Performance Optimization of IEEE 802.11ax UL OFDMA Random Access

Pengxue Liu, Yitong Li, and Dalong Zhang

***Abstract*—This paper presents an extensive analysis of the IEEE 802.11ax uplink orthogonal frequency-division multiple access (OFDMA)-based random access (UORA) mechanism, addressing inherent inefficiencies in channel access under varying network loads. Specifically, a mathematical model is developed to analyze the system performance of the 802.11ax UORA protocol, enabling the characterization of steady-state operating points under both saturated and unsaturated conditions. Key performance metrics, including system efficiency and mean access delay, are derived as functions of the steady-state operating points. Optimization of these performance metrics through the appropriate selection of backoff parameters is explored, with the analysis validated by simulation results. Additionally, the effects of access parameter heterogeneity, multi-link operation (MLO) and multiple resource unit (MRU) operation capabilities on the performance of IEEE 802.11ax UORA mechanism are further discussed.**

***Index Terms*—IEEE 802.11ax UORA, performance optimization, random access.**

I. Introduction

A. Background

Wi-FI technology [1] offers cost-effective, scalable local connectivity, enhancing indoor coverage and complementing cellular networks in environments such as smart homes, corporate offices, industrial sites, and public venues. Unlike centrally managed cellular networks [2], which are optimized for wide-area coverage and mobility, Wi-Fi networks are designed for local area communication and employ a decentralized random access mechanism, where stations (STAs) autonomously decide when to transmit data. However, this flexibility also introduces challenges, as evidenced by the well-known IEEE 802.11 distributed coordination function (DCF) protocol [3], which implements carrier sense multiple access with collision avoidance (CSMA/CA). This protocol illustrates key limitations when compared to the centralized access control in cellular networks. First, the absence of centralized scheduling results in competitive channel access, often causing network congestion and inefficient spectrum utilization in high-density scenarios. Second, Wi-Fi's spectrum efficiency declines significantly as multiple devices contend for the same channel [4], [5], unlike the optimized resource allocation in cellular systems. To mitigate these issues, the IEEE 802.11ax standard introduces several enhancements, the most notable being uplink (UL) multiuser (MU) orthogonal frequency-division multiple access (OFDMA) [6], [7].

The purpose of the MU-OFDMA feature is to enable multiple STAs to access the resource units (RUs) simultaneously, thus enhancing the efficiency of UL MU transmission. UL MU-OFDMA strategies can be classified into two primary categories: Scheduled access (SA) [8] and random access (RA) [9]. As the number of devices increases rapidly, monitoring the buffering status of all STAs by access points (APs) becomes more challenging, yet is crucial for efficient scheduling. This monitoring is pivotal because it allows APs to allocate bandwidth and time slots based on the real-time data needs and traffic conditions of each STA, which optimizes network performance and reduces latency. Consequently, the uplink OFDMA random access (UORA) mechanism has been introduced, enabling STAs with unknown buffering states to simultaneously participate in UL MU transmission. Recent studies have explored hybrid access (HA) mechanisms that integrate RA and SA strategies [10]–[13]. Upon data generation by an STA, a buffer status report (BSR) is generated and transmitted to the AP using the UORA protocol. Once the BSR has been successfully transmitted, the AP decodes it and schedules the corresponding STA for the data communication phase. However, the low efficiency of the UORA mechanism can hinder the overall effectiveness of the HA mechanism [12]. Consequently, comprehensive analysis and optimization of the UORA mechanism's performance is essential.

B. Related Works and Motivation

Extensive studies [14]–[17] have conducted performance analyses of the UORA protocol using the two-dimensional Markov chain model introduced in [3]. Studies [14] and [15] analyzed the performance of the UORA protocol under saturated conditions, while [16] investigated the coexistence of IEEE 802.11ax and legacy STAs in non-saturated scenarios. Additionally, [17] examined the impact of bursty traffic on network performance. Despite extensive investigation of the UORA protocol under various network conditions by these studies, they primarily focus on solving complex nonlinear

Manuscript received June 21, 2024; revised October 12, 2024; approved for publication by Shim, Byonghyo Division 4 Editor, November 21, 2024.

The work of D. Zhang was supported in part by the National Natural Science Foundation of China under Grant U22A2001, and in part by the Major Science and Technology Projects of Henan Province under Grant 241110210100. The work of Y. Li was supported by the National Natural Science Foundation of China under Grant 61801433.

P. Liu and Y. Li are with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China, email: pengxuelzzu@163.com, ieytli@zzu.edu.cn.

D. Zhang is with the School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450001, China, email: iedlzhang@zzu.edu.cn.

Y. Li is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2024.000069

1229-2470/24/$10.00 © 2024 KICS

equations to determine steady-state operating points, which poses challenges in achieving optimal network configurations.

The aforementioned studies primarily focus on the numerical calculating throughput and delay for given network configuration. However, how to adjust system parameters to dynamically optimize network performance remains an insufficiently addressed issue. In existing IEEE 802.11 DCF networks, many studies have shown that it can be optimize network performance by adjusting backoff parameters [18]–[23]. Similarly, the OFDMA backoff (OBO) mechanism in the IEEE 802.11ax UORA protocol requires tuning based on network conditions, which attracts much attention as well [24]–[27]. [24] introduced a cognitive backoff (CB) mechanism that dynamically adjusts the contention window (CW) size according to the measured conditional collision probability. Concurrently, the study in [25] developed a straightforward yet effective OBO control strategy (OBO-CTRL), enabling each STA to autonomously set its backoff value using a self-tunable parameter derived from transmission results. Building upon this, [26] proposed the Efficient OFDMA random access backoff (E-OBO) algorithm, which adapts the channel access probability in response to specific network conditions detected by the AP. Moreover, to minimize the wastage of RUs, [27] implemented the repeated competition (ReCo) concept within the IEEE 802.11ax framework, allowing STAs to engage in an additional backoff round if RUs remain unused post-initial contention. The results indicate that the proposed algorithm can significantly enhance the performance of the IEEE 802.11ax UORA mechanism.

Existing literature predominantly assumes a fixed backoff factor of $q = 1/2$ and a relatively small initial contention window for the IEEE 802.11ax UORA mechanism, which may result in suboptimal network performance or significant degradation as network size or traffic load increases under these default settings. Optimizing the backoff factor q thus becomes crucial. Distinct from previous methodologies, our study focuses on enhancing the IEEE 802.11ax UORA mechanism by theoretically analyzing and explicitly optimizing the backoff factor q within the existing standards.

The preceding discussions have primarily focused on homogeneous networks where all STAs possess identical access parameters. However, in practical network environments, the quality of service (QoS) requirements often vary among STAs. Thus, it is crucial to further analyze and optimize the performance of the heterogeneous IEEE 802.11ax UORA mechanism. Research on heterogeneous IEEE 802.11ax networks has been addressed in [28]–[30], where discussions included spatial reuse, fair channel sharing, and collision probability, considering diverse carrier sensing capabilities across STAs. Moreover, the current UORA scheme in the IEEE 802.11ax standard limits each STA to competitively access only one eligible random access resource unit (RA-RU) [6], [7], a restriction that becomes more pronounced in complex and demanding network environments. To mitigate these constraints and enhance network performance, the multi-link operation (MLO) and multiple resource unit (MRU) capabilities, which allow a single STA to maintain multiple links with APs and utilize multiple RUs, have been proposed for integration into the design of next-generation Wi-Fi protocols [31], [32], potentially improving overall spectral efficiency.

C. Contributions and Organization

Our contributions and main results of this work can be summarized as follows:

- An analytical framework is developed to analyze the performance of the IEEE 802.11ax UORA mechanism, characterizing the steady-state distribution based on the successful transmission probability of an HOL packet, p. Besides, depending on network conditions being unsaturated or saturated, two steady-state operating points are defined: The desired stable point p_L and the undesired stable point p_A, each explicitly derived as functions of the system parameters. Furthermore, the key performance metrics, including the system efficiency and mean access delay, are characterized, which are further optimized by properly selecting the backoff parameters. The results indicate that system performance is insensitive to the backoff parameter in unsaturated networks. In contrast, in saturated networks, system performance becomes closely dependent on backoff parameter.
- The effects of STA access parameter heterogeneity and the integration of MLO and MRU capabilities are discussed to provide insights into practical network design. Our analysis reveals that STAs with larger decrement values can complete the backoff process more rapidly and achieve superior performance in scenarios where different STAs decrement their backoff counters by varied amounts. Furthermore, with the incorporation of MLO and MRU capabilities, allowing STAs to randomly select multiple RUs for transmitting the same packet simultaneously can enhance system performance. However, if each STA randomly selects an excessive number of RUs, the probability of collision escalates significantly, thereby diminishing overall system performance.

The organization of this paper is as follows. In Section II, the IEEE 802.11ax UORA mechanism is described. The preliminary analysis is established in the Section III. The system efficiency and mean access delay are discussed in Section IV and Section V respectively. The impact of STAs' access parameter heterogeneity, MLO and MRU on the performance of IEEE 802.11ax UORA mechanism are presented in Section VI. Finally, the conclusion is summarized in Section VII. The main notations used are listed in Table I.

II. IEEE 802.11AX UPLINK OFDMA RANDOM ACCESS MECHANISM

This section provides a concise overview of the IEEE 802.11ax UORA mechanism, as defined in the IEEE 802.11ax standard. For further details on the IEEE 802.11ax UORA mechanism, please refer to [6].

The IEEE 802.11ax standard posits that an AP can continuously acquire channels by using a contention parameter with higher priority [33]. As depicted in Fig. 1, a basic service set (BSS) consists of a single AP transmitting trigger frames (TFs)

TABLE I
MAIN NOTATIONS.

