IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 September 2023, accepted 20 October 2023, date of publication 23 October 2023, date of current version 1 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327154

RESEARCH ARTICLE

Performance Analysis of Aggregation-Enabled IEEE 802.11 WLANs With Variable Aggregation Size

SHINNAZAR SEYTNAZAROV[1], (Member, IEEE), DONG GEUN JEONG[2], (Senior Member, IEEE), AND WHA SOOK JEON[3], (Senior Member, IEEE)

[1]Faculty of Computer Science and Engineering, Innopolis University, 420500 Innopolis, Russia
[2]Department of Electronics Engineering and Applied Communication Research Center, Hankuk University of Foreign Studies, Yongin 17035, South Korea
[3]Department of Computer Science and Engineering, Seoul National University, Seoul 08826, South Korea

Corresponding author: Wha Sook Jeon (wsjeon@snu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) through the Korean Government under Grant 2021R1A2C1005552 and Grant 2022R1A2C1004928.

ABSTRACT Aggregation-enabled wireless local area network (WLAN) technologies such as IEEE 802.11n and 802.11ac are deployed densely for wireless access to the high-speed Internet. Conventional approaches for performance modeling of 802.11n/ac assume fully saturated traffic, where stations always have enough packets to transmit aggregate frames of full or fixed aggregation size. However, such an assumption is unrealistic, since stations usually transmit aggregate frames with different aggregation sizes depending on several dynamic factors, such as offered traffic load, number of active stations, transmission rate, and random backoff time. In this work, we propose a new performance model of aggregation-enabled 802.11n/ac WLANs for different traffic loads and buffer sizes with the assumption that stations always have at least one packet to transmit. Unlike conventional models where stations always contend to transmit aggregate frames of maximum or other fixed aggregation size, the proposed model dynamically determines the aggregation size based on given offered traffic load, number of stations, transmission rate, and random backoff process. Performance evaluations showed that the proposed model produces remarkably accurate performance estimates for throughput, especially for higher traffic loads.

INDEX TERMS 802.11, 802.11n, 802.11ac, 802.11ax, A-MPDU, frame aggregation, Markov chain, performance modeling, WLAN.

I. INTRODUCTION

The IEEE 802.11 wireless local area network (WLAN) technology became a de facto standard for ubiquitous wireless Internet access in unlicensed frequency bands. To fulfill the persistent demand for high-speed and mobile Internet access, engineers have continuously improved IEEE 802.11 technology. Orthogonal frequency division multiplexing (OFDM), multiple-input/multiple-output (MIMO), and channel bonding mechanisms greatly improved data rate in the physical (PHY) layer of IEEE 802.11n and 802.11ac, while the packet aggregation mechanism was introduced to efficiently utilize the increased PHY data rates in the medium access control (MAC) layer. Due to its high efficiency, packet aggregation was also adopted in IEEE 802.11ax standard as well [1], [2], [3].

The packet aggregation allows an 802.11n/ac station to transmit up to 64 MAC protocol data units (MPDUs) in one large frame called aggregate MPDU (A-MPDU). The receiver station replies with a block acknowledgment (block-ACK) frame indicating a success/failure status of each MPDU in A-MPDU.

The associate editor coordinating the review of this manuscript and approving it for publication was Yafei Hou.

In the literature, most conventional performance models for aggregation-enabled 802.11n/ac WLANs assume fully saturated traffic loads, that is, each station always has an A-MPDU of maximum or other fixed aggregation size. For example, the models in [4], [5], [6], [7], [8], and [9] assume that all A-MPDUs have a maximum aggregation size, that is, 64 MPDUs. However, such an assumption rarely applies to real cases, since stations usually transmit A-MPDUs of different aggregation sizes depending on various dynamic factors (for example, offered load, random backoff time, etc.). Furthermore, using these models, it is impossible to produce basic curves of interest such as "*Throughput versus Offered load*".

