Delay Analysis of Unsaturated Heterogeneous Omnidirectional–Directional Small Cell Wireless Networks: The Case of RF-VLC Coexistence

Sihua Shao, *Student Member, IEEE*, and Abdallah Khreishah, *Member, IEEE*

***Abstract*— The coexistence of omnidirectional small cells (OSCs), such as RF small cells, and directional small cells (DSCs), such as visible-light communication cells, is investigated. The delay of two cases of such heterogeneous networks is evaluated. In the first case, resource allocated OSCs, such as RF femtocells, are considered. In the second case, contention-based OSCs, such as WiFi access point, are studied. For each case, two configurations are evaluated. In the first configuration, the non-aggregated scenario, any request is either allocated to OSC or DSC. While in the second configuration, the aggregated scenario, each request is split into two pieces, one is forwarded to OSC and the other is forwarded to DSC. For the first case, under Poisson request arrival process and exponential distribution of request size, the optimal traffic allocation ratio is derived for the non-aggregated scenario and it is mathematically proved that the aggregated scenario provides lower minimum average system delay than that of the non-aggregated scenario. For the second case, the average system delay is derived for both non-aggregated and aggregated scenarios, and extensive simulation results imply that, under certain conditions, the non-aggregated scenario outperforms the aggregated scenario due to the overhead caused by contention.**

***Index Terms*— Heterogeneous network (HetNet), delay, omnidirectional small cell (OSC), directional small cell (DSC), RF femtocall, WiFi, visible light communications (VLC), link aggregation.**

I. Introduction

DEMAND for ubiquitous wireless connectivity continues to grow due to the trend towards an always on culture, broad interest in mobile multimedia, and advancement towards the Internet of things. This demand stems from a multifaceted growth in the number of networked devices and the per-device data usage from novel applications (e.g., HD video, augmented reality, and cloud-based services). Forecasts from Cisco show Internet video accounting for 80% of all consumer Internet traffic by 2019 [1] while Qualcomm and Ericsson expect between 25 and 50 billion connected devices by 2020 [2], [3]. Next generation, or 5G, wireless networks will be challenged to provide the capacity needed to meet this growing demand. Compared to peak performance goals of previous generations, 5G goals include increasing the expected performance across non-uniform geographic traffic distributions. In particular, additional capacity is needed in dense urban environments and indoor environments where approximately 70% of IP-traffic occurs [4].

Heterogeneous wireless network, as a method to incorporate different access technologies, contains the potential capabilities of improving the efficiency of spectral resource utilization. Traffic offloading to omnidirectional small cells (OSCs), such as RF femtocells and WiFi WLANs, has already become an established technique for adding capacity to dense environments where macrocells are overloaded. Ultra-dense distributed directional small cells (DSCs), deployed in indoor environments, can supplement OSCs in areas like apartment complexes, coffee shops, and office spaces where device density and data demand are at their highest. These DSCs can be implemented by technologies like microwave [5], mmwave [6] and optical wireless. Optical wireless (OW) communication - specifically visible light communication (VLC) or LiFi [7] - is a directional communication technology that has gained interest within the research community in recent years. As an excellent candidate for 5G wireless communication, VLC provides ultra wide bandwidth and efficient energy utilization [8]. However, the weaknesses of VLC is the vulnerability to obstacles when compared to the omnidirectional RF communication.

In this work, we consider two cases of heterogenous OSC-DSC networks. One case is the coexistence of resource allocated OSCs (RAOSCs) and DSCs. A typical application of RAOSC is the RF femtocells [9], which are owned/controlled by a global entity (i.e., service provider). Therefore, interference can be mitigated in the provisioning process and multiple adjacent RF femtocells can perform downlink data transmission simultaneously without contention. This non-contention issue will be further discussed in Section II. The other case is the heterogenous network incorporating contention-based OSCs (CBOSCs) and DSCs. In contrast to RAOSC, CBOSC (such as WiFi AP) is purchased by local entities (i.e. home/business owners) and deployed in an ad-hoc manner such that interference is not planned. Particularly, WiFi networks employs the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocols to schedule the contention process. DSCs have a large reuse factor such

Manuscript received March 13, 2016; revised July 7, 2016; accepted September 17, 2016. Date of publication October 3, 2016; date of current version December 8, 2016. The associate editor coordinating the review of this paper and approving it for publication was C.-H. Lee.

The authors are with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: ss2536@njit.edu; abdallah@njit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2016.2614822

that the spectrum reuse can be easily implemented even in an indoor environment. Without the loss of generality, we use OSC and DSC notations instead of RF and VLC in the following description.

Many current research efforts have been paid towards developing heterogeneous networks incorporating both OSC and DSC. A protocol, considering OFDMA, vertical handover (VHO) and horizontal handover (HHO) mechanisms for mobile terminals (MTs) to enable the mobility of users among different VLC APs and OFDMA system, is proposed in [10]. The authors define a new metric, called spatial density, to evaluate the capacity of the heterogeneous network under the assumption of the Homogenous Poisson Point Process (HPPP) distribution of MTs. In [11], load balancing for hybrid VLC and WiFi system is optimized by both centralized and distributed resource-allocation algorithms while achieving proportional fairness. In [12], different RF-VLC heterogeneous network topologies, such as symmetric non-interfering, symmetric with interference and asymmetric, are briefly discussed. In [13], taking the advantage of wide coverage of RF and spatially reuse efficiency of VLC, a hybrid RF and VLC system is proposed to improve per user average and outage throughput.

Regarding the bandwidth aggregation, a thorough survey of approaches in heterogeneous wireless networks has been presented in [14]. The challenges and open research issues in the design of bandwidth aggregation system, ranging from MAC layer to application layer, have been investigated in detail. The benefits of bandwidth aggregation includes increased throughput, improved packet delivery, load balancing and seamless connectivity. This work also validates the feasibility of the heterogeneous OSC-DSC networks proposed here based on bandwidth aggregation. In [15], users connect to WiFi and VLC simultaneously. A parallel transmission MAC (PT-MAC) protocol containing CSMA/CA algorithm and the concept of parallel transmission are proposed. This protocol supports fairness among users in the hybrid VLC and WLAN network.

The above-mentioned works, which are primarily simulation-based studies, do not provide system-level implementation of the WiFi-LiFi systems. In our previous work [16]–[18], an aggregated WiFi-VLC system is presented and implemented using WiFi/VLC equipment and Linux Bonding driver. The realized WiFi-LiFi system aggregates a single WiFi link and a single VLC link, and provides improved throughput. This paper theoretically investigates *system delay*, a critical QoS metric especially for multimedia applications [19]. Here, system delay is defined as the amount of the time from the instant the request arrives at the AP to the instant that it successfully departs from the AP.

In [19], delay modeling of a hybrid WiFi-VLC system has been investigated. Each WiFi and VLC queue is observed as an M/D/1 queue, and the capacities with respect to the unstable delay points of WiFi only, asymmetric WiFi-VLC and hybrid WiFi-VLC systems are compared. An analytic model for evaluating the queueing delays and channel access times at nodes in 802.11 based WiFi networks is presented in [20]. The model provides closed form solutions for obtaining the values of the delay and queue length. This is done by modeling each node as a discrete time G/G/1 queue. However, these works do not investigate the delay modeling of a system with bandwidth aggregation. In other words, most of the existing heterogeneous works only study the networks without bandwidth aggregation (i.e. one request is either forwarded to one access technology or the other).

This paper characterizes the system delay of two cases of heterogeneous OSC-DSC wireless networks: (i) RAOSC-DSC; (ii) CBOSC-DSC. For each case, two configurations are taken into consideration. One of them is based on bandwidth aggregation and the other is not. The potential gain in terms of the minimum average system delay through aggregating the bandwidth of OSC and DSC is also evaluated. To the best of our knowledge, this work is the first to quantify the system performance of aggregation with respect to minimum average system delay. Note that investigating the delay performance of a heterogeneous system when *aggregation* is considered, is our major contribution, which differs from other existing works. The main contributions of this work include the following: (i) for the heterogeneous RAOSC-DSC wireless network, a generalized characterization of the system without bandwidth aggregation is derived in terms of the optimal ratio of traffic allocation and the minimum average system delay and a near-optimal characterization of the minimum average system delay of the system that utilizes bandwidth aggregation is proposed; (ii) for the heterogeneous RAOSC-DSC wireless network, it is also theoretically proved that the minimum average system delay of the system based on bandwidth aggregation is lower when compared to that of the system without bandwidth aggregation; (iii) for the heterogeneous CBOSC-DSC wireless network, the average system delay is derived for both the system without bandwidth aggregation and the system with bandwidth aggregation; (iv) for the heterogeneous CBOSC-DSC wireless network, extensive simulations are also conducted to indicate that under certain conditions, the system without bandwidth aggregation outperforms the system with bandwidth aggregation in terms of minimum average system delay.

II. System Model

A. Parameters

A recent measurement study [21] on traces of 3785 smart phone users from 145 countries over a four-month period shows that the ratio of download traffic to its upload traffic is 20:1. Therefore, in this paper, we investigate the downlink system delay of two cases of heterogeneous OSC-DSC wireless access networks:

- Case 1: heterogeneous RAOSC-DSC network,
- Case 2: heterogeneous CBOSC-DSC network.

Fig. 1 illustrates the network architecture for case 1. In the system model suggested, there are one RAOSC AP and N_1 DSC APs. Since OSC APs do not contend with each other, under homogeneous traffic distribution, the delay analysis of a single RAOSC AP can be easily extended to that of multiple RAOSC APs. Due to the fact that the DSCs have a large reuse factor [22], it is rational to assume that all the DSC

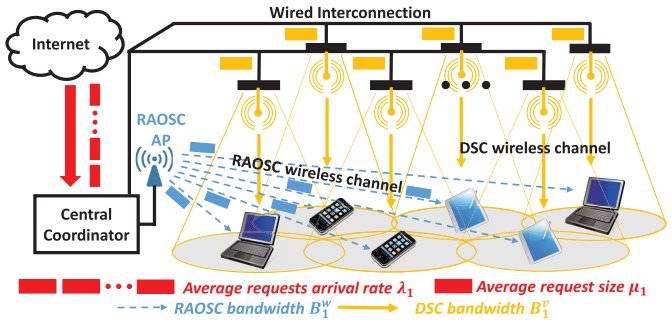


Fig. 1. Heterogeneous RAOSC-DSC network architecture.

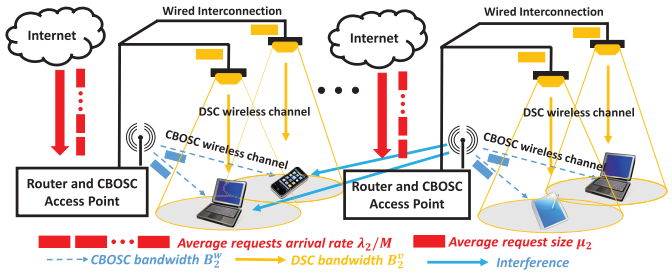


Fig. 2. Heterogeneous CBOSC-DSC network architecture.

links can be active simultaneously with negligible interference among them. Under the homogeneous traffic assumption, the traffic assigned to different DSC APs is evenly distributed. The requests arrival process to the central coordinator is a Poisson process [20], [23] with rate λ_1. One request here means one download session (e.g. a photo, a webpage, a video) from the Internet. For priority system [24], where each session forms a flow with a certain priority level and packets of lower priority start transmission only if no higher priority packet is waiting, Poisson arrival process is applicable due to the independency among a large number of arrival of requests. Since the requests are from different independent sources, it is assumed that the size of each request is exponentially distributed with mean μ_1. The downlink capacities of the RAOSC and the DSC are B_1^w and B_1^v, respectively, where $B_1^w < B_1^v$.

Fig. 2 illustrates the network architecture for case 2. In this case, there are M CBOSC APs and N_2 DSC APs, where $N_2 > M$. All of the M CBOSC APs are located in a single contention domain. The MAC scheme considered is IEEE 802.11 [25], which is implemented by using a Distributed Coordination Function based on the CSMA/CA protocol. The RTS/CTS exchange scheme, which is utilized to address the "hidden node" problem, is also taken into account. The 802.11 configurations will be described in details in Section IV. The blockage property of DSC is modeled as a successful transmission probability P_{succ} for each request. The whole request will be retransmitted once the transmission fails. The Ack-enabled mechanism [26] for DSC is considered. Under the homogeneous traffic assumption, the traffic assigned to different CBOSC and DSC APs are evenly distributed. The requests arrival process to each AP is a Poisson process with rate λ_2/M. The size of each request is exponentially distributed with mean μ_2. The downlink capacities of the CBOSC and the DSC are B_2^w and B_2^v, respectively.