Variables	Definition
n	Number of stations
M	Number of resource units
m	Cutoff phase of HOL packets
k	Backoff counter
W	Initial backoff window size
q	Backoff factor
t	the interval between two adjacent TFs
p	Steady-state probability of successful transmission of HOL packets
λ	Arrive rate of each station
$\hat{\lambda}_{\text{out}}$	Network throughput
$\hat{\lambda}_{\max}$	Maximum network throughput
$\hat{\lambda}_{e_{\text{out}}}$	System efficiency
$\hat{\lambda}_{e_{\max}}$	Maximum system efficiency

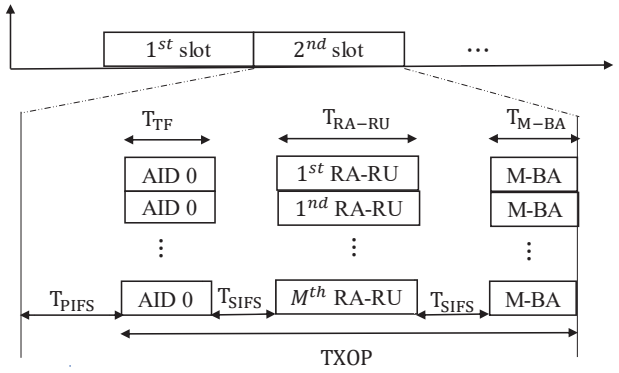


Fig. 1. Timing diagram of the IEEE 802.11ax UORA mechanism.

with M RUs to n STAs per slot, with n representing the number of STAs and M representing the number of RUs. The slot duration equals the sum of one point coordination function interframe space (PIFS) interval and one transmission opportunity (TXOP), in which the specific composition of TXOP is presented in Fig. 1. In alignment with previous studies [10], [14], [17], [34], it is assumed that both the number of RUs M and the length of time slots t remain constant at each trigger.

Within the IEEE 802.11ax UORA framework, after receiving a TF, each STA selects a random value, known as the OBO value, in a contention window of size OCW. If the OBO value selected by the STA does not exceed the number of available RA-RUs notified in the received TF, then the STA involved randomly chooses an RA-RU for transmission. Otherwise, the STA decrements the backoff counter by the number of RUs M and waits for the next TF. Note that OCW depends on the number of failed retransmissions encountered for the HOL packet. At the first transmission, contention window of size OCW is set to the minimum contention window OCW_{min}. After a successful transmission, the corresponding OCW is

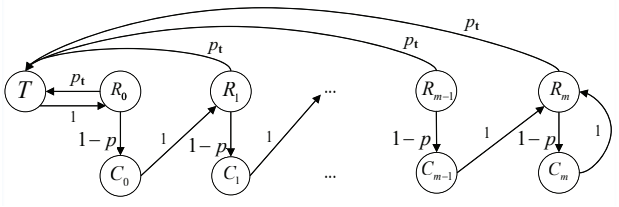


Fig. 2. Embedded Markov chain X_j of the state transition process of an individual HOL packet in IEEE 802.11ax UORA mechanism.

updated to its original value OCW_{min}. In the event of a collision, the OCW is increased incrementally until it reaches its maximum value $OCW_{\max}$. For simplicity, the notations $W = OCW_{min} + 1$ and $W_i = W \cdot q^{-i}$ $(i = 0, \cdots, m)$ are introduced to denote the number of possible OBO values in the initial and subsequent backoff stages, respectively, where q represents the backoff factor and m denotes the cutoff phase.

Note that both legacy IEEE 802.11 DCF and IEEE 802.11ax UORA mechanisms implement the binary exponential backoff (BEB) policy to manage the CW, with a fixed backoff factor $q = 0.5$. However, notable differences arise in their operation. Unlike the IEEE 802.11 DCF networks, where the decrement of the backoff counter is a straightforward count down by one after each idle time slot, the OBO value in the IEEE 802.11ax UORA mechanism is updated by reducing the number of available RUs immediately upon receiving a TF. This modification is designed to enable more dynamic and efficient resource management in environments with high network density.

III. PRELIMINARY ANALYSIS

In this paper, we consider a scenario where n STAs communicate with a single AP using the IEEE 802.11ax UORA mechanism. The communication channel is assumed to be error-free, indicating that transmission errors are exclusively attributable to collisions and are not influenced by the physical layer's channel quality. Each STA is characterized by a constant arrival rate λ and is equipped with an infinite buffer size.

A. System Model

This paper extends the analytical framework previously developed for 802.11 DCF networks [21] to analyze the IEEE 802.11ax UORA mechanism. Consistent with the methodologies outlined in [21], the status of a HOL packet transmitted at the jth transition is denoted by X_j and the epoch at which the jth transition occurs is represented by V_j. The behavior of each HOL packet is modeled as a discrete-time Markov renewal process (X, V)=$\{(X_j, Y_j), j = 0, 1, \cdots\}$ based on the backoff process.

As depicted in Fig. 2, the states of X_j are classified into the following three categories: (1) waiting to request (R_i), (2) collision state (C_i), and (3) successful state (T). A HOL packet transitions from the State R_i to the State T upon a successful transmission, while a failed transmission causes the packet to remain in the C_i state, and it will transition to

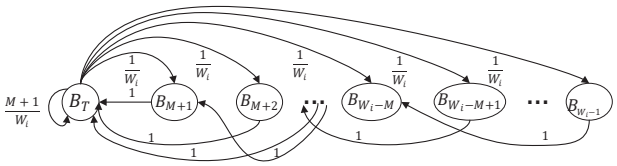


Fig. 3. State transition diagram of a State-R_i HOL packet in IEEE 802.11ax UORA mechanism, $i = 0, \cdots, m$.

the R_{i+1} state after the collision. The variable i represents the number of collisions experienced by the HOL packet, the continuous collision will cause i to increase continuously until the maximum backoff stage m is reached, as shown in Fig. 2. p_t denotes the probability of successful transmission at time slot t. The steady-state probability distribution of the embedded Markov chain in Fig. 2 can be expressed as

$$\pi_{R_i} = \begin{cases} (1-p)^i \pi_T, & i = 0, \cdots, m-1, \\ \frac{(1-p)^m}{p} \pi_T, & i = m, \end{cases} \tag{1}$$

and

$$\pi_{C_i} = \pi_{R_i} \cdot (1-p), \quad i = 0, \cdots, m, \tag{2}$$

where $p = \lim_{t \to \infty} p_t$. The holding time of the state X_j is $V_{j+1} - V_j$, i.e., the interval between successive switches. Specifically, in this paper, it is assumed that TFs are sent at fixed time intervals of T_I (i.e., one time slot), the holding time in successful transmission τ_T and the holding time in collision τ_C can then be written as

$$\tau_C = \tau_T = T_I. \tag{3}$$

While the mean holding time τ_{R_i} in state R_i is influenced by the backoff parameter and the number of RUs M. As shown in Fig. 3, the HOL packet's backoff counter is decremented by M as soon as it receives the TF for random access. Based on the above assumptions, the duration required in the ith backoff stage can be expressed as

$$d_i = \left\lceil \frac{k_0}{M} \right\rceil T_I, \quad i \in [0, m], \tag{4}$$

where k_0 is the initial backoff value randomly selected by the HOL packet at the ith backoff stage. Since k_0 follows a uniform distribution at $[0, W_i - 1]$, the mean holding time τ_{R_i} in state R_i can be given as

$$\tau_{R_i} = \sum_{k_0=0}^{W_i-1} \left\lceil \frac{k_0}{M} \right\rceil \frac{T_I}{W_i}, \quad i \in [0, m], \tag{5}$$

where $W_i = W \cdot q^{-i}$. Eventually, the steady-state probabilities $\widetilde{\pi}_j$ of Markov renewal process (X, V) can be acquired as

$$\widetilde{\pi}_j = \frac{\pi_j \cdot \tau_j}{\pi_T \tau_T + \sum_{i=0}^{m} \pi_{R_i} \tau_{R_i} + \sum_{i=0}^{m} \pi_{C_i} \tau_C}, \tag{6}$$

for $j \in S$, in which S is the state space of X. By substituting (1) and (2) into (6), the probability that the HOL packet is in state T can be derived as

$$\widetilde{\pi}_T = \frac{t}{\sum_{i=0}^{m-1} (1-p)^i \cdot \tau_{R_i} + \frac{(1-p)^m}{p} \cdot \tau_{R_m} + \frac{t}{p}}. \tag{7}$$

Note that the queue output is successful only when the HOL packet is in state T, so π_T also represents the service rate of each STA's queue. Let λ denote the input rate of each STA's queue. The offered load of each STA's queue, denoted as ρ, can be written as

$$\rho = \lambda / \widetilde{\pi}_T. \tag{8}$$

B. Steady-state Point Analysis

In this subsection, the limiting probability of successful transmission p is analyzed based on the above system model. Specifically, it is closely determined by the aggregate input rate $\hat{\lambda} = n\lambda$. As we will demonstrate in this subsection, when $\hat{\lambda} \leq Me^{-1}$, each STA's queue has a non-zero probability to be empty, and the network will operate at the unsaturated condition. As $\hat{\lambda}$ increases, however, all the STA's queue are busy with probability 1, and the network eventually becomes saturated. The rest of this subsection will focus on the derivation of the limiting probability of successful transmission p in both unsaturated and saturated scenarios.