In this work, our aim is to improve the current state of the existing modeling approaches for aggregation-enabled WLANs in [4], [5], [6], [7], [8], [9], [10], [11], [12], and [13] to model the performance more accurately for any offered load. To do so, we propose a new performance model of aggregation-enabled WLANs, assuming that stations always have at least one packet in their queues. Due to such an assumption, all stations continuously compete for channel access, making it easy to derive transmission and conditional collision probabilities. Since this assumption generates an additional packet whenever the station has no packet to send, the model produces a higher throughput than the real throughput, especially at small offered loads. But as offered load increases, the amount of such additionally generated packets decreases and the produced throughput becomes much similar to the real one. Most importantly, unlike the existing modeling approach, the aggregation size of A-MPDUs in our model is dynamically determined by several factors, such as the offered traffic load, the number of stations, the transmission rate, and the random backoff process. Moreover, it is possible to produce "*Throughput versus Offered load*" or "*Average aggregation size versus Offered load*" curves using the proposed model. The performance evaluations showed that the proposed model improves the accuracy of the performance modeling of aggregation-enabled WLANs compared to the existing modeling approach. The main contributions of this work are as follows.

- First, we thoroughly analyze the literature related to performance modeling of aggregation-enabled 802.11n/ac WLANs and legacy 802.11a/b/g WLANs.
- Then, we identify the major weaknesses of those models and group them according to their common properties. We also explain why there is a need for a more accurate and simpler model that can better reflect the operations of aggregation-enabled WLANs.
- Next, we propose a new three-dimensional Markov chain model based on simple assumptions where we derive steady-state probabilities, throughput, and aggregation size distribution. We also explain how to derive the throughput for the existing modeling approach.
- Finally, we evaluate the performance of the model in comparison to the simulation results and the existing modeling approach.

The remainder of this paper is organized as follows. Section II briefly reviews how the data is transmitted in aggregation-enabled WLANs, the A-MPDU and blockACK frames, the limiting factors of aggregation size, the existing performance models in the literature and describes the motivation of this work. Section III introduces the proposed performance analysis model. The aggregation size and throughput performance of the proposed model is evaluated and discussed in Section IV. Finally, Section V concludes this paper.

II. BACKGROUND

A. DISTRIBUTED COORDINATION FUNCTION

Data transfer in IEEE 802.11 WLANs, including IEEE 802.11n/ac, is handled by a distributed coordination function (DCF) protocol. Every station with a packet to send conducts carrier sensing to know if the medium is idle or busy. If the medium is sensed idle for a period called the DCF inter-frame space (DIFS), the station waits for additional random backoff counter slots before starting the transmission. The backoff counter is randomly selected from the current contention window size W_i, i.e., from $[0, W_i)$ interval, where $i \in [0, r]$ is the current backoff stage. The basic time unit is called a slot and has a duration of σ. For every idle slot, the station decrements its backoff counter by one. When the medium is detected to be busy, the backoff counter is frozen, and it continues to decrease only after the medium becomes idle for a DIFS period. When the backoff counter reaches 0, the station transmits the data frame [3].

The data frame can be an A-MPDU that includes multiple packets or an MPDU that includes a single packet. If the transmission is successful, following the short interframe space (SIFS) period, the receiver replies with an ACK frame when a single MPDU is received or blockACK frame when A-MPDU is received. When the sender receives the ACK/blockACK frame, it resets the backoff stage to 0, i.e., resets the contention window size to W_0 with a standard value of 16. If the sender does not receive ACK / blockACK within a timeout interval, it increments the backoff stage, i.e., doubles the contention window size as $W_{i+1} = \min(2W_i, W_{max})$, and starts a new backoff process. In other words, the station randomly chooses a new backoff counter value from $[0, W_{i+1})$, starts decrementing the backoff counter upon idle slots as explained previously, and finally retransmits the collided data frame when the counter expires. Here, W_{max} is the maximum contention window size. If the data frame successively collides and the sender does not receive an ACK/blockACK even at the last backoff stage (r), the sender drops/discards the data frame, resets the backoff stage to 0, and initiates another backoff process if there are other data to send [3].