TABLE I
THE DEFINITION OF SOME OF THE SYMBOLS

$\lambda_1(\lambda_2)$	Total request arrival rate for the heterogeneous RAOSC-DSC network(the heterogeneous CBOSC-DSC network)
$\mu_1(\mu_2)$	Mean size of request for the heterogeneous RAOSC-DSC network(the heterogeneous CBOSC-DSC network)
$B_1^w(B_2^w)$	RAOSC(CBOSC) bandwidth
$B_1^v(B_2^v)$	Case 1 DSC(Case 2 DSC) bandwidth
M	The number of CBOSC APs
$N_1(N_2)$	The number of Case 1 DSC(Case 2 DSC) APs
P_{succ}	The successful transmission rate for DSC links
$\alpha_1(\alpha_2)$	The percentage of traffic allocated to RAOSC(CBOSC)
$\beta_1(\beta_2)$	The proportion of the size of each request assigned to RAOSC(CBOSC)

For two cases of heterogeneous OSC-DSC wireless access networks, the system delay D performance is studied for two configurations: i) non-aggregated scenario and ii) aggregated scenario. In the non-aggregated scenario, any request is either allocated to the RAOSC/CBOSC or the DSC. In the aggregated scenario, each request is split into two pieces. One of them is forwarded to the RAOSC/CBOSC while the other is forwarded to one of the DSC APs. In the paper, one request means one download session (e.g. a photo, a webpage, a video) from the Internet. For the aggregated scenario, assume one request consists of 1000 packets, to implement aggregation, these 1000 packets are split into two sets - one contains β portion of packets and the other contains the remaining $(1-\beta)$ portion of packets. To aggregate the bandwidth of OSC and DSC, the β portion of packets will be transmitted through the OSC channel and simultaneously the $(1-\beta)$ portion of packets will be sent via the DSC channel. To implement such a heterogeneous system, one central coordinator is needed. The central coordinator is an additional device encompassing multiple functionalities, such as collecting the location and channel information of all APs and user terminals, computing the optimal traffic allocation ratio, and forwarding the data traffic to different APs. Most of the hybrid RF-VLC papers [13], [18], [19], [27], [28] have utilized the central coordinator in the system for performing the traffic allocation functionality. Also the cost of the central coordinator is usually cheap, such as banana pi [29]. As a result, the system delay of each request is the maximum of i) time spent by the piece of request in RAOSC/CBOSC and ii) time spent by the piece of request in DSC. The system delay of the requests in RAOSC, CBOSC and DSC are represented by D_{RAOSC}, D_{CBOSC} and D_{DSC}, respectively. New metrics $\alpha_1(\alpha_2)$ and $\beta_1(\beta_2)$ are defined for two cases, to represent the traffic allocation ratio and request splitting ratio for non-aggregated and aggregated scenarios, respectively. These four factors will be discussed in detail in Section III and Section IV. The main notations are summarized in Table. I.

B. Overview of Typical Omnidirectional Non-Contention and Contention Wireless Networks

As we discussed earlier, a typical example of omnidirectional non-contention wireless network is the

RF femtocell network. RF femtocell is a small and low-power cellular base station, typically designed for coverage and capacity improvement. One of the most critical issues from deploying RF femtocells is the potential interference among femtocells and macrocells [30]. However, femtocells can incorporate interference mitigation techniques-detecting macrocells, adjusting power and scrambling codes accordingly [31] to eliminate the potential interference. The interference management among neighboring femtocells and among femtocells and macrocells are also investigated in [32]. Clustering of femtocells [33], [34], fractional frequency reuse (FFR) and resource partitioning [35], [36], and cognitive approaches [37] can be employed to mitigate the inter-femtocells interference. Since femtocells are deployed by service provider, who has the priority of manipulating the frequency, power, and location of all the femtocells, the above-mentioned interference mitigation techniques can be applied without contention issue. With interference issue solved, the neighboring RF femtocells can perform downlink data transmission at the same time without worrying about the contention process even at the cell edge.

For omnidirectional contention-based wireless network, a typical example is WiFi network. Since each WiFi AP is normally deployed independently without coordination with the neighboring WiFi APs, the interference among WiFi APs will inevitably trigger the contention process when the adjacent WiFi APs perform the downlink data transmission simultaneously. The CSMA/CA based MAC protocol of IEEE 802.11 [25] is designed to mitigate the collisions due to multiple WiFi APs transmitting on a shared channel. In a WiFi network employing CSMA/CA MAC protocol, each WiFi AP with a packet to transmit will first sense the channel during a Distributed Inter-frame Space (DIFS) to decide whether it is idle or busy. If the channel is idle, the WiFi AP proceeds with the transmission. If the channel is busy, the WiFi AP defers the transmission until the channel becomes idle. The WiFi AP then initializes its backoff timer with a randomly chosen backoff period and decrements this timer every time it senses the channel to be idle. The timer stops decreasing once the channel becomes busy and the decrementing process will be restarted again after DIFS idle sensing. The WiFi AP attempts to transmit once the timer reaches zero. The backoff mechanism and the definition of contention window will be discussed later in Section IV.

III. System Delay Analysis for Heterogeneous RAOSC-DSC Network

This section presents the mathematical derivation of the minimum average system delay of the non-aggregated scenario for heterogeneous RAOSC-DSC networks when negligible blockage rate of DSC is considered. It provides a theoretical proof that under this case the performance of the aggregated scenario is always better than that of the non-aggregated scenario in terms of the minimum average system delay. For the evaluation of the minimum average system delay of the aggregated scenario, an efficient solution is proposed. This solution is shown to achieve less than 3% close to the optimal solution. The comparison between the empirical results of the aggregated scenario and the delay performance of the non-aggregated scenario is also presented. In the end, when non-negligible blockage rate of DSC is assumed, we use simulation results to evaluate the minimum average system delay of the aggregated and non-aggregated scenarios.

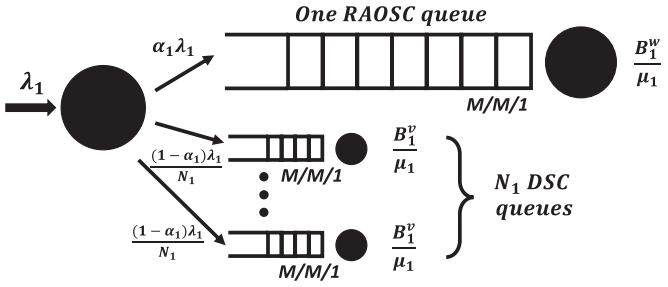


Fig. 3. Queuing model representing the non-aggregated system model for heterogenous RAOSC-DSC networks.

A. The Non-Aggregated Scenario

Let α_1 denote the percentage of requests allocated to RAOSC. The non-aggregated scenario can be represented by the queuing model shown in Fig. 3. Due to the assumption that requests are randomly forwarded to RAOSC and DSC, the requests arrival to each queue is still a Poisson process. Requests arrive to RAOSC and DSC queues with mean rates $\alpha_1\lambda_1$ and $(1-\alpha_1)\lambda_1/N_1$, respectively. The average service time of RAOSC and DSC queue are exponentially distributed with means B_1^w/μ_1 and B_1^v/μ_1, respectively. Thus, each RAOSC and DSC queue is characterized by the M/M/1 queuing model.

Theorem 1: In the non-aggregated system model, the minimum average system delay is

$$D_{min\_non\_agg} = \begin{cases} \dfrac{\mu_1 N_1}{B_1^v N_1 - \lambda_1\mu_1}, & \text{if } \dfrac{B_1^v N_1}{\lambda_1\mu_1}(1-\sqrt{\gamma N_1}) \geq 1 \\ \dfrac{\lambda_1\mu_1(1+N_1) - B_1^v N_1(1-\sqrt{\gamma N_1})^2}{\lambda_1[B_1^v N_1(\gamma+1) - \lambda_1\mu_1]}, & \text{otherwise} \end{cases}$$

Proof: The optimization problem for minimizing the average system delay is formulated as follows:

$$\begin{aligned} \text{Objective:} \quad & \min\ \alpha_1 D_{RAOSC} + (1-\alpha_1) D_{DSC} \\ & s.t.\ 0 \leq \alpha_1 \leq 1 \\ & \alpha_1\lambda_1 < B_1^w/\mu_1 \qquad (1) \\ & (1-\alpha_1)\lambda_1/N_1 < B_1^v/\mu_1 \qquad (2) \end{aligned}$$

In order to find the candidate minimum points, the average system delay as a function is described as follows:

$$\begin{aligned} D(\alpha_1) &= \alpha_1 D_{RAOSC} + (1-\alpha_1) D_{DSC} \\ &= \frac{\alpha_1}{B_1^w/\mu_1 - \alpha_1\lambda_1} + \frac{1-\alpha_1}{B_1^v/\mu_1 - (1-\alpha_1)\lambda_1/N_1} \end{aligned}$$

$D(\alpha_1)$ is continuous in $(1 - B_1^v N_1/(\lambda_1\mu_1), B_1^w/(\lambda_1\mu_1))$. From constraints (1) and (2), we have $1 - B_1^v N_1/(\lambda_1\mu_1) < 0$

and $B_1^w/(\lambda_1\mu_1) > 1$. Hence, $D(\alpha_1)$ is continuous in [0,1]. The derivative of $D(\alpha_1)$ is

$$D'(\alpha_1) = \frac{a\alpha_1^2 + b\alpha_1 + c}{f^2(\alpha_1)}, \text{ where}$$
$$a = \lambda_1^2(B_1^w - B_1^v N_1^2),$$
$$b = \frac{2\lambda_1 B_1^w(B_1^v N_1 - \lambda_1\mu_1 + B_1^v N_1^2)}{\mu_1},$$
$$c = [B_1^w((B_1^v)^2 N_1^2 - 2\lambda_1\mu_1 B_1^v N_1 + \lambda_1^2\mu_1^2 - B_1^w B_1^v N_1^2)]/\mu_1^2,$$
$$f(\alpha_1) = \sqrt{\mu_1}(-\lambda_1\alpha_1 + \frac{B_1^w}{\mu_1})(\frac{\lambda_1\alpha_1}{N_1} + \frac{B_1^v}{\mu_1} - \frac{\lambda_1}{N_1}).$$

It is found that $f^2(\alpha_1) \neq 0$ when α_1 is in [0,1]. Since $a < 0$ and $b^2 - 4ac > 0$, $D'(\alpha_1)$ has two zero points $\alpha_1(1)$ and $\alpha_1(2)$

$$\alpha_1(1) = \frac{\lambda_1\mu_1\sqrt{\gamma}/(B_1^v N_1) + \sqrt{\gamma}(\sqrt{\gamma N_1} - 1)}{\lambda_1\mu_1(\sqrt{\gamma} + \sqrt{N_1})/(B_1^v N_1)} \quad (3)$$

$$\alpha_1(2) = \frac{\sqrt{\gamma}[1 - B_1^v N_1(\sqrt{\gamma N_1} + 1)/(\lambda_1\mu_1)]}{\sqrt{\gamma} - \sqrt{N_1}} \quad (4)$$

$$\alpha_1(2) - \alpha_1(1) = \frac{2\sqrt{\gamma N_1}[1 - B_1^v N_1(\gamma + 1)/(\lambda_1\mu_1)]}{\gamma - N_1} \quad (5)$$

where $\gamma = B_1^w/(B_1^v N_1)$ and $\gamma < 1$. In (3), the numerator is less than $\lambda_1\mu_1/(B_1^v N_1)$ and the denominator is greater than $\lambda_1\mu_1/(B_1^v N_1)$. Thus, this proves $\alpha_1(1) < 1$. In (4), the numerator and the denominator are both less than zero. This proves that $\alpha_1(2) > 0$. In (5), since the numerator and denominator are both less than zero, $\alpha_1(2)$ is greater than $\alpha_1(1)$. This means that i) $D'(\alpha_1) < 0$ when $\alpha_1 < \alpha_1(1)$ or $\alpha_1 > \alpha_1(2)$; ii) $D'(\alpha_1) > 0$ when $\alpha_1(1) < \alpha_1 < \alpha_1(2)$.