1) Steady-state point of unsaturated conditions: Precisely, for every HOL packet, its successful transmission requires all other $n-1$ STAs that do not compete for the channel, i.e., all other $n-1$ STAs's queue are empty, or non-empty but the HOL packets are not requested to be transmit. According to the backoff mechanism of IEEE 802.11ax UORA mechanism, the HOL packet can request transmission only when its corresponding backoff counter is not larger than the number of RU M. In the unsaturated condition, the probability that a STA's queue is empty is given by $1 - \rho$, and the probability that a STA has a HOL packet but does not send the packet is denoted by $\rho(1 - \sum_{i=0}^{m} \widetilde{\pi}_{R_i} r_i)$. As presented in Appendix A, the probability of the HOL packet requesting transmission under the condition that the backoff counter k is not greater than the number of RUs M can be calculated as follows:

$$r_i = \frac{W_i}{W_i - \frac{M}{2} \left\lfloor \frac{W_i - 1}{M} \right\rfloor^2 + \left(W_i - 1 - \frac{M}{2}\right) \left\lfloor \frac{W_i - 1}{M} \right\rfloor}, \tag{9}$$

for $i = 0, \cdots, m$. Therefore, by combining (8) and (9), the steady-state probability that the HOL packet making a successful transmission, denoted as p, can be written as

$$\begin{aligned} p &= \left\{ 1 - \rho + \rho \left(1 - \sum_{i=0}^{m} \widetilde{\pi}_{R_i} \frac{r_i}{M} \right) \right\}^{n-1} \\ &\overset{\text{for large } n}{\approx} \exp\left\{ -\hat{\lambda} \sum_{i=0}^{m} \frac{\widetilde{\pi}_{R_i} r_i}{\widetilde{\pi}_T M} \right\}. \end{aligned} \tag{10}$$

Finally, by combining equations (1)–(2), (5)–(6), and (9)–(10) can be further given as

$$\begin{aligned} p &= \exp\left\{ -\hat{\lambda} \sum_{i=0}^{m} \frac{\pi_{R_i} \tau_{R_i} r_i}{\pi_T \tau_T M} \right\} \\ &= \exp\left\{ \frac{-\hat{\lambda}}{Mp} \right\}. \end{aligned} \tag{11}$$

The derivation of (11) is provided in detail in Appendix B. Specifically, (11) has two non-zero roots:

$$p_L = \exp\left\{\mathbb{W}_0(-\hat{\lambda}/M)\right\}, \tag{12}$$

and

$$p_S = \exp\left\{\mathbb{W}_{-1}(-\hat{\lambda}/M)\right\}, \tag{13}$$

if the aggregate input rate $\hat{\lambda}$ does not exceed

$$\hat{\lambda}_{\max} = Me^{-1}, \tag{14}$$

where $\mathbb{W}_0(\cdot)$ and $\mathbb{W}_{-1}(\cdot)$ in (12)–(13) represents two branches of the Lambert W function [35].

2) Steady-state point of saturated conditions: With the aggregate input rate $\hat{\lambda}$ increasing, the queues of all STAs are non-empty and eventually the network will become saturated. In this case, the steady-state probability p can be expressed as

$$p = \left\{1 - \sum_{i=0}^{m} \frac{\widetilde{\pi}_{R_i} r_i}{M}\right\}^{n-1} \overset{\text{with large } n}{\approx} \exp\left\{-n \sum_{i=0}^{m} \frac{\widetilde{\pi}_{R_i} r_i}{M}\right\}. \tag{15}$$

By substituing (1)–(2), (5)–(6), and (9) into (15), we have

$$p = \exp\left\{\frac{-n/M}{p\sum_{i=0}^{m-1}(1-p)^i \cdot \frac{\tau_{R_i}}{t} + (1-p)^m \cdot \frac{\tau_{R_m}}{t} + 1}\right\}, \tag{16}$$

where $i = 0, \cdots, m$. With $q = 1/2$, It can be proved that $1-p$ in this paper is consistent with (10) in [10].

Specifically, when the initial backoff window size W is large enough and the number of RU M is relatively small, by combining (15) and Appendix B, it can be further obtained that

$$\begin{aligned} p_A &\overset{W \gg M}{\approx} \exp\left\{\frac{-n/M}{p\sum_{i=0}^{m-1}(1-p)^i \cdot \frac{W_i}{2M} + (1-p)^m \frac{W_m}{2M} + \frac{3}{2}}\right\} \\ &= \exp\left\{\frac{-2 \cdot n}{W\left(\frac{pq}{p+q-1} - \left(\frac{pq}{p+q-1} - 1\right)\left(\frac{1-p}{q}\right)^m\right) + 3M}\right\}. \end{aligned} \tag{17}$$

It can be clearly seen from (17) that p_A is closely related to the number of STAs n and backoff parameters, where the backoff parameters include the cutoff phase m, the initial backoff window size W, the number of RU M and the backoff factor q. When $m = \infty$, p_A can be explicity expressed as

$$p_A^{m=\infty} \overset{\text{for large } W}{\approx} \frac{2n(1-q)/(Wq)}{\mathbf{W}_0(2n(1-q)/(Wq)\exp(2n/(Wq)))}. \tag{18}$$

C. Simulation Results

In this subsection, simulation results are presented to verify the theoretical analysis. In this paper, event-driven simulations are conducted using MATLAB, with each simulation run consisting of 10^6 time slots. The configuration adheres to the parameters established by the system model. Specifically, the number of RUs, denoted by M, is set to 9 and 37, corresponding to the upper limits of RA-RUs allowed in the 20 MHz and 80 MHz frequency bands, respectively, as specified in the IEEE 802.11ax standard. The steady-state probability of successful transmission of HOL packets, denoted by p, is calculated as the ratio of successfully transmitted packets to the total number of transmitted packets.

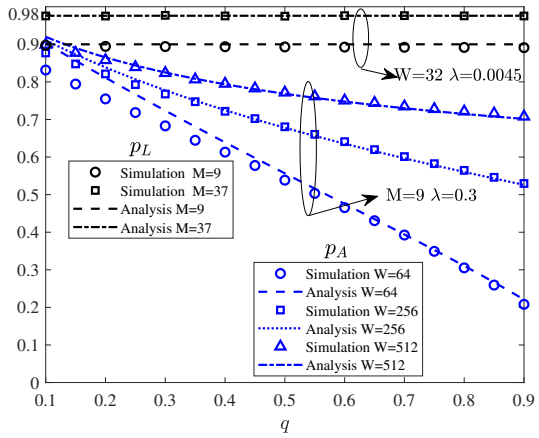


Fig. 4. Steady-state operating points versus backoff factor q in IEEE 802.11ax UORA mechanism. $n = 200$ and $m = \infty$.

Fig. 4 illustrates the variation in steady-state probability of successful transmission for HOL packets as a function of the backoff factor q. It is evident from Fig. 4 that at $\hat{\lambda} = 0.9$ and $n = 500$, the network operates at the desired stable point p_L, which remains invariant with respect to q. Furthermore, Fig. 4 demonstrates a monotonic increase in p_L with the number of RUs, M, aligning with (12). Conversely, at the undesired stable point p_A, depicted in Fig. 4, the undesired stable point for cutoff phase $m = \infty$ $p_A^{m=\infty}$ demonstrate a monotonic decline with an increase in backoff factor q, or an increase with a larger initial backoff window size W. This behavior contrasts with the desired stable point p_L, indicating a significant dependence of p_A on system parameters.

IV. Analysis and Optimization of System Efficiency

In this section, the system efficiency $\hat{\lambda}_{e_{\text{out}}}$, which defines the average number of access requests that can be successfully sent in each RU per time slot, is optimized. Note that network throughput $\hat{\lambda}_{\text{out}}$ represents the average number of access requests successfully transmitted in each time slot. As we will demonstrate, the system efficiency is closely determined by the aggregate input rate $\hat{\lambda}$ and the backoff factor q.

A. Stable Region and Maximum System Efficiency

When the aggregate input rate $\hat{\lambda}$ is low, the network will initially operates at the desired stable point p_L, and a stable throughput $\hat{\lambda}_{\text{out}} = \hat{\lambda}$ can be achieved. As $\hat{\lambda}$ increases, the network operating point may shift to the undesired stable point p_A, and network throughput is given by $\hat{\lambda}_{\text{out}} = n \cdot \widetilde{\pi}_T = -M \cdot p_A \ln p_A$. Accordingly, the corresponding system efficiency can be denoted as

$$\hat{\lambda}_{e_{\text{out}}} = \frac{\hat{\lambda}_{\text{out}}}{M} = -p_A \ln p_A. \tag{19}$$

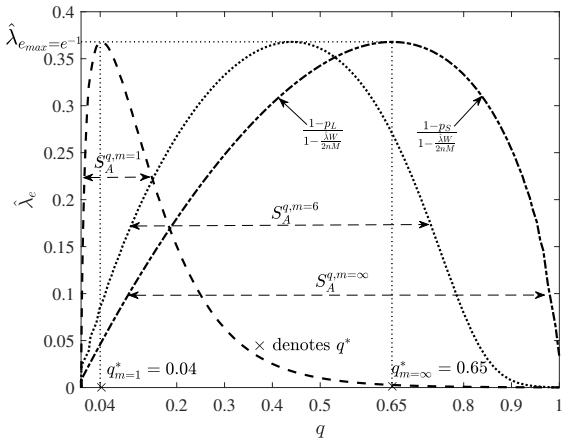


Fig. 5. Stable region of backoff factor q under various cutoff phase m. $n = 500$, $M = 9$, $W = 64$.

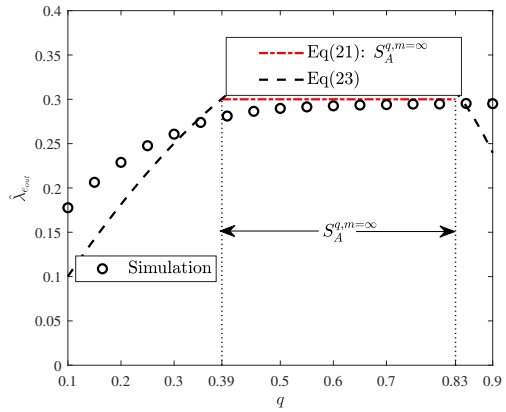


Fig. 6. System Efficiency $\hat{\lambda}_{e_{\text{out}}}$ versus backoff factor q. $\hat{\lambda} = 2.7$, $\hat{\lambda}_e = \hat{\lambda}/M = 0.3$, $n = 500$, $M = 9$, $W = 32$, $m = \infty$.