There are two types of medium access schemes in 802.11 WLANs. The first one is called a basic access scheme, which was explained in the previous two paragraphs, where the sender immediately transmits a data frame upon expiration of the backoff counter, and the receiver replies with an

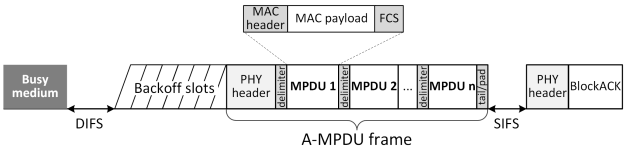


FIGURE 1. A-MPDU and blockACK frames.

ACK/blockACK if the transmission was successful. The second is the request-to-send (RTS)/clear-to-send (CTS) access scheme, where the sender sends a short frame called RTS upon the backoff timer expiration; if the RTS is successful, the receiver replies with another short frame called CTS after a SIFS duration. If the sender successfully receives the CTS, it finally transmits the data frame after another SIFS duration; if the sender does not receive the CTS within the CTS timeout interval, the sender increments the backoff stage and retries the transmission [3].

B. A-MPDU AGGREGATION IN IEEE 802.11n/ac

One of the major improvements throughout the evolution of IEEE 802.11 technology was A-MPDU aggregation and blockACK mechanisms. The IEEE 802.11n standard introduced A-MPDU aggregation, where a sender is allowed to send up to 64 MPDUs in one large frame called A-MDPU [1]. As shown in Figure 1, the sender, once its backoff timer expires, composes an A-MPDU from pending MPDUs in its queue and starts transmitting. The MPDUs are guarded with delimiters, and each MPDU has its MAC header with a sequence number (SN) and a frame checksum (FCS) fields.

Upon receiving the A-MPDU, the receiver can determine if each MPDU is received correctly using its FCS field. If at least one of the MPDUs was received correctly, the receiver replies with a blockACK frame indicating the SNs of the MPDUs that were successfully received. After receiving the blockACK, the sender retransmits the failed MPDUs (if there are any) together with new MPDUs (if there are any) in another A-MPDU upon next access to the channel [3].

C. DETERMINING FACTORS FOR THE AGGREGATION SIZE

As it is clear from the discussion of the previous two subsections, the station composes an A-MPDU and transmits it after the backoff timer expires. Thus, under unsaturated traffic conditions, the aggregation size of each A-MPDU depends on the number of pending MPDUs in the queue (i.e., the current queue size), which changes depending on current offered load at the station, the number of competing stations in the network, the PHY transmission rate, and a random backoff time.

Let us now address the other factors that play a role in determining the aggregation size. Apart from a maximum aggregation size of 64, the 802.11 standard puts the following limitations while composing an A-MPDU [1], [2], [3]:

- The maximum size for an A-MPDU is limited to 64 Kbytes in 802.11n and 1024 Kbytes in 802.11ac.
- The maximum duration for an A-MPDU is limited to 10 ms in 802.11n and 5.484 ms in 802.11ac.

The above two standard limitations restrict the aggregation size further. For example, suppose that all arriving MPDUs have sizes of about 1.2 Kbytes. Then the 802.11n station cannot include 64 such MPDUs in one A-MPDU, since it would violate the 64 Kbytes maximum A-MPDU size limit.

In addition to the limitations put by standards, there are also real implementations that introduce their limitations. For example, *ath9k* driver for Atheros 802.11n cards, limits the aggregation size to 32 packets and the A-MPDU duration to 4 ms [14]. Now consider another case, where the currently optimal PHY transmission rate at the 802.11n station is only 6.5 Mbps. Then the maximum A-MPDU size is limited to 6.5 Mbps $\times$ 4.0 ms $\approx$ 25.4 Kbytes instead of the 64 Kbytes limit put by the standard.