The discussion is divided into four cases: i) $0 < \alpha_1(1) < 1$ and $0 < \alpha_1(2) < 1$; ii) $\alpha_1(1) \leq 0$ and $0 < \alpha_1(2) < 1$; iii) $0 < \alpha_1(1) < 1$ and $\alpha_1(2) \geq 1$; iv) $\alpha_1(1) \leq 0$ and $\alpha_1(2) \geq 1$. In case i) and iii), for the first case, $D'(\alpha_1)$ is negative in the range of $[0, \alpha_1(1))$ and $(\alpha_1(2), 1]$, and positive in the range of $(\alpha_1(1), \alpha_1(2))$. Also because $D(0) < D(1)$, thus $D_{min}(\alpha_1) = D(\alpha_1(1))$. For the third case, $D'(\alpha_1)$ is negative in the range of $[0, \alpha_1(1))$ and positive in the range of $(\alpha_1(1), 1]$. Therefore, $D_{min}(\alpha_1) = D(\alpha_1(1))$. In case ii) and iv), $D_{min}(\alpha_1) = D(0)$ because $D(0) < D(1)$. After substituting $\alpha_1 = 0$ and $\alpha_1 = \alpha_1(1)$ into $D(\alpha_1)$, it is found that

$$D(0) = \frac{\mu_1 N_1}{B_1^v N_1 - \lambda_1\mu_1} \text{ and}$$
$$D(\alpha_1(1)) = \frac{\lambda)1\mu_1(1 + N_1) - B_1^v N_1(1 - \sqrt{\gamma N_1})^2}{\lambda_1[B_1^v N_1(\gamma + 1) - \lambda_1\mu_1]}$$

Note that $D_{min\_non\_agg} = D(\alpha_1(1))$ iff $\alpha_1(1) > 0$. It means that $\frac{B_1^v N_1}{\lambda_1\mu_1}(1 - \sqrt{\gamma N_1}) < 1$. ■

B. The Aggregated Scenario

Let β_1 denote the proportion of the size of each request that is allocated to the RAOSC. The aggregated scenario can be represented by the queuing model shown in Fig. 4. Assuming that the requests arrival are randomly and evenly distributed to each DSC queue, the requests arrival process to each DSC queue is still a Poisson process. The average

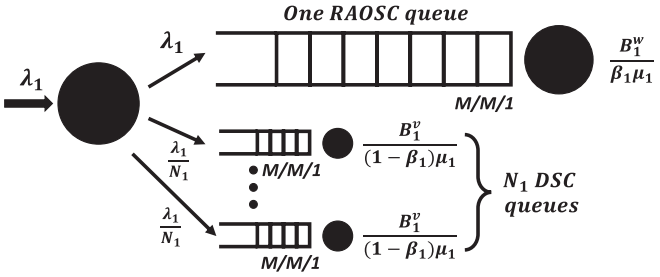


Fig. 4. Queuing model representing the aggregated system model for heterogeneous RAOSC-DSC networks.

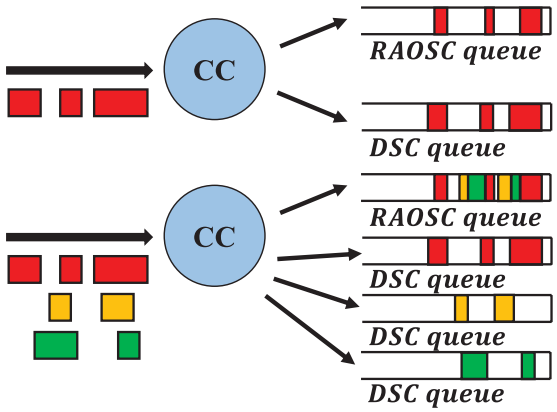


Fig. 5. Requests distribution in the aggregated scenario for $N_1 = 1$ and $N_1 > 1$.

requests arrival rates for RAOSC and DSC are λ_1 and λ_1/N_1, respectively. The average serving rates of RAOSC and DSC are $B_1^w/(\beta_1\mu_1)$ and $B_1^v/[(1-\beta_1)\mu_1]$, respectively. Similar to the non-aggregated scenario, each RAOSC and DSC queue can be characterized by the M/M/1 queuing model. The objective of the optimization problem can be expressed as minimizing $E[\max(D_{RAOSC}, D_{DSC})]$.

Fig. 5 represents the requests distribution to RAOSC and DSC queues for $N_1 = 1$ and $N_1 > 1$. In Fig. 5, it can be seen that when $N_1 = 1$, the delay of the DSC queue is fully correlated to that of the RAOSC queue. Therefore, achieving the objective value of minimizing $E[\max(D_{RAOSC}, D_{DSC})]$ is equivalent to obtaining the optimal β_1 from $E[D_{RAOSC}] = E[D_{DSC}]$. However, when $N_1 > 1$, the RAOSC queue contains different colored pieces of request, which are split from the requests flowing to different DSC APs. Each color represents a data stream destined to one DSC AP. The arrival times and the sizes of different colored pieces of request are independent while those of the same colored pieces of request are completely correlated. Specifically, due to the existence of yellow and green pieces of request (in Fig. 5) in the RAOSC queue, the departure times of the red pieces of request in the RAOSC queue and the DSC queue are neither independent nor completely correlated. Hence, the complexity of computing the optimal β_1 is severely exacerbated. Instead of searching for the optimal β_1 by minimizing $E[\max(D_{RAOSC}, D_{DSC})]$, the objective is

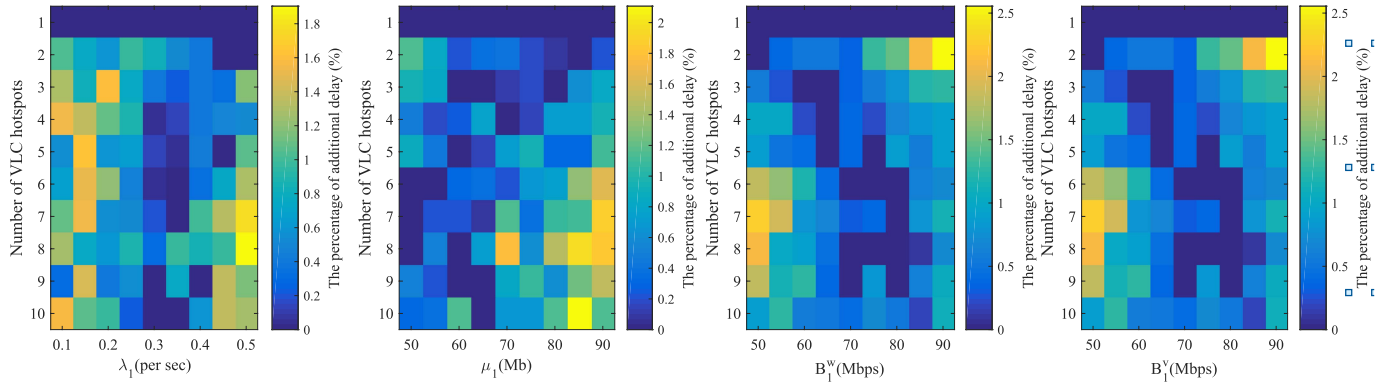


Fig. 6. The percentages of additional delay caused by approximation in terms of (a) λ_1; (b) μ_1; (c) B_1^w; (d) B_1^v, with N_1 varied from 1 to 10.

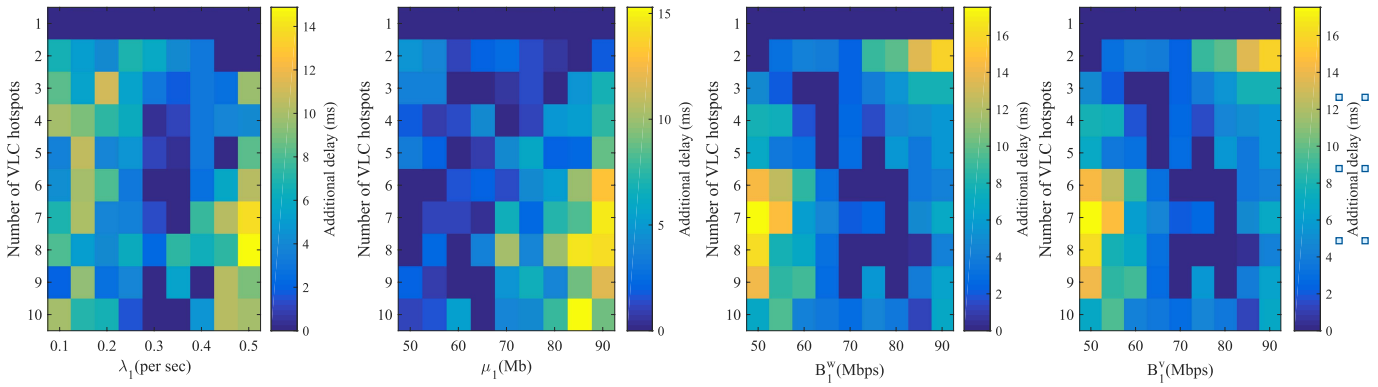


Fig. 7. The amount of additional delay caused by approximation in terms of (a) λ_1; (b) μ_1; (c) B_1^w; (d) B_1^v, with N_1 varied from 1 to 10.

simplified as minimizing $\max(E[D_{RAOSC}], E[D_{DSC}])$. For instance, let us assume that the delays of three pieces of request in RAOSC are 1, 2 and 3 seconds respectively, and the delays of the corresponding three pieces of request in DSC are 2 seconds for all. As such, the objective value of $E[\max(D_{RAOSC}, D_{DSC})]$ will be 2.33 seconds while the objective value of $\max(E[D_{RAOSC}], E[D_{DSC}])$ will be 2 seconds, which provides an underestimation of the traffic load. When the RAOSC queue is overwhelmed, approximated $E[D_{RAOSC}]$ will be lower than the real average request delay and vice versa. The error value has been further validated not to exceed 3% by the simulation results. To determine the approximated value of the optimal β_1 from the objective of minimizing $\max(E[D_{RAOSC}], E[D_{DSC}])$, we make $E[D_{RAOSC}] = E[D_{DSC}]$. Therefore, the approximated value of β_1 is, $\beta_1 = (-b - \sqrt{b^2 - 4ac})/(2a)$, where $a = \lambda_1\mu_1(1 - 1/N_1)$, $b = -[B_1^w + B_1^v + \lambda_1\mu_1(1 - 1/N_1)]$, and $c = B_1^w$.

By simulating the aggregated scenario with the approximated β_1, the percentages of additional delay caused by approximation are shown in Fig. 6. The values of the $\lambda_1, \mu_1, B_1^w, B_1^v$ are initially set as 0.5/s, 90 Mb, 50 Mpbs, 100 Mbps, respectively. In each plot, one of these four parameters is varied while keeping the other three fixed to the initial values. With N_1 varied from 1 to 10, it is noticed that the percentage of the maximum additional delay is 2.7%, which is less than 3%. Figs. 6 (a)-(c), show that, as λ_1, μ_1 and B_1^w increase, the percentage of the additional delay decreases initially and increases after reaching the minimum level. However, in Fig. 6 (d), the percentage of the delay penalty does not change much. Figs. 6 (a)-(c) show that the percentage of additional delay has the minimum values when $\lambda_1 \approx 0.33$, $\mu_1 \approx 58$ and $B_1^w \approx 70$, respectively. When $\lambda_1 < 0.33$, $\mu_1 < 58$ and $B_1^w > 70$, the approximation approach overestimates the congestion level of RAOSC and causes additional traffic load allocated to DSC, and vice versa. Note that when $N_1 = 1$, the approximated solution proposed here will lead to the exact minimum average system delay of the aggregated scenario because the delay of requests at each queue are fully correlated. The explicit additional delay values are shown in Fig. 7.

C. Theoretical Analysis

Theorem 2: Under our heterogeneous RAOSC-DSC network model, the aggregated scenario has a lower minimum average system delay than that of the non-aggregated scenario.