When the network operates at p_A, in order to achieve a stable throughput $\hat{\lambda}_{\text{out}} = \hat{\lambda}$, it is necessary to ensure that the aggregate service rate is not less than the aggregate input rate $\hat{\lambda}$, i.e., $n\widetilde{\pi}_T \geq \hat{\lambda}$. Therefore, it can be further inferred that in order to obtain a stable throughput at p_A, p_A has to satisfy the following condition:

$$p_S \leq p_A \leq p_L, \tag{20}$$

where p_L and p_S can be found in (12)–(13).

According to (17), it can be clearly observed that p_A is closely related to the backoff parameters, including the backoff factor q and the cutoff phase m. Specifically, Let $S_A^q = \{q | p_S \leq p_A \leq p_L\}$ denote the stable region with respect to the backoff factor q, in which the network throughput $\hat{\lambda}_{\text{out}}$ is equal to the aggregate input rate $\hat{\lambda}$. According to (12)–(13), it can be easily observed that p_L decreases as $\hat{\lambda}$ increases and p_S increases with $\hat{\lambda}$. Therefore, the stable region S_A^q will gradually decrease as the aggregate input rate $\hat{\lambda}$ increases, and it will be reduced to a single point when $\hat{\lambda}$ approaches to the maximum network throughput $\hat{\lambda}_{\max}$. As $\hat{\lambda} > \hat{\lambda}_{\max}$, S_A^q becomes to an empty set.

The range of the stable region is also affected by the cutoff phase m. Fig. 5 demonstrates how the stable region varies with the cutoff phase m. It can be observed that the stable region will be enlarged as the cutoff phase m increases. From an intuitive perspective, the larger the cutoff phase, the more space STAs can have to backoff to alleviate frequent collisions and enhance the network throughput. Consequently, to achieve stable throughput, the selection of the backoff factor q exhibits low sensitivity. Specifically, with $m < \infty$, p_A is an implicit function of q, and we can only obtain the numerical result of the stable region $S_A^{q,m}$ according to (17). On the other hand, with $m = \infty$, the undesired stable point $p_A^{m=\infty}$ can be obtained according to (18). The corresponding stable region can be explicitly written as

$$S_A^{q,m=\infty} = \left[\frac{1-p_L}{1-\frac{\hat{\lambda}W}{2nM}}, \frac{1-p_S}{1-\frac{\hat{\lambda}W}{2nM}} \right]. \tag{21}$$

According to (21), it can be observed that the stable range $S_A^{q,m=\infty}$ shrinks as the aggregate input rate $\hat{\lambda}$ increases. When the aggregate input rate $\hat{\lambda}$ increases to $\hat{\lambda}_{\max} = M \cdot e^{-1}$, $S_A^{q,m=\infty}$ eventually narrows to a single point

$$q^*_{m=\infty} = \frac{1-e^{-1}}{1-\frac{W}{2n}e^{-1}}. \tag{22}$$

Note that $q^*_{m=\infty}$ should not exceed 1, which requires that $W \leq 2n$. Otherwise, the maximum network throughput $\hat{\lambda}_{\max}$ cannot be achieved. With $q \in S_A^{q,m=\infty}$ and aggregate input rate $\hat{\lambda} \leq \hat{\lambda}_{\max}$, we can get a stable throughput $\hat{\lambda}_{\text{out}} = \hat{\lambda}$ at p_A. With $q \notin S_A^{q,m=\infty}$, the network throughput will be lower than the aggregate input rate $\hat{\lambda}$ and is related to the aggregate service rate. By combining (18) and (19), the system efficiency with $q \notin S_A^{q,m=\infty}$ is given by

$$\begin{aligned}\hat{\lambda}_{e_{\text{out}}}^{m=\infty} =& \frac{2n(q-1)}{qW} \\ &+ \frac{4Mn^2(1-q)/(W^2q^2)}{\mathbb{W}_0(2n(1-q)/(Wq)\cdot exp(2n/W/q))}.\end{aligned} \tag{23}$$

As the aggregate input rate $\hat{\lambda}$ exceeds the maximum network throughput $\hat{\lambda}_{\max}$, a stable throughput $\hat{\lambda}_{\text{out}} = \hat{\lambda}$ cannot be achieved. According to (19), the maximum system efficiency is given by $\hat{\lambda}_{e_{\max}} = e^{-1}$ when p_A is equal to e^{-1}. As a result, the optimal backoff factor q^* to achieve $\hat{\lambda}_{e_{\max}}$ is the root of the following equation

$$2n = W\left(\frac{e^{-1}q^*}{e^{-1}+q^*-1} - \left(\frac{e^{-1}q^*}{e^{-1}+q^*-1}-1\right)\left(\frac{1-e^{-1}}{q^*}\right)^m\right)+3M. \tag{24}$$

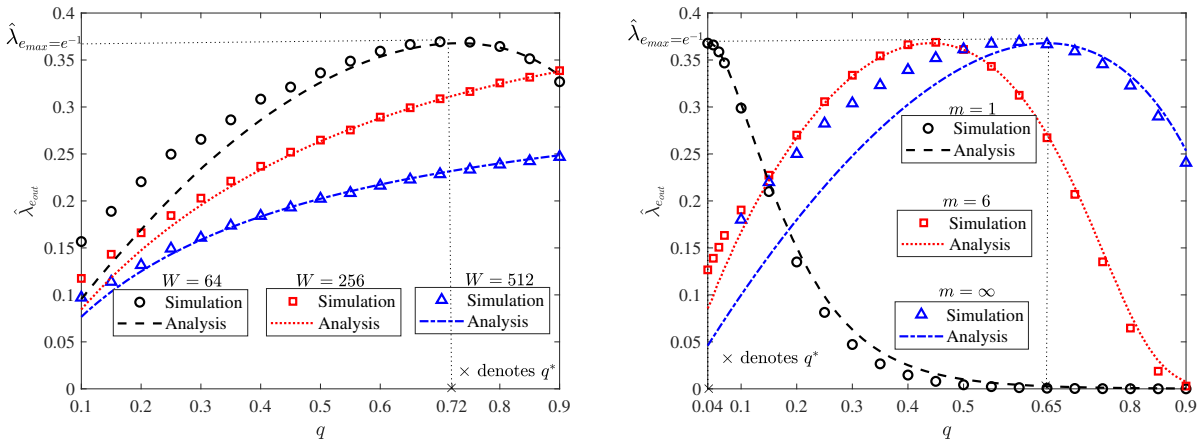
Specifically, the optimal backoff factor $q^*_{m=1}$ and $q^*_{m=\infty}$ for the cases that cutoff phase $m = 1$ and $m = \infty$ can be respectively denoted as[1]

$$q^*_{m=1} = \frac{1-e^{-1}}{\frac{2n-3M}{W} - e^{-1}}, \tag{25}$$

and

$$q^*_{m=\infty} = \frac{1-e^{-1}}{1-\frac{We^{-1}}{2n-3M}}. \tag{26}$$

[1]By following the assumption in (17), when W is large enough and M is relatively small, (26) can be further reduced to (22).



(a) (b)

Fig. 7. System efficiency $\hat{\lambda}_{e_{\text{out}}}$ versus backoff factor q in saturated IEEE 802.11ax UORA networks. $\lambda = 0.3$. (a) System efficient $\hat{\lambda}_{e_{\text{out}}}$ with various initial backoff window size W. $n = 100$, $M = 9$, $m = \infty$. (b) System efficiency $\hat{\lambda}_{e_{\text{out}}}$ with various cutoff phase m. $n = 500$, $M = 9$, $W = 64$.

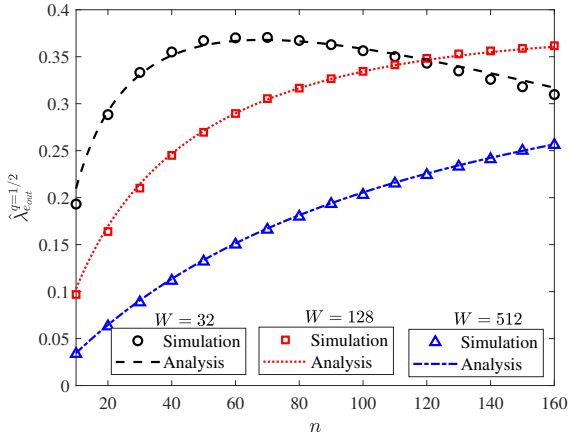


Fig. 8. System Efficiency $\hat{\lambda}_{e_{\text{out}}}^{q=1/2}$ versus the number of STAs n in saturated IEEE 802.11ax UORA networks. $\lambda = 1$, $M = 9$, $m = 3$.

B. Simulation Results

The above analysis is verified by the simulation results presented in this subsection. The simulation setting follows the system model described in Section III and thus the details are omitted here for brevity. In simulations, the system efficiency is obtained by calculating the ratio of the sum of all successfully transmitted packets to the total number of RUs M for all time slots.

