal{L}_{I_i^{u_k}}(s)$ is given by

$$\mathcal{L}_{I_i^{u_k}}(s) = \begin{cases} \exp\left(\frac{bs}{1-b} \frac{q_k A_{0k}}{\sum_{n=1}^{K} q_n A_{0n}} \left(\frac{P_m B_m}{P_s B_s} \right)^{1-b} a_0 \, \mathrm{E}_{L_{0k}} \left[L_{0k}^{b} \right] C_b \left(\frac{P_m s}{L_{0k}^{1} P_s} \right) \right), & k = i; \\ \exp\left(\frac{bs}{1-b} q_k a_0 \mathrm{E}_{L_{0k}} \left[L_{0k}^{b} \right] C_b \left(\frac{s}{L_{0k}^{1}} \right) \right), & k \neq i. \end{cases} \tag{E.3}$$

Similarly, the interference from MUs associated with type-k SBSs at the typical type-i SBS is given by

$$\mathcal{L}_{u;2ji;k}(s) = \exp\left(\frac{bs}{1-b} a_k \mathrm{E}_{L_k} \left[L_k^{b} \right] C_b \left(\frac{s}{L_k^{1}} \right) \right). \tag{E.4}$$

This proof is concluded by substituting (E.2) and (E.4) into (E.1).

F Papers Submitted and Under Preparation

Yongqiang Zhang, Mustafa A. Kishk, and Mohamed-Slim Alouini, "Computation Offloading and Service Caching in Heterogeneous MEC Wireless Networks", *IEEE Transactions on Mobile Computing*, submitted.

Yongqiang Zhang, Mustafa A. Kishk, and Mohamed-Slim Alouini, "A Survey on Integrated Access and Backhaul Networks", *Frontiers in Communications and Networks*, to appear.