Proof: The average system delays of the non-aggregated and the aggregated scenarios are

$$\begin{aligned} E[D_{non\_agg}] &= \frac{\alpha_1}{B_1^w/\mu_1 - \alpha_1\lambda_1} \\ &\quad + \frac{1-\alpha_1}{B_1^v/\mu_1 - (1-\alpha_1)\lambda_1/N_1} \\ E[D_{agg}] &= E[\max(D_{RAOSC}, D_{DSC})] \\ &= E[D_{RAOSC}] + E[D_{DSC}] \\ &\quad - E[\min(D_{RAOSC}, D_{DSC})] \end{aligned}$$

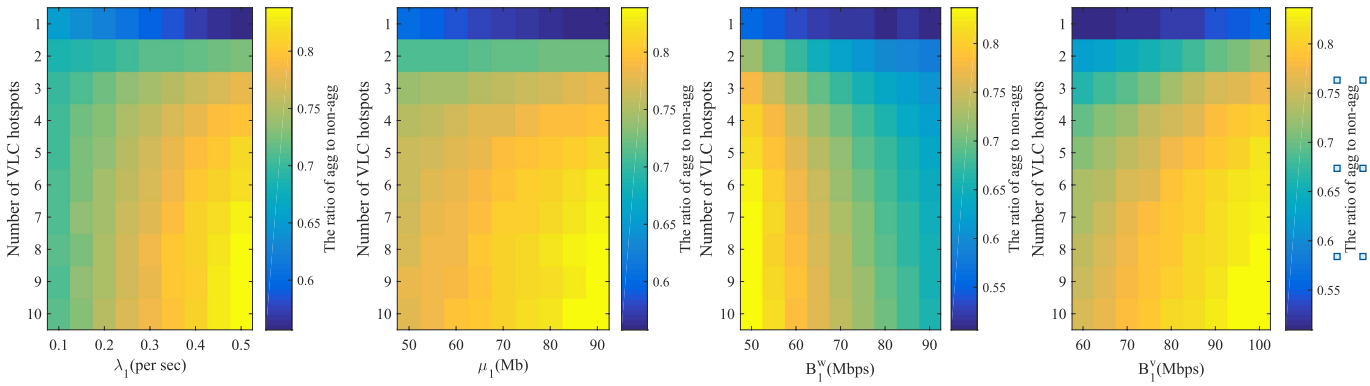


Fig. 8. The ratio of the approximated minimum average system delay of the aggregated scenario to the minimum average system delay of the non-aggregated scenario in terms of (a) λ_1; (b) μ_1; (c) B_1^w; (d) B_1^v, with N_1 varied from 1 to 10.

Note that, for aggregated scenario,

$$E[D_{RAOSC}] = \frac{1}{\frac{B_1^w}{\beta_1\mu_1} - \lambda_1} = \frac{\beta_1}{\frac{B_1^w}{\mu_1} - \beta_1\lambda_1}$$

$$E[D_{DSC}] = \frac{1}{\frac{B_1^v}{(1-\beta_1)\mu_1} - \frac{\lambda_1}{N_1}} = \frac{1-\beta_1}{\frac{B_1^v}{\mu_1} - \frac{(1-\beta_1)\lambda_1}{N_1}}$$

When $\alpha_1 = \beta_1$, since $E[\min(D_{RAOSC}, D_{DSC})]$ is greater than zero, we always have $E[D_{non\_agg}] > E[D_{agg}]$. Therefore, the minimum average system delay of the aggregated scenario is lower than that of the non-aggregated scenario. ■

D. Empirical Analysis

When applying the approximation method, the following question should be addressed: is the resulting minimum average system delay with approximated β_1 of the aggregated scenario still lower than that of the non-aggregated scenario? To further investigate the comparison between the non-aggregated and the aggregated scenarios, the analytical results obtained when applying the non-aggregated scenario are compared with the simulation results obtained when applying the approximated aggregated scenario. The ratio of the approximated minimum average system delay of the aggregated scenario to the minimum average system delay of the non-aggregated scenario is used to demonstrate the viability of the approximation approach. Fig. 8 illustrates the comparison. The values of λ_1, μ_1, B_1^w, B_1^v and N_1 are the same as those in Fig. 6. As such, based on the simulation parameters, the approximated minimum average system delay of the aggregated scenario is at least 16% lower than that of the non-aggregated scenario. The aggregation has diminishing gains over the non-aggregated scenario as the number of DSC APs increases and the ratio of RAOSC bandwidth to DSC bandwidth decreases. This is due to the additional RAOSC capacity which leads to decreasing the effect per DSC AP. Besides, the benefit of aggregating RAOSC and DSC becomes less evident as λ_1 and μ_1 increases. This is because increasing traffic load reduces the effect of efficient bandwidth utilization provided by aggregation.

E. Extension to Non-Negligible Blockage Rate of DSC

As it will be discussed in the next section, the queuing model of DSC would be changed to M/G/1 if non-zero blockage rate is considered. As a result, it would be very difficult to mathematically derive the minimum average system delay of the non-aggregated scheme for heterogeneous RAOSC-DSC networks and also very complicated to theoretically compare the performance of the aggregated scheme and that of the non-aggregated scheme in terms of the minimum average system delay. Note that the mathematical derivation and theoretical comparison are both performed in the first case (i.e. RAOSC-DSC) when negligible blockage rate is considered.

To evaluate the RAOSC-DSC case when non-negligible blockage rate of DSC is assumed, we perform simulations with the settings similar to that of the negligible blockage rate case, but change the blockage rate from 0 to 0.1 and 0.2. The simulation results of RAOSC-DSC case are shown in Fig. 9 and Fig. 10, respectively. Comparing the results in Fig. 9 and Fig. 10 to the results in Fig. 8, we observe that the variation trend of the ratio of the minimum average system delay of the aggregation scenario to that of the non-aggregation scenario are very similar. As it is expected, the only difference is that when non-zero blockage rate is considered for the DSC channels, the benefit of performing aggregation increases. This is consistent with the simulation results in the Fig. 8. As the bandwidth of DSC decreases, which is similar to increase the blockage rate of DSC channel, the gain of performing aggregation is enhancing. Therefore, the same conclusion when blockage is not considered can be drawn when blockage is considered.

IV. System Delay Analysis for Heterogeneous CBOSC-DSC Network

In this section, we first model the system delay of the non-aggregated and the aggregated scenarios for heterogeneous CBOSC-DSC networks. To validate our analytical model, we conduct extensive simulations based on the system model presented in Section II. We also observe from the simulation results that, under certain conditions, the non-aggregated scenario outperforms the aggregated one in terms of minimum

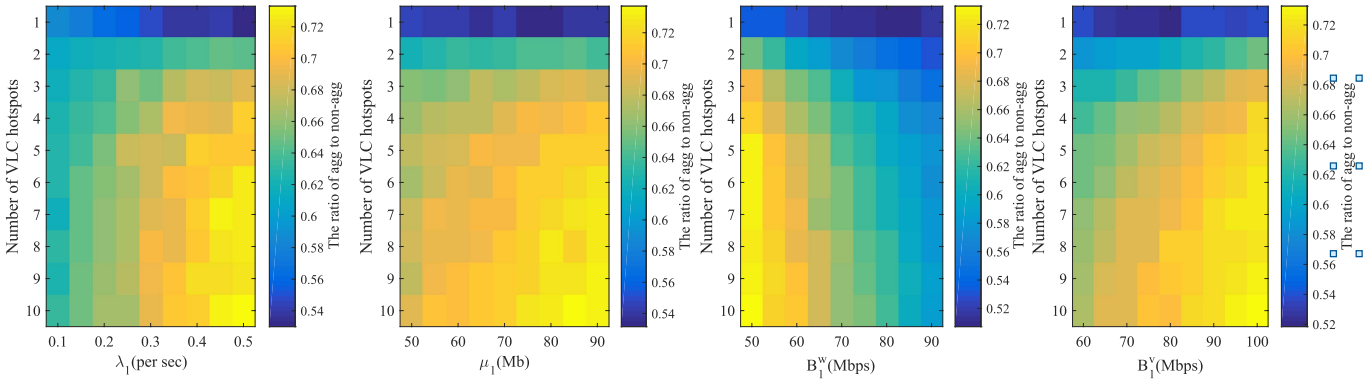


Fig. 9. For the case of RAOSC-DSC, when blockage rate of DSC is 0.1, the ratio of the minimum average system delay of the aggregated scenario to that of the non-aggregated scenario in terms of (a) λ_1; (b) μ_1; (c) B_1^w; (d) B_1^v, with N_1 varied from 1 to 10.

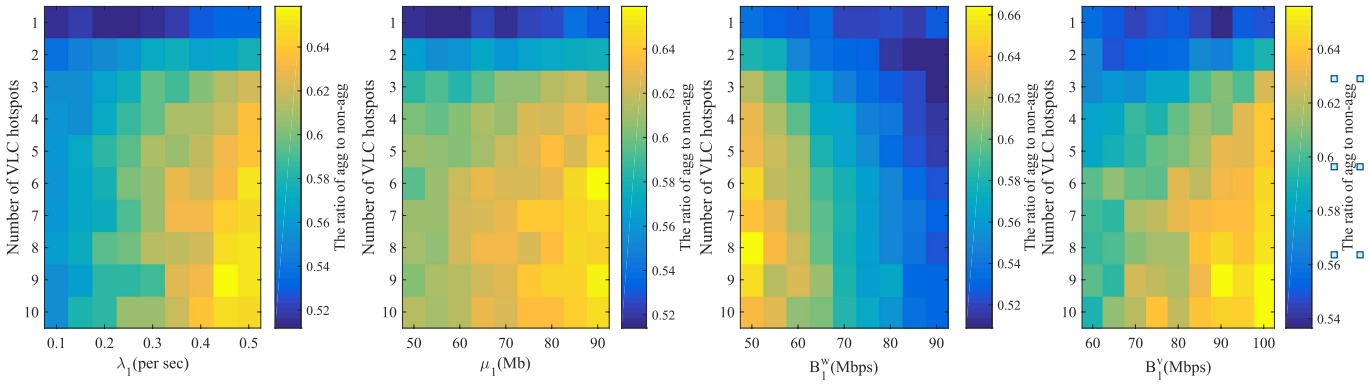


Fig. 10. For the case of RAOSC-DSC, when blockage rate of DSC is 0.2, the ratio of the minimum average system delay of the aggregated scenario to that of the non-aggregated scenario in terms of (a) λ_1; (b) μ_1; (c) B_1^w; (d) B_1^v, with N_1 varied from 1 to 10.

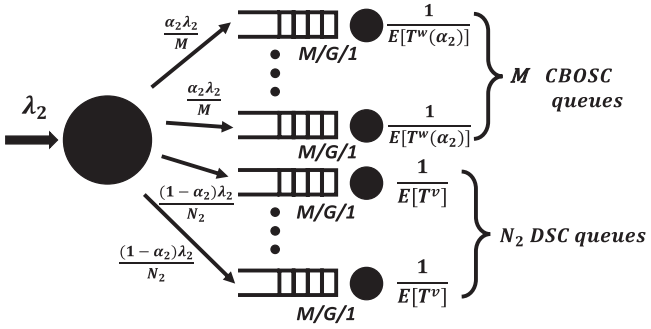


Fig. 11. Queuing model representing the non-aggregated system model for heterogeneous CBOSC-DSC networks.

average system delay. This is due to the fact that the delay penalty introduced by aggregation when contention and back-off mechanism is utilized surpasses the benefit of splitting the request.

A. The Non-Aggregated Scenario

Let α_2 denote the percentage of requests allocated to CBOSC. The non-aggregated scenario can be represented by the queuing model in Fig. 11. Similar to the analysis for heterogeneous RAOSC-DSC networks, the request arrival process to each queue is still a Poisson process. However, since the contention and backoff of 802.11 protocols are considered when modeling the CBOSC network, the service time of each CBOSC queue $T^w(\alpha_2)$ depends on the traffic load allocated to CBOSC. Also, for DSC queues, due to the consideration of the blockage, the distribution of the service time of each request T^v is not memoryless. Therefore, the M/G/1 queuing model is utilized to characterize each CBOSC and DSC queue. In order to fully characterize the delay of the resulting M/G/1 model, we need to derive the expectation and the second moment of the service time of the resulting M/G/1 model.