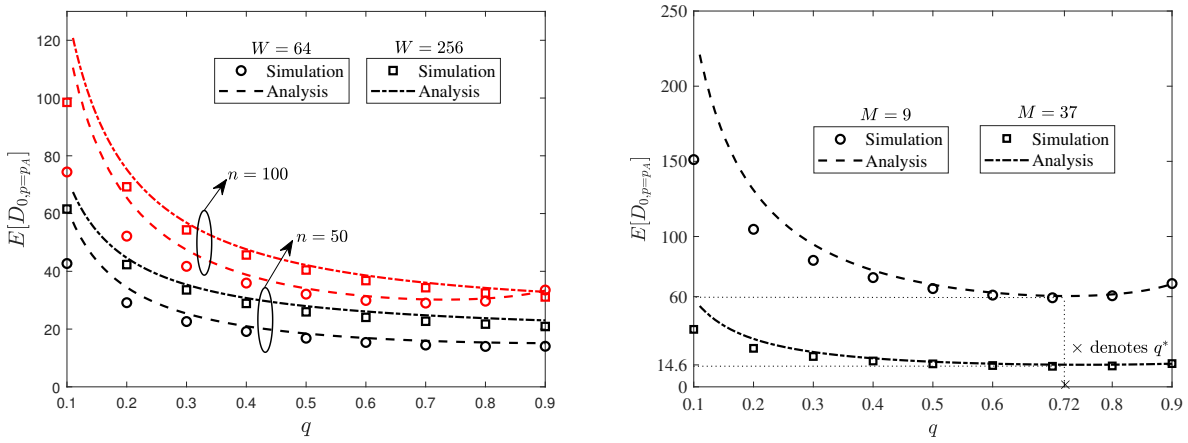
Fig. 6 illustrates the variation of system efficiency with the backoff factor q when the aggregate input rate $\hat{\lambda} = 2.7$. As demonstrated in Section IV, a stable throughput $\hat{\lambda}_{\text{out}} = \hat{\lambda}$ can be achieved when $p = p_A$ and q is within the stable region S_A^q. Outside this region, the network becomes unstable, resulting in a throughput $\hat{\lambda}_{\text{out}}$ that is lower than $\hat{\lambda}$. The stable region $S_A^{q,m=\infty}$ and the network throughput for $q \notin S_A^{q,m=\infty}$, with cutoff phase $m = \infty$, are presented in equations (21) and (23), respectively. These analyses are corroborated by simulation results depicted in Fig. 6.

Fig. 7 elucidates the influence of system parameters, such as the initial backoff window size W and the cutoff phase m, on the system efficiency in scenarios where $\hat{\lambda} > \hat{\lambda}_{\max}$. In this case, the network operates at the undesired stable point p_A and the stable region S_A^q is empty. Fig. 7(a) illustrates how the system efficiency varies with the backoff factor q under different values of W. Notably, for $W = 64 < 2n$, it can clearly observed that the maximum system efficiency is achieved at the optimal backoff factor q^* derived in (26). Conversely, for $W > 2n$, it can be observed from Fig. 7(a) that the maximum system efficiency $\hat{\lambda}_{e_{\max}}$ derived in Section IV cannot be achieved. Instead, the system efficiency monotonically increases with backoff factor q and is maximized at $q = 1$. Fig. 7(b) illustrates the effect of cutoff phase m on system efficiency. As shown in Fig. 7(b), the optimal backoff factor q^* significantly increases as the cutoff phase m increases. In the scenario where m approaches infinity, the fairness among all the STAs is compromised owing to the "capture effect" [36], elucidating the observed discrepancy between the simulation results and analytical predictions.

Fig. 8 further presents the system efficiency versus the number of STAs n with $q = 1/2$, which is adopted in current IEEE 802.11ax UORA mechanism. It can be seen from Fig. 8, with $m = 3$ and $W = 32$, when the number of STAs n increases to 120, the probability of collisions between STAs increases, leading to a significant reduction in system efficiency. When W is large, fewer STAs can successfully complete the backoff process, resulting in a serious underutilisation of RUs, and thus the system efficiency is much lower than the maximum value e^{-1}. Thus, the selection of system parameters is very important for system performance.

V. Analysis and Optimization of Access Delay

In this section, the derivation of access delays at p_L and p_A is presented, along with a discussion on the optimization of the average access delay at p_A through the appropriate selection



(a) (b)

Fig. 9. Mean access delay $E[D_{0,p=p_A}]$ (in unit of t) versus backoff factor q in saturated IEEE 802.11ax UORA networks. $\lambda = 1$, $m = \infty$. (a) Mean access delay $E[D_{0,p=p_A}]$ with various W and n. $M = 9$. (b) Mean access delay $E[D_{0,p=p_A}]$ with various the number of RUs M. $n = 200$, $W = 128$.

of the backoff factor q. The average access delay, denoted as $E[D_0]$, is defined as the expected number of time slots required for the successful transmission of HOL packet.

A. Mean Access Delay

For analytical convenience, the notation Y_i is introduced to represent the sojourn time of the HOL packet in state R_i, and D_i is used to denote the duration from state R_i to the completion of service, where i ranges from 0 to m. According to the Markov chain presented in Fig. 3, we have

$$D_i = \begin{cases} Y_i + t, & \text{with probability } p, \\ Y_i + t + D_{i+1}, & \text{with probability } 1-p\ , \end{cases} \tag{27}$$

for $i = 0, \cdots, m-1$ and

$$D_m = \begin{cases} Y_m + t, & \text{with probability } p, \\ Y_m + t + D_m, & \text{with probability } 1-p\ . \end{cases} \tag{28}$$

It is important to highlight that D_0 represents the service time of the HOL packet, which is equivalently referred to as the access delay. The probability generating function of D_0 is denoted by $G_{D_0}(z)$. It can be established that

$$G'_{D_0}(1) = \frac{1}{p}t + \sum_{i=0}^{m-1}(1-p)^i G'_{Y_i}(1) + \frac{(1-p)^m}{p} G'_{Y_m}(1), \tag{29}$$

where

$$G'_{Y_i}(1) = G'_{Y_m}(1) = \tau_{R_i}. \tag{30}$$

By combining (29) and (30), the average access delay $E[D_0]$ (in unit of t) can be written as

$$\begin{aligned} E[D_0] =& G'_{D_0}(1) \\ =& \frac{3}{2p} + \frac{W}{2M} \\ & \cdot \left(\frac{1}{1-\frac{1-p}{q}} + \left(\frac{1}{p} - \frac{1}{1-\frac{1-p}{q}} \right) \cdot \left(\frac{1-p}{q} \right)^m \right). \end{aligned} \tag{31}$$

According to (31), it can be clearly observed that the average access delay $E[D_0]$ is closely related to the backoff parameters, such as the backoff factor q, the initial backoff window size W and the cutoff phase m. Specifically, by combining (12) and (31), the mean access delay at the desired stable point P_L can be written as

$$\begin{aligned} E[D_{0,p=p_L}] =& \frac{3}{2e^{\mathbb{W}_0(\frac{-\hat{\lambda}}{M})}} + \frac{W}{2M} \cdot \frac{1}{1 - \frac{1-e^{\mathbb{W}_0(-\frac{\hat{\lambda}}{M})}}{q}} \\ & + \frac{W}{2M}\left(\frac{1}{e^{\mathbb{W}_0(\frac{-\hat{\lambda}}{M})}} - \frac{q}{e^{\mathbb{W}_0(-\frac{\hat{\lambda}}{M})} + q - 1} \right) \left(\frac{1-e^{\mathbb{W}_0(\frac{-\hat{\lambda}}{M})}}{q} \right)^m. \end{aligned} \tag{32}$$

It can be easily observed from (32) that when aggregate input rate $\hat{\lambda}$ becomes significantly low, specifically when $\hat{\lambda} \to 0$, the mean access delay tends towards $\lim_{\hat{\lambda}\to 0} E[D_{0,p=p_L}] = 3/2 + W/2M$. On the other hand, according to (16) and (31), the mean access delay at the undesired stable point P_A $E[D_{0,p=p_A}]$ can be written as

$$E[D_{0,p=p_A}] = \frac{n}{M} \cdot \frac{1}{-p_A \ln p_A}, \tag{33}$$

which is minimized when $p_A = e^{-1}$. Therefore, the minimum mean access delay for $p = p_A$ is given by $\min_{p_A} E[D_{0,p=p_A}] = ne/M$. The corresponding optimal backoff factor q^* can be further obtained by solving (24).

B. Simulation Results

In this subsection, simulation results are shown to verify the preceding analysis. The simulation setting is consistent with the system model. The mean access delay is calculated by dividing the total access delay of all successfully transmitted packets by the number of successfully transmitted packets.

Fig. 9 presents how the mean access delay at $p = p_A$ varies with the backoff parameters. As shown in the Fig. 9(a), the average access delay for $p = p_A$ $E[D_{0,p=p_A}]$ demonstrate a monotonic increase with an increase in number of STAs n

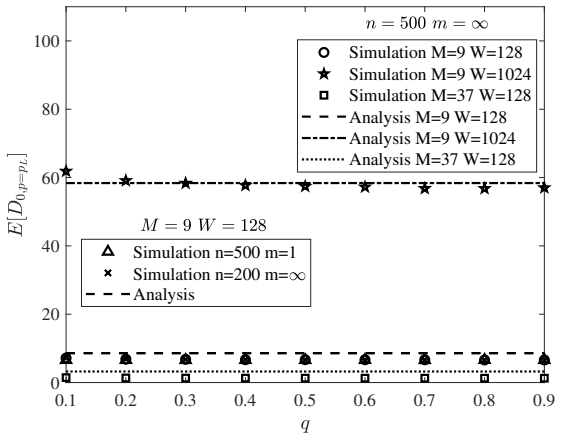


Fig. 10. Mean access delay $E[D_{0,p=p_L}]$ (in unit of t) versus backoff factor q in unsaturated IEEE 802.11ax UORA networks. $\hat{\lambda} = 0.1$.

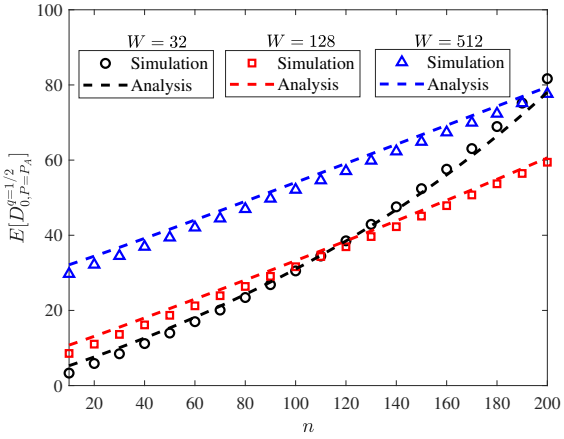


Fig. 11. Mean access delay $E[D^{q=1/2}_{0,p=p_A}]$ versus the number of STAs n in saturated IEEE 802.11ax UORA networks. $\lambda = 1$, $M = 9$, $m = 3$.

and the initial backoff window size W. In a dense network, the probability of collision between STAs increases. Also, the time slot required for the STAs to complete a backoff increases when W is large. Both of the above scenarios lead to an increase in the average access delay. From Fig. 9(b), with $m = \infty$, $n = 200$ and $W = 128$, the mean access delay decreases as the number of RUs M increases, which is consistent with (33). When the number of RUs M is large, the STAs may complete the backoff faster, resulting in a reduction in the delay required for the STA to successfully transmit. Furthermore, it can clearly observed that the minimum average access delay ne/M is achieved at the optimal backoff factor q^* derived in (24). As a result, the selection of backoff parameters is critical to the mean access delay when $p = p_A$.

Fig. 10 illustrates the influence of system parameters on the average access delay at $p = p_L$. As depicted in Fig. 10, with $\hat{\lambda} = 0.1$, the mean access delay at p_L $E[D_{0,p=p_L}]$ increases with the initial backoff window size W and decreases with the number of RUs M. Note the average access delay at p_L $E[D_{0,p=p_L}]$ is insensitive to the number of STAs n and the cutoff phase m, because the HOL packets hardly encounter any collisions if p_L is approach to one when $\hat{\lambda} \to 0$. The above phenomena align with the theoretical analyses in Section V.

Fig. 11 further presents the mean access delay versus the number of STAs n with $q = 1/2$. Similar to Fig. 8, we can see from Fig. 11 that the average access delay is also affected by the number of STAs n and the initial backoff window size W.

VI. Discussions

So far we have focused on the performance analysis of homogeneous IEEE 802.11ax UORA mechanism where all the STAs have the same backoff parameters. However, in real-world implementations of contention-based protocols, the heterogeneity in STAs' access parameters introduces disparities, resulting in serious unfairness among STAs. Consequently, the performance implications of IEEE 802.11ax UORA within heterogeneous cases remain largely unexplored. In addition, to further enhance throughput and reduce latency, it is necessary to combine the 802.11ax OFDMA UORA protocol with the MLO and MRU capabilities introduced by the next-generation Wi-Fi technologies [31], [32]. In this section, we will further discuss the impact of STAs' access parameter heterogeneity, MLO and MRU on the performance of saturated IEEE 802.11ax UORA networks with the BEB mechanism.

A. The Effect of STAs' Access Parameter Heterogeneity on IEEE 802.11ax UORA Mechanism

In this subsection, the impact of STAs' access parameter heterogeneity on network performance is explored by analyzing the effects of decrementing the packet's backoff counter by M upon receiving the TF. Specifically, the network is segmented into K groups, with each STA within a group k sharing identical backoff parameters, where $k = 1, 2, \cdots, K$. We denote the number of STAs in group k as n_k, with each STA in group k decrementing its backoff counter by M_k. These decrement values may vary from one group to another, although the actual number of RUs utilized for access requests remains constant at M. As outlined in Section III-B, for a HOL packet in group g to be successfully transmitted, the remaining $n_g - 1$ STAs in the same group and the STAs in the other $K - 1$ groups must be in state R_i, where $i = 0, \cdots, m$, and not transmitting any requests. The limiting probability of successful transmission of a HOL packet in such a heterogeneous network, employing the binary exponential backoff (BEB) strategy, is denoted as p_B^M and can be expressed as

$$p_B^M = \left\{1 - \sum_{i=0}^{m} \frac{\widetilde{\pi}_{R_i}^{M=M_g} r_i^{M=M_g}}{M}\right\}^{n_g - 1} \cdot \prod_{k=1, k\neq g}^{K} \left\{1 - \sum_{i=0}^{m} \frac{\widetilde{\pi}_{R_i}^{M=M_k} r_i^{M=M_k}}{M}\right\}^{n_k}. \quad (34)$$

With large values of number of STAs in each group, (34) can be further rewritten as

$$p_B^M \overset{W \geq M}{\approx} \exp\left\{\sum_{k=1}^{K}\left(\frac{-2 \cdot n_k/M}{\frac{W}{M_k}\left(\frac{p_B^M}{2p_B^M - 1} - \left(\frac{p_B^M}{2p_B^M - 1} - 1\right)(2(1 - p_B^M))^m\right)}\right)\right\}. \quad (35)$$

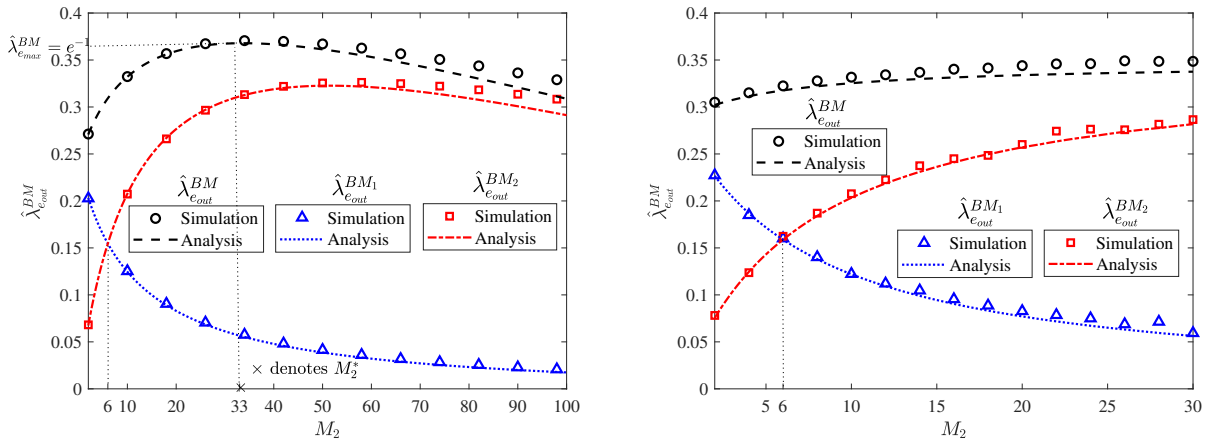


Fig. 12. Overall system efficiency $\hat{\lambda}^{BM}_{e_{\text{out}}}$ and system efficiency of each group $\hat{\lambda}^{BM_k}_{e_{\text{out}}}$ of a two-group saturated IEEE 802.11ax UORA networks under various M_2. $\lambda = 1$, $W = 128$, $M = 9$, $M_1 = 6$. (a) $n_1 = n_2 = 50$, $m = 3$. (b) $n_1 = n_2 = 100$, $m = \infty$.

According to (7), (19) and (35), the overall system efficiency in the heterogeneous case, which defined as the total system efficiency of K groups, can be obtained as

$$\hat{\lambda}^{BM}_{e_{\text{out}}} = \sum_{k=1}^{K} \frac{n_k \cdot \widetilde{\pi}_T^{M=M_k}}{M} = -p_B^M \ln p_B^M, \tag{36}$$

and the system efficiency of each group, $\hat{\lambda}^{BM_k}_{e_{\text{out}}}$, can be written as

$$\begin{aligned}\hat{\lambda}^{BM_k}_{e_{\text{out}}} &= \frac{n_k \cdot \widetilde{\pi}_T^{M=M_k}}{M} \\ &= \frac{2p_B^M \cdot n_k/M}{\frac{W}{M_k}\left(\frac{p_B^M}{2p_B^M-1} - \left(\frac{p_B^M}{2p_B^M-1}-1\right)\left(2(1-p_B^M)\right)^m\right)}.\end{aligned} \tag{37}$$

Specifically, we extend the previous analysis to a scenario involving two groups. According to (35) and (37), the system efficiency ratio between the two groups is given by

$$\frac{\hat{\lambda}^{BM_1}_{e_{\text{out}}}}{\hat{\lambda}^{BM_2}_{e_{\text{out}}}} = \frac{n_1 M_1}{n_2 M_2}. \tag{38}$$

According to (36), the maximum system efficiency, $\hat{\lambda}^{BM}_{e_{\max}}$, can attain e^{-1} when $p_B^M = e^{-1}$. By substituting $p^M_{B,K=2} = e^{-1}$ into (35), the decrement values of backoff counter of these two groups to achieve the maximum system efficiency can be further revealed. For instance, with the decrement value M_1 for group 1 held constant, the optimal decrement value for group 2 to achieve the maximum system efficiency $\hat{\lambda}^{BM}_{e_{\max}}$ is given by