The minimum and maximum contention window size associated with backoffs are denoted by CW_{min} and CW_{max}, respectively. In 802.11 protocol, m is defined as $m = \log_2(CW_{max}/CW_{min})$. For instance, $CW_{min} = 16$ slots and $CW_{max} = 1024$ slots, and thus $m = 6$ for 802.11n protocol. In the following analysis, since RTS/CTS exchange is considered, we denote the probability that an RTS transmission results in a collision by p. Following the same approach in [20] [20, eq. (5)], the average number of backoff slots experienced by a request at a CBOSC AP can be expressed as

$$\bar{W} = \frac{1 - p - p(2p)^m}{1 - 2p} \frac{CW_{min}}{2}. \tag{6}$$

Denote the duration consumed by a collision by $T_c = DIFS + \sigma_{RTS}$, where Distributed Inter-Frame Space (DIFS)

is utilized to sense the idle channel and $\sigma_{RTS} = l_{RTS}/B_2^w$ is the transmission delay of an RTS packet. Given the average request arrival rate as $\frac{\alpha_2\lambda_2}{M}$ and the average time to transmit a request in CBOSC queue as $\frac{\mu_2}{B_2^w}$, the collision probability can be expressed as follows according to [20] [20, eq. (11)]

$$p = 1 - \left(1 - \frac{\frac{\alpha_2\lambda_2}{M}[1 + \frac{1}{\bar{W}}(\frac{\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)})]}{1 - \frac{\alpha_2\lambda_2}{M}(M-1)[\frac{\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)}]}\right)^{M-1}. \tag{7}$$

By substituting (6) into (7), the collision rate p can be obtained by numerical methods.

Denote the queue utilization rate of each CBOSC AP as ρ, then according to [20] [20, eq. (10)], we have

$$\rho = \frac{\frac{\alpha_2\lambda_2}{M}[\frac{\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)} + \bar{W}]}{1 - \frac{\alpha_2\lambda_2}{M}(M-1)[\frac{\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)}]}.$$

Next, we start deriving the probability density function (pdf) of the request service time, which is from the instant that the request reaches the head to the queue to the instant that the request departs from the queue. The pdf of the backoff slots (BO), following a successful transmission of a request at a CBOSC AP, is represented by

$$\begin{aligned} P[BO = i] = {} & \rho(1-p)U_{1,CW_{min}}(i) + p(1-p) \\ & \times [U_{1,CW_{min}} * U_{1,2CW_{min}}(i)] \\ & + \ldots + (p)^m(1-p)[U_{1,CW_{min}} * U_{1,2CW_{min}} \\ & * \ldots * U_{1,2^mCW_{min}}](i)], \end{aligned}$$

where $U_{a,b}$ denotes the pdf of a uniform distribution between a and b, and $*$ represents the convolution operation.

To evaluate the portion of service time resulted from the successful transmissions and collisions of the contending CBOSC APs, we denote q as the probability that one of the remaining $M-1$ CBOSC APs attempts to transmit in a given slot, and q_c as the probability that a collision occurs in a slot given that at least one of the $M-1$ CBOSC APs attempts to transmit in that slot. According to [20] [20, eqs. (13) and (15)]), we have

$$q = 1 - (1 - \frac{\rho}{\bar{W}})^{M-1},$$

and

$$q_c = \frac{1 - (1 - \frac{\rho}{\bar{W}})^{M-1} - \frac{(M-1)\rho}{\bar{W}}(1 - \frac{\rho}{\bar{W}})^{M-1}}{1 - (1 - \frac{\rho}{\bar{W}})^{M-1}}.$$

Assume that in the i backoff slots, j slots are followed by transmission attempts of the other $M-1$ CBOSC APs and k out of j slots are followed by collisions, then $j-k$ slots are followed by successful transmissions of the $M-1$ CBOSC APs. Since the summation of $j-k$ i.i.d. exponential random variables (i.e. transmission time of a request $\frac{\mu_2}{B_2^w}$) is a gamma random variable, the contribution of $j-k$ successful transmissions to the service time can be expressed as a gamma distribution

$$l^{(j-k)}(x) = \frac{1}{(j-k-1)!\,(\frac{B_2^w}{\mu_2})^{j-k}} x^{j-k-1} e^{-\frac{\mu_2 x}{B_2^w}}.$$

Then the pdf of the channel access delay experienced by a request is given by

$$\begin{aligned} P[Y = s] = {} & \sum_i^\infty \sum_j^i \sum_k^j l^{(j-k)}(x)\binom{i}{j} q^i(1-q)^{i-j} \\ & \times \binom{j}{k} q_c^k (1-q_c)^{j-k} P[BO = i] I(s), \end{aligned} \tag{8}$$

where $\binom{i}{j}q^i(1-q)^{i-j}$ represents the probability that j out of i slots are followed by transmission attempt from the $M-1$ CBOSC APs, $\binom{j}{k}q_c^k(1-q_c)^{j-k}$ represents the probability that k out of j slots are followed by collisions, and $I(s)$ is an indicator function which equals 1 when $s = x + i + kT_c$ and 0 otherwise.

Denote the moment generating function (mgf) of the channel access delay by $M_Y(t)$, the mgf of the total service time $M_R(t)$, including the channel access delay and request transmission time, is given by

$$M_R(t) = M_Y(t)(1 - t(\frac{B_2^w}{\mu_2})^{-1})^{-1},$$

where $(1 - t(\frac{B_2^w}{\mu_2})^{-1})^{-1}$ represents the mgf of an exponential random variable with mean $\frac{\mu_2}{B_2^w}$. Then the second moment and the mean of the total service time T^w can be obtained by differentiating $M_R(t)$ with respect to t and setting $t = 0$ as follows

$$E[(T^w)^2] = \frac{\mathrm{d}^2 M_R(t)}{\mathrm{d}t^2}(0), \;\; E[T^w] = \frac{\mathrm{d}M_R(t)}{\mathrm{d}t}(0).$$

According to Pollaczek-Khinchine formula, the expected system delay of CBOSC queues is given by

$$E[D_{CBOSC}] = \frac{\frac{\alpha_2\lambda_2}{M}E[(T^w)^2]}{2(1-\rho)} + E[T^w].$$

For DSC queues, in order to fully characterize the average system delay of requests, we need to derive the expectation and the second moment of the service time of the resulting M/G/1 model. Recall that the probability of successful transmission is denoted by P_{succ} and packet drop due to buffer limitation is not considered. Although in some cases, a packet may be dropped after a certain number of unsuccessful retransmissions, the error caused by this infinite extension is negligible since $P_{succ}(1 - P_{succ})^{n-1} \to 0$ as n increases. Therefore, the expected service time of a request in DSC queues is

$$\begin{aligned} E[T^v] = {} & \frac{\mu_2}{B_2^v}[P_{succ} + 2P_{succ}(1 - P_{succ}) \\ & + \ldots + nP_{succ}(1 - P_{succ})^{n-1} + \ldots] \\ = {} & \frac{\mu_2}{B_2^v P_{succ}}. \end{aligned}$$

Suppose a request's transmission time is v and the number of transmission attempts is u, then the total service time of this request is uv. Thus, the second moment of the service time of a request in DSC queues is

$$\begin{aligned} E[(T^v)^2] = {} & \sum_v^\infty \sum_u^\infty \frac{B_2^v}{\mu_2} e^{-\frac{B_2^v}{\mu_2}v} P_{succ} \\ & \times (1 - P_{succ})^{u-1}(uv)^2. \end{aligned}$$

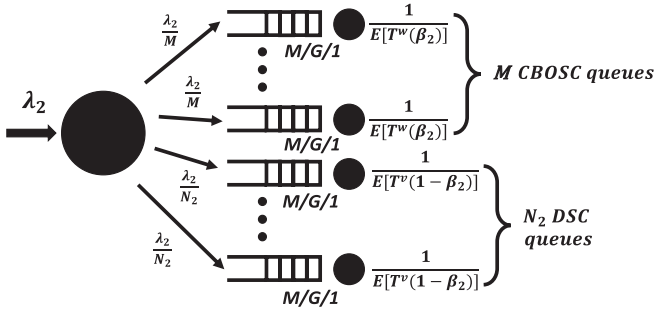


Fig. 12. Queuing model representing the aggregated system model for heterogeneous CBOSC-DSC networks.

According to Pollaczek-Khinchine formula, the expected system delay of DSC queues is given by

$$E[D_{DSC}] = \frac{\frac{(1-\alpha_2)\lambda_2}{N_2} E[(T^v)^2]}{2(1 - \frac{(1-\alpha_2)\lambda_2}{N_2} E[T^v])} + E[T^v].$$

Since α_2 portion of the requests are allocated to CBOSC networks and $1 - \alpha_2$ portion of requests are allocated to DSC networks, the average system delay of the heterogeneous CBOSC-DSC networks based on the non-aggregated scenario is given by

$$D_{non\_agg} = \alpha_2 E[D_{CBOSC}] + (1 - \alpha_2) E[D_{DSC}].$$

B. The Aggregated Scenario

Let β_2 denote the proportion of the size of each request that is allocated to the CBOSC. The aggregated scenario can be represented by the queuing model in Fig. 12. Similar to the non-aggregated scenario for heterogeneous CBOSC-DSC networks, the request arrival process of each CBOSC or DSC queue can be described by a Poisson process, and the distribution of service time are not memoryless for both CBOSC and DSC queues. Therefore, we use the M/G/1 queuing model to characterize the system delay of each CBOSC and DSC queue.

For the derivation of the system delay for the aggregated scenario, we only describe the parameters p, ρ, $l^{(j-k)}(x)$, $M_R(t)$, $E[D_{CBOSC}]$, $E[T^v]$, $E[(T^v)^2]$ and $E[D_{DSC}]$ with different expressions when comparing them to those of the non-aggregated scenario. Given the average request arrival rate of CBOSC queues as $\frac{\lambda_2}{M}$ and the average time to transmit a request in CBOSC queue as $\frac{\beta_2\mu_2}{B_2^w}$, the collision probability, queue utilization and the contribution of $j - k$ successful transmissions to the service time can be expressed as follows

$$p = 1 - \left(1 - \frac{\frac{\lambda_2}{M}[1 + \frac{1}{\bar{W}}(\frac{\beta_2\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)})]}{1 - \frac{\lambda_2}{M}(M-1)[\frac{\beta_2\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)}]}\right)^{M-1}, \tag{9}$$

$$\rho = \frac{\frac{\lambda_2}{M}[\frac{\beta_2\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)} + \bar{W}]}{1 - \frac{\lambda_2}{M}(M-1)[\frac{\beta_2\mu_2}{B_2^w} + T_c\frac{p}{2(1-p)}]}, \tag{10}$$

$$l^{(j-k)}(x) = \frac{1}{(j-k-1)!\,(\frac{B_2^w}{\beta_2\mu_2})^{j-k}} x^{j-k-1} e^{-\frac{\beta_2\mu_2 x}{B_2^w}}. \tag{11}$$

Substitute (9), (10) and (11) into (8), the pdf of the channel access delay can be obtained. Then the mgf of the total service time is expressed as follows

$$M_R(t) = M_Y(t)(1 - t(\frac{B_2^w}{\beta_2\mu_2})^{-1})^{-1}.$$

Similar to the non-aggregated scenario, the expected service time of a request in CBOSC queues is

$$E[D_{CBOSC}] = \frac{\frac{\lambda_2}{M} E[(T^w)^2]}{2(1-\rho)} + E[T^w].$$

For DSC queues, the expectation and the second moment of the service time are

$$E[T^v] = \frac{\beta_2\mu_2}{B_2^v P_{succ}} \text{ and}$$

$$E[(T^v)^2] = \sum_{v}^{\infty}\sum_{u}^{\infty} \frac{B_2^v}{\beta_2\mu_2} e^{-\frac{B_2^v}{\beta_2\mu_2}v} P_{succ} \times (1 - P_{succ})^{u-1}(uv)^2.$$