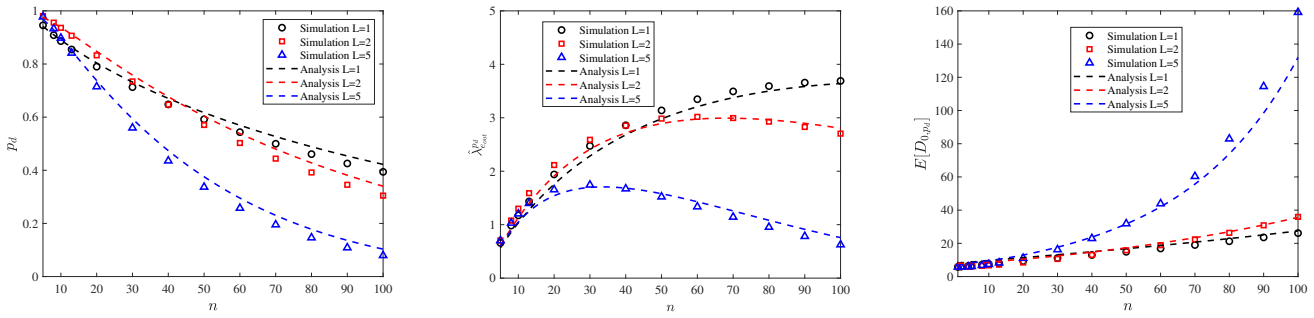
$$M_2^* = \left\lceil \frac{MW\left(\frac{e^{-1}}{2e^{-1}-1} - \left(\frac{e^{-1}}{e^{-1}-1}-1\right)\left(2(1-e^{-1})\right)^m\right) - 2n_1M_1}{2n_2} \right\rceil. \tag{39}$$

As observed from Fig. 12, the system efficiency is the same for both groups when $M_1 = M_2 = 6$. From Fig. 12, the system efficiency of Group 2 rises as the corresponding decrement values of backoff counter M_2 increases, while that of Group 1 decreases as M_2 increases. This trend is intuitive: a higher M_2 affords more transmission opportunities for the STAs in the second group n_2, thereby boosting the system efficiency of this group, denoted as $\hat{\lambda}^{BM_2}_{e_{\text{out}}}$. Consequently, the overall system efficiency $\hat{\lambda}^{BM}_{e_{\text{out}}}$ also increases. Conversely, the transmission requests from STAs in the first group, represented by $\hat{\lambda}^{BM_1}_{e_{\text{out}}}$, are reduced. Additionally, as shown in Fig. 12, the efficiency ratio between the two groups is $(n_1M_1)/(n_2M_2)$, which aligns with (38). Fig. 12(a) demonstrates that, with $n_1 = n_2 = 50$ and cutoff phase $m = 3$, the system efficiency reaches its peak when M_2 is set to $M_2^* = 33$, as indicated in (39). From Fig. 12(b), with $n_1 = n_2 = 100$ and $m = \infty$, the overall system efficiency is not sensitive to changes in M_2. These observations underscore the critical importance of selecting appropriate M_k values in a heterogeneous network.

B. The Impact of MLO and MRU on IEEE 802.11ax UORA Mechanism

In this subsection, we explore how MLO and MRU capabilities can enhance the performance of the IEEE 802.11ax UORA mechanism. Specifically, we consider a scenario where each STA equipped with multi-link capabilities can randomly select L RUs to transmit the same packet simultaneously upon receiving the TF, and the HOL packet is considered to be transmitted successfully as long as one of the RUs does not have a collision. This operation is referred to as packet duplication (PD) mode in MLO [31], [32]. For simplicity, we term this enhanced approach as the UORA-based multi-opportunity randomized access (UORA-based MORA) mechanism.

According to PD mode mechanism, successful transmission of a given HOL packet requires that the other $n-1$ STAs are either in state R_i and not transmitting any request, or in state C_i, with collisions occurring only in j RUs, where $j \leq L$. The limiting probability of successful transmission in the UORA-



(a) (b) (c)

Fig. 13. Limiting probablitity of successful transmission p_d, system efficiency $\hat{\lambda}_{e_{\text{out}}}^{p_d}$ and mean access delay $E[D_{0,p_d}]$ versus the number of STAs n in UORA-BASED MORA mechanism. $\lambda = 1$, $M = 10$, $m = 1$, $W = 128$, $q = 1/2$. (a) Limiting probability p_d with various L. (b) System efficiency $\hat{\lambda}_{e_{\text{out}}}^{p_d}$ with various L. (c) Mean access delay $E[D_{0,p_d}]$ with various L.

BASED MORA mechanism, p_d, is given by:

$$p_d = \sum_{j=1}^{L} (-1)^{j+1} \binom{L}{j} \cdot \left[(1-p_t) + p_t \left(\frac{\binom{M-j}{L}}{\binom{M}{L}} \right) \right]^{n-1}, \quad (40)$$

where $p_t = \sum_{i=0}^{m} \widetilde{\pi}_{R_i}^{p=p_d} \cdot r_i$. When $L = 1$, (40) is consistent with (15).

By combining (7), (31) and (40), the corresponding system efficiency $\hat{\lambda}_{e_{\text{out}}}^{p_d}$ and mean access delay $E[D_{0,p_d}]$ can be written as

$$\hat{\lambda}_{e_{\text{out}}}^{p_d} = \frac{n\widetilde{\pi}_T^{p=p_d}}{M} = \frac{np_d}{\frac{W}{2M}\left(\frac{p_d q}{p_d+q-1} - \left(\frac{p_d q}{p_d+q-1} - 1\right)\left(\frac{1-p_d}{q}\right)^m\right) + \frac{3}{2}}, \quad (41)$$

and

$$E[D_{0,p_d}] = \frac{3}{2p_d} + \frac{W}{2M}\left(\frac{1}{1-\frac{1-p_d}{q}} + \left(\frac{1}{p_d} - \frac{1}{1-\frac{1-p_d}{q}}\right)\left(\frac{1-p_d}{q}\right)^m\right), \quad (42)$$

respectively.

Fig. 13 illustrates the performance of the UORA-based MORA mechanism versus the number of STAs n, with the number of RUs fixed at $M = 10$. The simulation results, as depicted in Fig. 13, corroborate the theoretical analysis outlined in (40), (41), and (42). It is evident from the figure that when the number of STAs n is below 40, employing $L = 2$ yields better network performance than $L = 1$. However, as n increases, the performance advantage of $L = 2$ diminishes. Intuitively, with fewer STAs, the lower collision probability and underutilization of RUs contribute to enhanced performance for $L = 2$, increasing both the successful transmission probability p_d and the system efficiency $\hat{\lambda}_{e_{\text{out}}}^{p_d}$. Furthermore, this configuration also demonstrates reduced mean access delays compared to $L = 1$. It is important to note, however, that setting L to excessively high values, such as $L = 5$, significantly increases collision probabilities, leading to marked performance degradation. As n grows, the UORA-based MORA mechanism with $L \neq 1$ tends to aggravate conflicts among STAs, resulting in lower overall system performance than the standard IEEE 802.11ax UORA mechanism with $L = 1$.

VII. Conclusion

This paper presents a comprehensive mathematical model to analyze the impact of backoff parameters on the performance of IEEE 802.11ax UORA mechanism, elucidating the critical relationship between steady-state points and system parameters. The analysis clarifies how to optimize the network performance, including system efficiency and mean access delay by properly selecting the backoff factor q when the network operates at different steady-state points. The effects of access parameter heterogeneity, the capabilities of MLO and MRU are further discussed, which provide practical insights for advancing next-generation Wi-Fi network design.

Appendix A
Derivation of (9)

We can see from Fig. 3 that when the backoff counter selected by the HOL packet is greater than M, the HOL packet will be in state R_i. As long as the HOL packet receives the TF, it will decrement by M until the backoff counter decrements to T ($T \leq M$), and the HOL packet will leave the state R_i and request transmission. As a result, accoording to the Markov chain presented in Fig. 3, we have

$$\begin{cases} f_{B_k} = \left(\left\lfloor \frac{W_i-1-k}{M} \right\rfloor + 1\right) \frac{1}{W_i} f_{B_T}, \\ f_{B_T} + \sum_{k=M+1}^{W_i-1} f_{B_k} = 1, \end{cases} \quad (43)$$

where f_{B_k} denotes the conditional probability of the HOL packet can request transmission when the selected backoff counter is k ($k \in [0, W_i - 1]$).

To simplify presentation, let

$$\begin{aligned} X_i &= \sum_{k=M+1}^{W_i-1} \left(\left\lfloor \frac{W_i - 1 - k}{M} \right\rfloor + 1 \right) \\ &= -\frac{M}{2} \left\lfloor \frac{W_i - 1}{M} \right\rfloor^2 + \left(W_i - 1 - \frac{M}{2} \right) \left\lfloor \frac{W_i - 1}{M} \right\rfloor, \end{aligned} \quad (44)$$

then, (43) can be rewritten as

$$\begin{cases} f_{B_k} = \left(\left\lfloor \frac{W_i-1-k}{M} \right\rfloor + 1\right) \frac{1}{W_i} f_{B_T}, \\ f_{B_T} + \frac{X_i}{W_i} f_{B_T} = 1. \end{cases} \quad (45)$$

By substituing (44) into (45), the conditional probability r_i (i.e., f_{B_T}) that a HOL packet makes a transmission request at the given its backoff counter less than or equal to M (i.e., the selected backoff counter is T) can be obtained as

$$\begin{aligned} r_i &= \frac{W_i}{W_i + X_i} \\ &= \frac{W_i}{W_i - \frac{M}{2}\left\lfloor \frac{W_i-1}{M} \right\rfloor^2 + \left(W_i - 1 - \frac{M}{2}\right)\left\lfloor \frac{W_i-1}{M} \right\rfloor}. \end{aligned} \quad (46)$$

APPENDIX B
DERIVATION OF (11)

The mean holding time τ_{R_i} in (5) can be rewritten as

$$\begin{aligned} \tau_{R_i} &= \sum_{k_0=0}^{W_i-1} \left\lceil \frac{k_0}{M} \right\rceil \frac{t}{W_i} \quad i \in [0, m] \\ &= \frac{t}{W_i} \cdot \frac{\left\lceil \frac{W_i-1}{M} \right\rceil \cdot \left(\left\lceil \frac{W_i-1}{M} \right\rceil - 1\right)}{2} \cdot M \\ &\quad + \frac{t}{W_i} \left\lceil \frac{W_i-1}{m} \right\rceil \cdot \left(W_i - 1 - \left(\left\lceil \frac{W_i-1}{M} \right\rceil - 1 \right) \cdot M \right) \\ &= \frac{t}{W_i} \cdot Y_i, \end{aligned} \quad (47)$$

where

$$Y_i = \frac{-M}{2} \left\lceil \frac{W_i-1}{M} \right\rceil^2 + \left(W_i - 1 + \frac{M}{2} \right) \left\lceil \frac{W_i-1}{M} \right\rceil + 1. \quad (48)$$

Because whether W_i is greater than M depends not only on W and m, but also on the value of q, it is difficult to derive the exact root of the mean holding time τ_{R_i}. Therefore, this paper only considers the case of $W_0 > M$. When the initial backoff window size W is large and the number of RUs M is small. Y_i and (44) can be rewritten as

$$Y_i \approx \frac{W_i^2}{2M} \cdot + \frac{W_i}{2}, X_i \approx \frac{W_i^2}{2M} - \frac{W_i}{2}. \quad (49)$$

By combining (46) and (47), we can get

$$\begin{aligned} \tau_{R_i} \cdot r_i &= \frac{t}{W_i} \cdot Y_i \cdot \frac{W_i}{W_i + X_i} \\ &= t \cdot \frac{Y_i}{W_i + X_i} \\ &\approx t, \end{aligned} \quad (50)$$

with which (11) can further be obtained.

REFERENCES

[1] J. Kurose and K. Ross, *Computer networks: A top down approach featuring the internet*. Pearson Addison Wesley, 2010.

[2] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[3] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas in Commun.*, vol. 18, no. 3, pp. 535–547, 2000.

[4] H. A. Omar *et al.*, "A survey on high efficiency wireless local area networks: Next generation wifi," *IEEE Commun. Surveys Tut.*, vol. 18, no. 4, pp. 2315–2344, 2016.

[5] D.-J. Deng, S.-Y. Lien, J. Lee, and K.-C. Chen, "On quality-of-service provisioning in IEEE 802.11ax WLANs," *IEEE Access*, vol. 4, pp. 6086–6104, 2016.

[6] *IEEE Approved Draft Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Enhancements for High Efficiency WLAN*, IEEE Std. P802.11ax/D8.0, Oct. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9353436

[7] E. Mozaffariahrar, F. Theoleyre, and M. Menth, "A survey of Wi-Fi 6: Technologies, advances, and challenges," *Future Internet*, vol. 14, no. 10, p. 293, 2022.

[8] B. Bellalta and K. Kosek-Szott, "AP-initiated multi-user transmissions in IEEE 802.11ax WLANs," *Ad Hoc Netw.*, vol. 85, pp. 145–159, 2019.

[9] K. Kosek-Szott, S. Szott, and F. Dressler, "Improving IEEE 802.11 ax UORA performance: Comparison of reinforcement learning and heuristic approaches," *IEEE Access*, vol. 10, pp. 120285–120295, 2022.

[10] H. Yang, D.-J. Deng, and K.-C. Chen, "Performance analysis of IEEE 802.11 ax UL OFDMA-based random access mechanism," in *Proc. IEEE GLOBECOM*, 2017.

[11] G. Naik, S. Bhattarai, and J.-M. Park, "Performance analysis of uplink multi-user OFDMA in IEEE 802.11 ax," in *Proc. IEEE ICC*, 2018.

[12] D. Xie, J. Zhang, A. Tang, and X. Wang, "Multi-dimensional busy-tone arbitration for OFDMA random access in IEEE 802.11 ax," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4080–4094, 2020.

[13] D. Magrin, S. Avallone, S. Roy, and M. Zorzi, "Performance evaluation of 802.11ax OFDMA through theoretical analysis and simulations," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5070–5083, 2023.

[14] T. Uwai *et al.*, "Performance evaluation of OFDMA random access in IEEE802. 11ax," in *Proc. IEEE ISPACS*, 2016.

[15] L. Lanante, H. O. T. Uwai, Y. Nagao, M. Kurosaki, and C. Ghosh, "Performance analysis of the 802.11 ax UL OFDMA random access protocol in dense networks," in *Proc. IEEE ICC*, 2017.

[16] K.-H. Lee, "Performance analysis of the IEEE 802.11 ax MAC protocol for heterogeneous Wi-Fi networks in non-saturated conditions," *Sensors*, vol. 19, no. 7, p. 1540, 2019.

[17] R.-G. Cheng, C.-M. Yang, B. S. Firmansyah, and R. Harwahyu, "Uplink OFDMA-based random access mechanism with bursty arrivals for IEEE 802.11 ax systems," *IEEE Netw. Lett.*, vol. 4, no. 1, pp. 34–38, 2021.

[18] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 protocol: Design and performance evaluation of an adaptive backoff mechanism," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, pp. 1774–1786, 2000.

[19] A. L. Toledo, T. Vercauteren, and X. Wang, "Adaptive optimization of IEEE 802.11 DCF based on Bayesian estimation of the number of competing terminals," *IEEE Trans. Mobile Comput.*, vol. 5, no. 9, pp. 1283–1296, 2006.

[20] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Netw.*, vol. 8, no. 6, pp. 785–799, 2000.

[21] L. Dai and X. Sun, "A unified analysis of IEEE 802.11 DCF networks: Stability, throughput, and delay," *IEEE Trans. Mobile Comput.*, vol. 12, no. 8, pp. 1558–1572, 2012.

[22] M. Karaca, S. Bastani, and B. Landfeldt, "Modifying backoff freezing mechanism to optimize dense IEEE 802.11 networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9470–9482, 2017.

[23] Y.-W. Kuo and T.-L. Tsai, "Design and evaluation of a contention-based high throughput MAC with delay guarantee for infrastructured IEEE 802.11 WLANs," *J. Commun. Netw.*, vol. 15, no. 6, pp. 606–613, 2013.

[24] N. Shahin, R. Ali, S. W. Kim, and Y.-T. Kim, "Cognitive backoff mechanism for IEEE802.11ax high-efficiency WLANs," *J. Commun. Netw.*, vol. 21, no. 2, pp. 158–167, 2019.

[25] Y. Kim, L. Kwon, and E.-C. Park, "OFDMA backoff control scheme for improving channel efficiency in the dynamic network environment of IEEE 802.11 ax WLANs," *Sensors*, vol. 21, no. 15, p. 5111, 2021.

[26] K. Kosek-Szott and K. Domino, "An efficient backoff procedure for IEEE 802.11 ax uplink OFDMA-based random access," *IEEE Access*, vol. 10, pp. 8855–8863, 2022.

[27] A. Baiocchi, D. Garlisi, A. L. Valvo, G. Santaromita, and I. Tinnirello, "'Good to repeat': Making random access near-optimal with repeated contentions," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 712–726, 2019.

[28] J. Mvulla, E.-C. Park, M. Adnan, and J.-H. Son, "Analysis of asymmetric hidden node problem in IEEE 802.11 ax heterogeneous WLANs," in *Proc. ICTC*, 2015, pp. 539–544.

[29] J. Mvulla and E.-C. Park, "Enhanced dual carrier sensing with transmission time control for fair spatial reuse in heterogeneous and dense WLANs," *IEEE Access*, vol. 6, pp. 22140–22155, 2018.
[30] M. Z. Ali, J. Mišić, and V. B. Mišić, "Impact of hidden nodes on uplink transmission in IEEE 802.11 ax heterogeneous network," in *Proc. IWCMC*, 2018.
[31] *IEEE Draft Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment: Enhancements for Extremely High Throughput (EHT)*, IEEE Std. P802.11be/D5.0, Jan. 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10381585
[32] Q. Zhang, P. Schulz, and G. Fettweis, "Optimizing real-time responsiveness in IIoT: A dynamic approach for WiFi OFDMA uplink transmissions," in *Proc. IEEE VTC2023-Fall*, 2023.
[33] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11 ax high efficiency WLANs," *IEEE Commun. Surveys Tut.*, vol. 21, no. 1, pp. 197–216, 2018.
[34] T. Uwai *et al.*, "Adaptive backoff mechanism for OFDMA random access with finite service period in IEEE802. 11ax," in *Proc. IEEE CSCN*, 2016.
[35] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, pp. 329–359, 1996.
[36] K. K. Ramakrishnan and H. Yang, "The Ethernet capture effect: Analysis and solution," in *Proc. IEEE LCM*, 1994.

Pengxue Liu received the B.Eng. degree in Applied Electronic Technology Education from Hubei University of Education, Wuhan, China, in 2017, where she is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China. Her research interests include modeling and optimization of next-generation WiFi networks, and reinforcement learning for networking.

Yitong Li received the B.Eng. and Ph.D. degrees in Electrical Engineering from the City University of Hong Kong, Hong Kong, in 2011 and 2016, respectively. He is currently an Associate Professor with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China. His research interests include the performance evaluation and optimization of wireless random access networks.

Dalong Zhang is currently a Professor with the School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, China. He was a Visiting Senior Researcher with the Chinese University of Hong Kong, Hong Kong, China, from 2018 to 2019. As a Coordinator, Project Manager and Principal Investigator, he has been involved in the National Natural Science Foundation of China, National Science and Technology Major Project of the Ministry of Science and Technology of China and other research projects. His research interests include network resilience assessment, distributed network security, and wireless communication security.