The expectation of the system delay of the DSC queues is

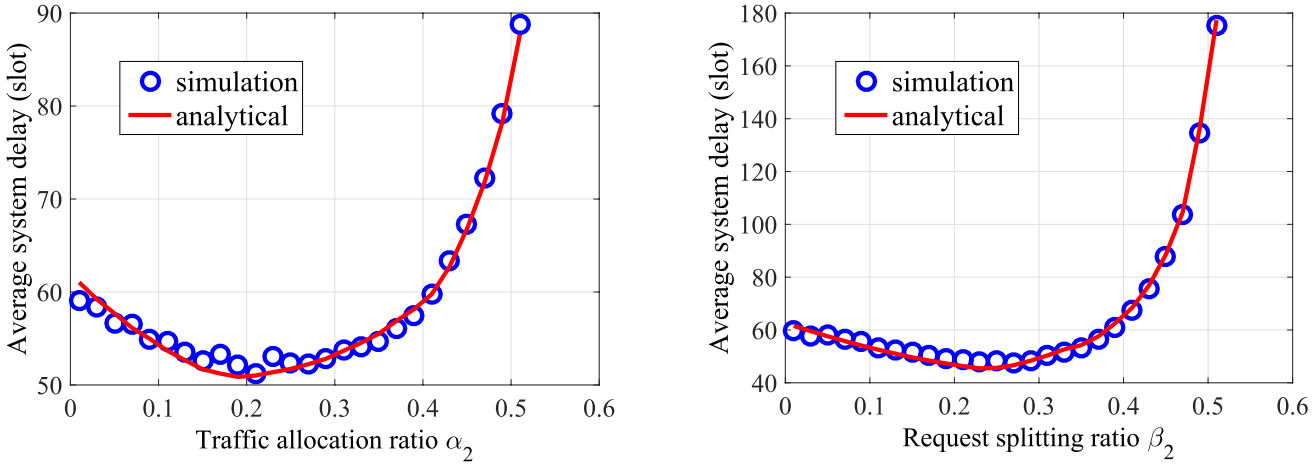
$$E[D_{DSC}] = \frac{\frac{\lambda_2}{N_2} E[(T^v)^2]}{2(1 - \frac{\lambda_2}{N_2} E[T^v])} + E[T^v].$$

Similar to the approximation for the aggregated scenario in heterogeneous RAOSC-DSC networks, the average system delay of the heterogeneous CBOSC-DSC networks based on the aggregated scenario is approximated by

$$D_{agg} = \begin{cases} E[D_{CBOSC}], & \text{if } E[D_{CBOSC}] \geq E[D_{DSC}], \\ E[D_{DSC}], & \text{otherwise.} \end{cases}$$

C. Empirical Analysis

To validate our analytical model and compare the system delay performance of heterogeneous CBOSC-DSC networks under non-aggregated and aggregated scenarios, we conduct extensive simulations under the homogeneous traffic assumptions. The final system delay is averaged over 100,000 simulated requests. For the simulation settings, we consider a 8 × 10 meters room. There are 10 CBOSC APs located in a single contention domain (i.e. each pair of CBOSC APs have non-negligible interference between each other). For 802.11 a/g/n, the minimum and maximum contention window sizes [38] are 16 slots and 1024 slots, respectively. Referring to [20], the 802.11 MAC settings, including RTS size, CTS size, DIFS and slot size, are set to 44 bytes, 38 bytes, 50 μsec and 20 μsec, respectively. In the room, there are 20 DSC APs mounted on the 2.5 meters height ceiling in grid structure, where each DSC AP is serving a 2 × 2 meters square area. Each adjacent 4 DSC APs are using different frequency. In other words, the reuse factor is 4. Each DSC AP has 5 MHz bandwidth and is using 4-PAM as the modulation scheme. The maximum optical power of each DSC AP is set to 0.5 Watt. The Gaussian noise value is calculated based on the parameters in [39] and is set to 4.7×10^{-14} A^2. The semi-angle at half power, area of detector, optical filter gain and refractive index are all set to the same as the parameter in [39]. For 4-PAM, the



(a) non-aggregated scenario (b) aggregated scenario

Fig. 13. Comparison between the simulation and analytical results of the average system delays for (a) non-aggregated scenario; (b) aggregated scenario.

TABLE II
VALUES OF THE PARAMETERS USED IN THE SIMULATION

Room size	8×10×2.5 meters
Number of CBOSC APs	10
CBOSC bandwidth	20 Mbps
Minimum contention window	16
Maximum contention window	1024
RTS size	44 bytes
CTS size	38 bytes
DIFS	50 μsec
Slot size	20 μsec
Number of DSC APs	20
Reuse factor of DSC	4
Modulation scheme	4-PAM
Maximum optical power	0.5 Watt
Noise level	4.7×10^{-14} A^2
Field of view	40 degrees
DSC bandwidth	10 Mbps
P_{succ} of DSC	0.5
request arrival rate	0.05/slot
mean request size	1000 bytes

required minimum SNR value for achieving 10^{-3} bit error rate is 19.80 dB [40]. Based on the setting, the SNR value for the user terminals located at the boundary of each AP's coverage is 25.78 dB, which satisfies the minimum requirement of 4-PAM. The field of view (FOV) of optical receivers is set to 40 degrees. which means that for each DSC AP, the signals from the closest interfering AP will not be received by the serving user terminals. Therefore, each DSC AP can achieve 10 Mbps throughput. Within each 2 × 2 meters square area served by each DSC AP, based on the practical settings given above, the data rate of a user terminal will be the same no matter where it is located. The uniformly distributed blockage rate is set to 0.5. All the parameter settings for CBOSC and DSC networks are given in Table II.

In Fig. 13, we vary the traffic allocation ratio α_2 for the non-aggregated scenario and the request splitting ratio β_2 for the aggregated scenario, and compare the simulation and analytical results for the average system delay. For both scenarios, we can see the close match between the analytical and simulation results. The simulation results are the average system delay over all the simulated requests. If the number of simulated requests is large enough, the simulation results are expected to converge to the analytical results. Refer to (9) in [20] as follows,

$$\frac{1}{\mu} = \rho(N-1)[T_S + T_C\frac{p}{2(1-p)}] + \bar{W} + T_S + T_C\frac{p}{2(1-p)}$$

the factor of 2 in the denominator of $T_C\frac{p}{2(1-p)}$ represents the first degree approximation that only two nodes are involved in a collision. The first degree approximation underestimates the collision effect, thus under some cases (i.e. three or more nodes collide), the simulation result is expected to be above the analytical one. On the other hand, refer to (6) in [20] as follows,

$$p = 1 - P[SE]^{N-1}$$

where $P[SE]$ denotes the probability that a node does not transmit in a slot, the assumption behind (6) in [20] is that the event that a node does not transmit in a slot is independent of similar decisions by the other nodes. The decoupling approximation overestimates the collision probability, therefore under some cases (i.e. a node does not transmit is correlated to the similar decisions of the other nodes), the simulation result is expected to be below the analytical one. As expected, there exist optimal values of α_2 and β_2 that will lead to the minimum average system delay of the heterogeneous CBOSC-DSC network. With α_2 and β_2 lower than the optimal values, the DSC network will contribute more delay penalty to the average system delay. However, since the contention and backoff mechanism is not utilized in DSC, the average system delay will not approach to infinity even if α_2 and β_2 are equal to 0. In contrast, as α_2 and β_2 increase above the optimal value, the CBOSC queues will be saturated quickly, which leads to infinite average system delay.

In Fig. 14, the values of $\lambda_2, \mu_2, B_2^w, B_2^v$ are initially set to the values in Table II. In each plot, one of these four

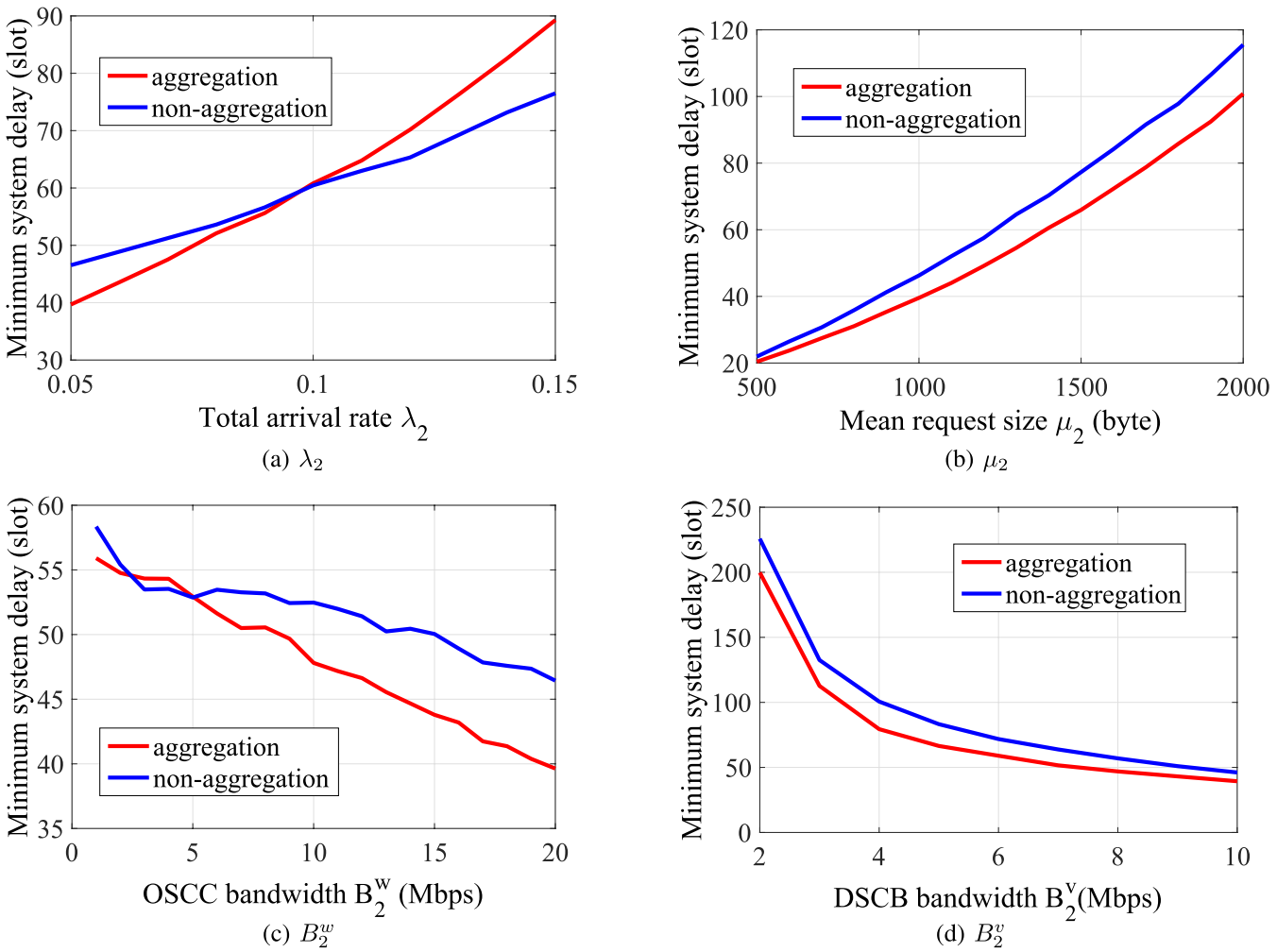


Fig. 14. Comparison between the average system delays of non-aggregated scenario and aggregated scenario in terms of (a) λ_2; (b) μ_2; (c) B_2^w; (d) B_2^v, when $M = 10$ and $N_2 = 20$.

parameters is varied while keeping the other three fixed at the initial values. In Fig. 14 (a), it is observed that the average system delay of aggregated scenario is not always lower than that of the non-aggregated scenario. This is the major difference from the simulation results of heterogeneous RAOSC-DSC networks, where contention and backoff mechanism is not utilized. As the request arrival rate increases, the backoff penalty brought by aggregation will surpass the benefit from splitting the requests. Therefore, in heterogeneous networks where contention and backoff mechanism is applied, under certain conditions, the non-aggregated scenario outperforms the aggregated scenario in terms of average system delay. In Fig. 14 (b), as the mean request size increases, the gap between aggregation and non-aggregation increases. These results are opposite to the results of Fig. 8 (b). The reason is that as the mean request size decreases, the benefit brought from aggregation becomes less evident than the backoff penalty. In Fig. 14 (c) and Fig. 14 (d), the results are consistent with the results of Fig. 8 (c) and (d). As the CBOSC bandwidth increases, the collision probability of the CBOSC network decreases. Thus, the delay penalty effect brought by aggregation is diminishes. As the DSC bandwidth increases, similar to the heterogeneous RAOSC-DSC network, the benefit gain of aggregated scenario is slightly reduced. This is because the increase in the DSC bandwidth leads to smaller optimal α_2 and β_2, which will reduce the gap between the delay performance of non-aggregated scenario and aggregated scenario.

To evaluate the effect of the number of APs on the system delay performance of the heterogeneous CBOSC-DSC network, we reduce the number of CBOSC APs M from 10 to 2 and the number of DSC APs N_2 from 20 to 4. The comparisons between non-aggregated scenario and aggregated scenario in terms of λ_2, μ_2, B_2^w, B_2^v are performed again and the simulation results are shown in Fig. 15. Compared to the simulation results when $M = 10$ and $N_2 = 20$, the average system delays are higher when $M = 2$ and $N_2 = 4$. This is because the total network capacity is reduced when the number of APs decreases. We also observe that when $M = 10$ and $N_2 = 20$, the benefit gain of aggregated scenario over non-aggregated scenario is less than 20%; while this benefit gain increases up to 40% when $M = 2$ and $N_2 = 4$. In addition, we set the values of λ_2, μ_2, B_2^w, B_2^v to 0.05/slot, 1000 bytes, 20 Mbps, 10 Mbps, respectively. The number of CBOSC APs

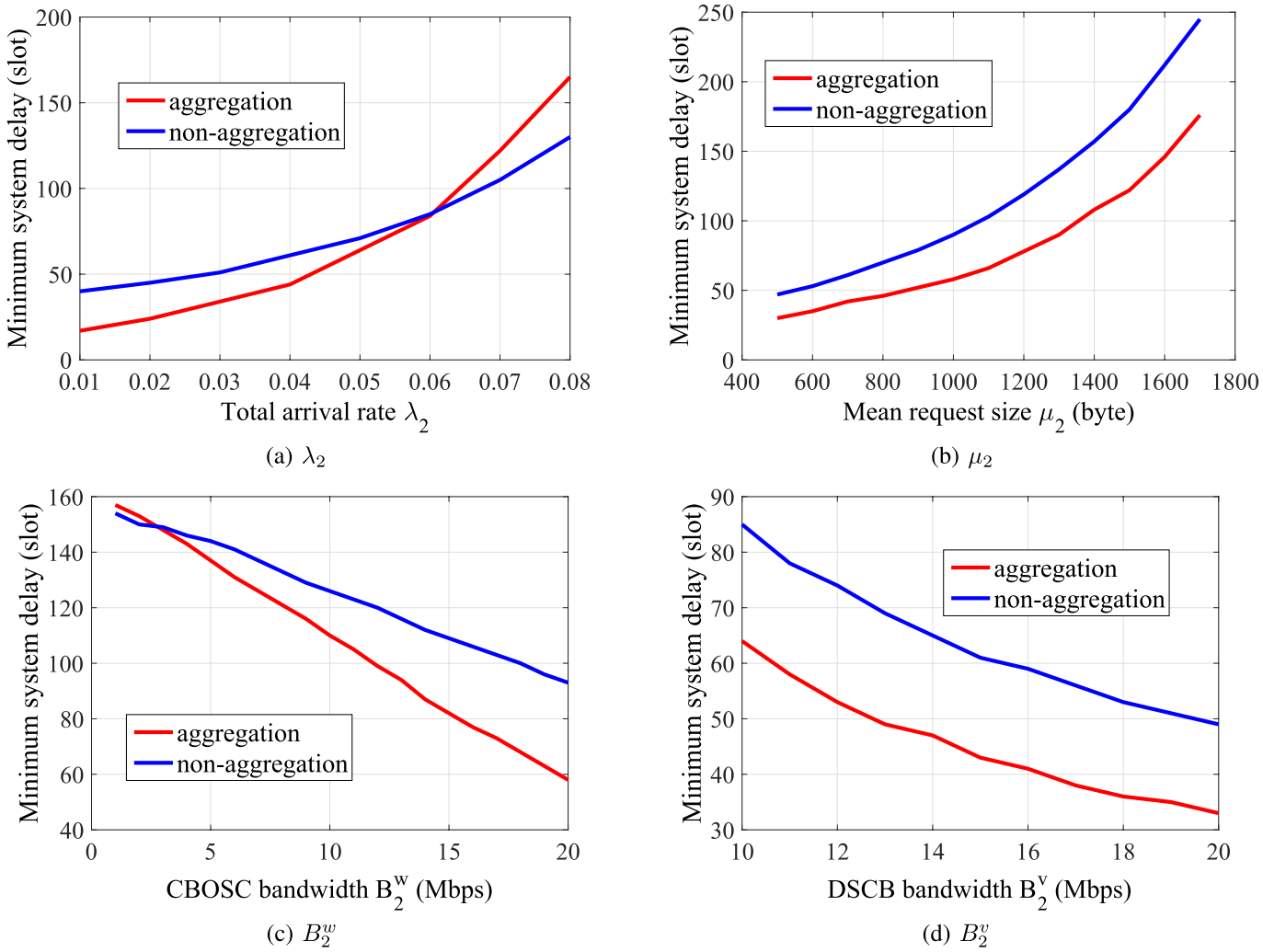


Fig. 15. Comparison between the average system delays of non-aggregated scenario and aggregated scenario in terms of (a) λ_2; (b) μ_2; (c) B_2^w; (d) B_2^v, when $M = 2$ and $N_2 = 4$.

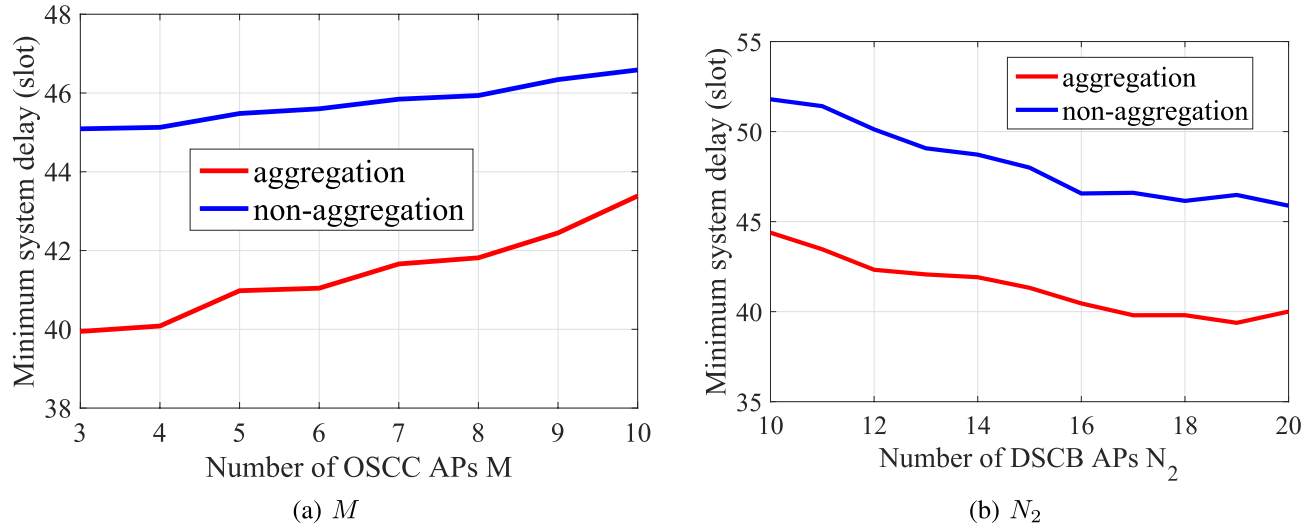
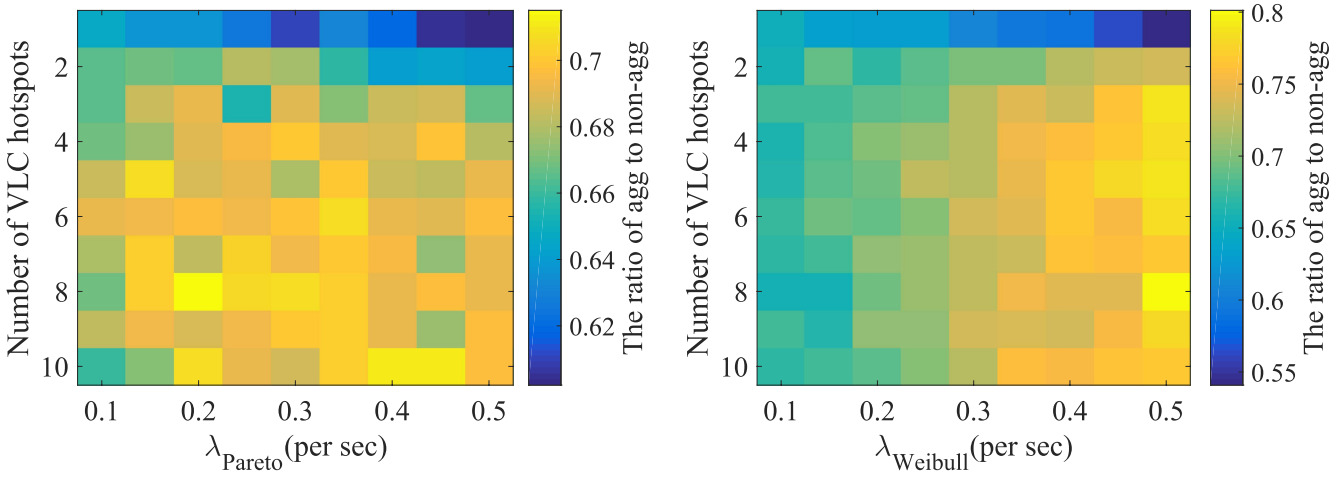


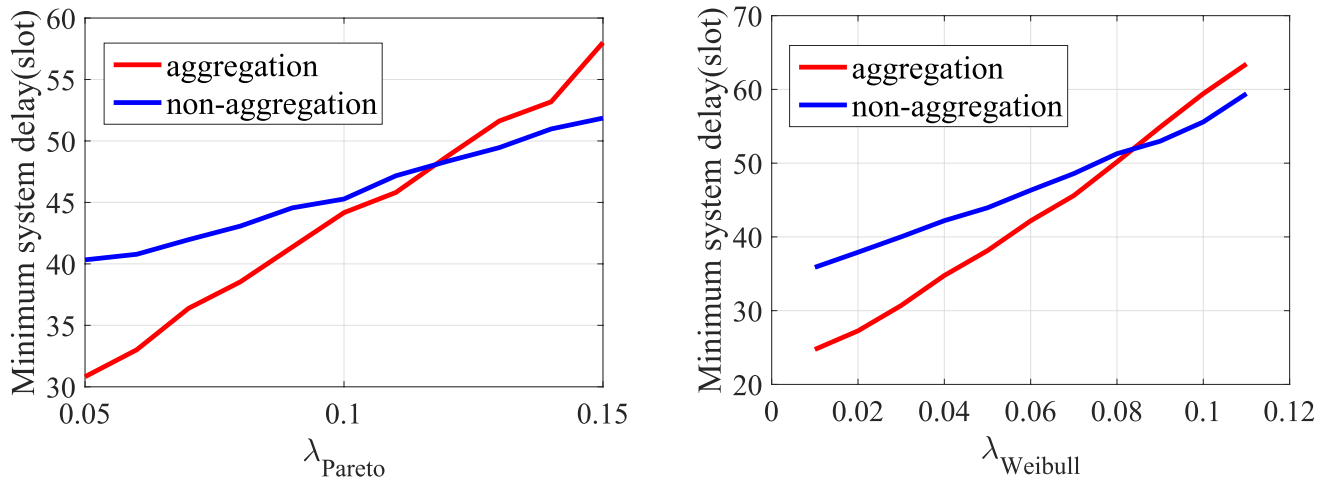
Fig. 16. Comparison between the minimum average system delays of non-aggregated scenario and aggregated scenario in terms of (a) the number of CBOSC APs M; (b) the number of DSC APs N_2.

M are varied from 3 to 10 while fixing the number of DSC APs N_2 to 20. The simulation results are shown in Fig. 16 (a). As it is expected, the gap between aggregation and non-aggregation is decreasing when the number of CBOSC APs M increases. This is because with certain value of total request arrival rate, mean request size, CBOSC and DSC



(a) Generalized Pareto distribution (b) Weibull distribution

Fig. 17. For the case of RAOSC-DSC, the ratio of the minimum average system delay of the aggregated scenario to that of the non-aggregated scenario in terms of (a) λ_{pareto} for generalized pareto distribution and (b) $\lambda_{weibull}$ for weibull distribution, with N_1 varied from 1 to 10.



(a) Generalized Pareto distribution (b) Weibull distribution

Fig. 18. For the case of CBOSC-DSC, comparison between the minimum average system delay of non-aggregated scenario and aggregated scenario in terms of (a) λ_{pareto} for generalized pareto distribution and (b) $\lambda_{weibull}$ for weibull distribution, when $M = 10$ and $N_2 = 20$.

bandwidth, the collision probability of CBOSC network is increasing as the number of CBOSC APs increases. In particular, the backoff penalty of aggregated scenario is dominating as the number of CBOSC APs increases. Therefore, the benefit gain of aggregated scenario over non-aggregated scenario becomes dominant when the number of CBOSC APs is small. In Fig. 16 (b), the number of DSC APs N_2 are varied from 10 to 20 while fixing the number of CBOSC APs M to 10. We observe that the gap between aggregation and non-aggregation does not change much when the number of DSC APs N_2 varies. However, the minimum average system delay of the two scenarios are both decreasing as N_2 increases. This is due to the additional network capacity added by increasing number of DSC APs.

Furthermore, to evaluate the effect of other distribution of arrival process on our approach, we investigate two other distributions of interarrival time by simulations - generalized pareto distribution [41] and weibull distribution [42]. The pdf of the generalized pareto distribution is as follows:

$$\begin{aligned} y_{pareto} &= f(x|k, \lambda_{pareto}, \theta) \\ &= \lambda_{pareto}(1 + k(x-\theta)\lambda_{pareto})^{-1-\frac{1}{k}} \end{aligned}$$

where k is the shape parameter, λ_{pareto} is the reciprocal of the scale parameter and θ is the threshold parameter.

The pdf of the weibull distribution is shown as follows:

$$\begin{aligned} y_{weibull} &= f(x|\lambda_{weibull}, b) \\ &= \lambda_{weibull} b(\lambda_{weibull} x)^{b-1} e^{-(\lambda_{weibull} x)^b} \end{aligned}$$

where $\lambda_{weibull}$ is the reciprocal of the scale parameter and b is the shape parameter.

In the simulation, under the assumption of generalized pareto distribution of interarrival time, we set $k = 1$ and $\theta = \lambda k$. Under the assumption of weibull distribution of

interarrival time, we set $b = 1.5$. Similar to the evaluation performed above, the minimum average system delay performance of non-aggregated scenario and aggregated scenario is evaluated for the RAOSC-DSC case and CBOSC-DSC case. The other simulation settings are the same as the settings above. The simulation results are shown in Fig. 17 and Fig. 18. It can be observed that based on the given simulation settings for the case of RAOSC-DSC the minimum average system delay of the aggregated scenario is still always lower than that of the non-aggregated scenario, while for the case of CBOSC-DSC, the minimum average system delay of the aggregated scenario is lower than that of the non-aggregated scenario for light traffic condition and vise versa. These results are consistent with the results based on the assumption of Poisson arrival process.

V. Conclusion

In this paper, two cases of heterogeneous OSC-DSC wireless networks are considered for aggregation and non-aggregation scenarios. In the first case, the heterogeneous RAOSC-DSC network is investigated. Given the assumptions that requests arrive according to Poisson process and the request size is exponentially distributed, it is proved that the minimum average system delay of the aggregated scenario is always lower than that of the non-aggregated scenario. An efficient method is proposed to approximate the optimal requests splitting ratio in the aggregated scenario. The analytical results when applying the non-aggregated scenario and simulation results when applying the aggregation system are also presented. In the second case, the heterogeneous CBOSC-DSC network is studied. The average system delay is derived for both the non-aggregated and aggregated scenarios. Extensive simulation results imply that, when contention and backoff mechanism is considered, the non-aggregated scenario outperforms the aggregated one under certain conditions. This is because the backoff penalty caused by aggregation exceeds the benefit from splitting the request.

References

[1] *Cisco visual networking index: Forecast and methodology, 2014-2019*, Cisco, San Jose, CA, USA, 2015.

[2] *Mobility Report on the Pulse of the Networked Society*, Ericsson, Stockholm, Sweden, 2015.

[3] *An Internet of Everything that Works for Everyone*, Qualcomm, San Diego, CA, USA, May 2015.

[4] *Visible Light Communication (VLC)—A Potential Solution to the Global Wireless Spectrum Shortage*, G. Research, London, U.K., 2011.

[5] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1994.

[6] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, "REX: A randomized exclusive region based scheduling scheme for mmWave WPANs with directional antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 113–121, Jan. 2010.

[7] S. Wu, H. Wang, and C. H. Youn, "Visible light communications for 5G wireless networking systems: From fixed to mobile communications," *IEEE Netw.*, vol. 28, no. 6, pp. 41–45, Nov. 2014.

[8] S. Shao, A. Khreishah, and I. Khalil, "Joint link scheduling and brightness control for greening VLC–based indoor access networks," *J. Opt. Commun. Netw.*, vol. 8, no. 3, pp. 148–161, 2016.

[9] X. Ortiz and A. Kaul, "Small cells: Outdoor Pico and micro markets, 3G/4G solutions for metro and rural deployments," *ABI Res.*, vol. 5, 2011.

[10] X. Bao, X. Zhu, T. Song, and Y. Ou, "Protocol design and capacity analysis in hybrid network of visible light communication and OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 63, no. 4, pp. 1770–1778, May 2014.

[11] X. Li, R. Zhang, and L. Hanzo, "Cooperative load balancing in hybrid visible light communications and WiFi," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1319–1329, Apr. 2015.

[12] T. D. C. Little and M. Rahaim, "Network topologies for mixed RF-VLC HetNets," in *Proc. IEEE Summer Topicals Meeting Ser.*, Jul. 2015, pp. 163–164.

[13] D. A. Basnayaka and H. Haas, "Hybrid RF and VLC Systems: Improving user data rate performance of VLC Systems," in *Proc. IEEE VTC Spring*, 2015, pp. 1–5.

[14] A. L. Ramaboli, O. E. Falowo, and A. H. Chan, "Bandwidth aggregation in heterogeneous wireless networks: A survey of current approaches and issues," *J. Netw. Comput. Appl.*, vol. 35, no. 6, pp. 1674–1690, Nov. 2012.

[15] W. Guo, Q. Li, H.-Y. Yu, and J.-H. Liu, "A parallel transmission MAC protocol in hybrid VLC-RF network," *J. Commun.*, vol. 10, no. 1, pp. 80–85, 2015.

[16] S. Shao *et al.*, "An indoor hybrid WiFi-VLC internet access system," in *Proc. IEEE 11th Int. Conf. Mobile Ad Hoc Sensor Syst.*, Oct. 2014, pp. 569–574.

[17] S. Shao *et al.*, "Design and analysis of a visible-light-communication enhanced wifi system," *Opt. Commun. Netw.*, vol. 7, no. 10, pp. 960–973, 2015.

[18] M. Ayyash *et al.*, "Coexistence of WiFi and LiFi toward 5G: Concepts, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 64–71, Feb. 2016.

[19] M. B. Rahaim, A. M. Vegni, and T. D. Little, "A hybrid radio frequency and broadcast visible light communication system," in *Proc. IEEE GLOBECOM*, 2011, pp. 792–796.

[20] O. Tickoo and B. Sikdar, "Modeling queueing and channel access delay in unsaturated IEEE 802.11 random access MAC based wireless networks," *IEEE/ACM Trans. Netw*, vol. 16, no. 4, pp. 878–891, Aug. 2008.

[21] N. Ding, D. Wagner, X. Chen, A. Pathak, Y. C. Hu, and A. Rice, "Characterizing and modeling the impact of wireless signal strength on smartphone battery drain," *ACM SIGMETRICS*, vol. 41, no. 1, 2013, pp. 29–40, 2013.

[22] J. Vucic, C. Kottke, S. Nerreter, K.-D. Langer, and J. W. Walewski, "513 Mbit/s visible light communications link based on DMT-modulation of a white LED," *JLT, IEEE/OSA*, vol. 28, no. 24, pp. 3512–3518, Dec. 2010.

[23] M. Baz, P. D. Mitchell, and D. A. J. Pearce, "Analysis of queuing delay and medium access distribution over wireless multihop PANs," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 2972–2990, Jul. 2015.

[24] *Queueing Theory Tutorial*, accessed on Jan. 15, 2016. [Online]. Available: http://web.mit.edu/dimitrib/www/OPNET_Full_Presentation.ppt

[25] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Computer Society LAN MAN Standards Committee, 1997.

[26] S. K. Nobar, K. A. Mehr, and J. M. Niya, "Comprehensive performance analysis of IEEE 802.15.7 CSMA/CA mechanism for saturated traffic," *JOCN, IEEE/OSA*, vol. 7, no. 2, pp. 62–73, Feb. 2015.

[27] Z. Huang and Y. Ji, "Design and demonstration of room division multiplexing-based hybrid vlc network," *Chin. Opt. Lett.*, vol. 11, no. 6, p. 060603, 2013.

[28] R. Zhang, J. Wang, Z. Wang, Z. Xu, C. Zhao, and L. Hanzo, "Visible light communications in heterogeneous networks: Paving the way for user-centric design," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 8–16, Apr. 2015.

[29] B. Pi. *Banana Pi-A Highend Single-Board Computer*, accessed on Jun. 16, 2016. [Online]. Available: http://www.bananapi.org/

[30] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 41–48, Sep. 2009.

[31] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A Stackelberg game approach," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 538–549, Apr. 2012.

[32] N. Saquib, E. Hossain, L. B. Le, and D. I. Kim, "Interference management in OFDMA femtocell networks: Issues and approaches," *IEEE Wireless Commun.*, vol. 19, no. 3, pp. 86–95, Jun. 2012.

[33] H. Li, X. Xu, D. Hu, X. Qu, X. Tao, and P. Zhang, "Graph method based clustering strategy for femtocell interference management and spectrum efficiency improvement," in *Proc. IEEE WiCOM*, Oct. 2010, pp. 1–5.

[34] H. Widiarti, S.-Y. Pyun, and D.-H. Cho, "Interference mitigation based on femtocells grouping in low duty operation," in *Proc. IEEE VTC 2010-Fall*, 2010, pp. 1–5.
[35] H.-C. Lee, D.-C. Oh, and Y.-H. Lee, "Mitigation of inter-femtocell interference with adaptive fractional frequency reuse," in *Proc. IEEE ICC*, May 2010, pp. 1–5.
[36] T.-H. Kim and T.-J. Lee, "Throughput enhancement of macro and femto networks by frequency reuse and pilot sensing," in *Proc. IEEE IPCCC*, Dec. 2008, pp. 390–394.
[37] L. Zhang, L. Yang, and T. Yang, "Cognitive interference management for LTE-A femtocells with distributed carrier selection," in *Proc. IEEE VTC 2010-Fall*, Sep. 2010, pp. 1–5.
[38] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
[39] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.
[40] S. Hranilovic, *Wireless Optical Communication System*. Berlin, Germany: Springer Science + Business Media, 2006.
[41] B. C. Arnold, *Pareto distribution*. Hoboken, NJ, USA: Wiley, 2015.
[42] R. P. Covert and G. C. Philip, "An EOQ model for items with Weibull distribution deterioration," *AIIE Trans.*, vol. 5, no. 4, pp. 323–326, 1973.

Sihua Shao (S'14) received the B.S. degree in electrical and information engineering from the South China University of Technology in 2011, and the M.S. degree in electrical and information engineering from The Hong Kong Polytechnic University in 2012. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology. His current research interests include wireless communication, visible light communication, and heterogeneous network.

Abdallah Khreishah (M'09) received the B.S. degree in computer engineering from the Jordan University of Science and Technology in 2004 and the M.S. and Ph.D. degrees in electrical and computer engineering from Purdue University in 2006 and 2010. While pursuing the Ph.D. studies, he was with NEESCOM. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology. His research interests fall in the areas of visible-light communication, green networking, network coding, wireless networks, and network security. He is the Chair of the North Jersey IEEE EMBS Chapter.