In summary, in practice, the aggregation size of A-MPDUs varies dynamically depending on not only offered traffic load, random backoff process, number of stations, and PHY transmission rate but also standard and implementation-specific limitations and packet size.

D. EXISTING PERFORMANCE MODELS OF AGGREGATION-ENABLED WLANs IN LITERATURE

There are many works aimed at developing the performance model of aggregation-enabled WLANs. Almost all of them extend the seminal legacy 802.11a/b/g DCF model proposed by Bianchi [15]. In the Bianchi model, stochastic processes of backoff stage and backoff counter evolution compose a two-dimensional (2)-D) Markov chain with the assumptions of constant collision probability, saturated traffic, and an ideal channel environment. Moreover, Bianchi assumed that the stations decrement their backoff counters at the beginning of each slot time. However, as Tinnirello et al. stated in [16], a closer look at the standard specification revealed that the stations decrement their backoff counters only at the end of idle slots. Due to such a backoff counter decrement rule, the first backoff slot immediately following the successful transmission plus idle DIFS can be accessed only by the transmitting station if it chooses 0 as the next backoff counter value. In other words, only the station which has just finished successful transmission can access the first backoff slot with a probability of $1/W_0$. Besides, the first backoff slot followed by collision plus an idle extended IFS (EIFS) duration is not accessible by any of the stations. Tinnirello et al. called these two types of slots *anomalous slots* and extended the Bianchi model to reflect their impact on performance. We ask readers to refer to [16] for more information on anomalous slots.

In the literature, several works model the performance of legacy IEEE 802.11a/b/g WLANs with unsaturated traffic considerations. Most of them extend the Bianchi's model. For example, in [17], [18], and [19], the authors extended Bianchi's model to account for a particular case of unsaturated traffic assuming a *bufferless station*, i.e., a station can have at most one packet at a time. But in practice,

802.11 stations usually have buffers for at least a few dozen packets. Few other works extended Bianchi's model to address the unsaturated traffic case with finite buffer [20], [21]. To achieve it, the authors introduced a third dimension in the Markov chain to track the evolution of queue size. However, the complexity of the model increased substantially because new packet arrivals are accounted for at the end of each slot, i.e., the queue size can change after each slot, resulting in an enormous number of transitions between the states of the different queue size. Such modeling approach makes it difficult to formulate the steady-state probabilities in terms of concise equations.

Let us now discuss the existing performance modeling approaches of aggregation-enabled WLANs. In [4], Lin and Wong extended the Bianchi's model to analyze the throughput for error-prone channels with an assumption of fully saturated traffic where all stations always have enough packets to send A-MPDUs of maximum or other fixed aggregation size. In [5], Li et al. proposed a new aggregation scheme similar to A-MPDU and extended the Bianchi's model to analyze the performance of the scheme with a fully saturated traffic assumption where the stations always transmit full-sized aggregate frames.

In [6], Hajlaoui et al. proposed an algorithm that adapts the MPDU size for noisy channel and adjusted Bianchi's model to analyze the performance for a saturated traffic condition. Seytnazarov et al. [7] proposed a scheme to transmit multiple 20MHz PHY PDUs (PPDUs) in parallel instead of a single wideband PDU and analyzed the performance of the scheme using Bianchi's model. In [8], the authors extended the Tinnirello's model [16] to include frame aggregation.

In [10] and [11], the authors proposed new automatic repeat request (ARQ) schemes and analyzed their throughput performance with a fully saturated traffic assumption. In [12], Seytnazarov et al. proposed a new model that reflects the impact of the standard ARQ protocol on the size of the aggregation and the throughput under erroneous channel and saturated traffic conditions. In [9] and [13], the authors proposed a model similar to the one in [12], but including the probability of packet drop after an exhausted retry limit.

All the above modeling approaches for aggregation-enabled WLANs consider a fully saturated traffic condition where every station has enough packets to compose an A-MPDU of maximum aggregation size [4], [5], [6], [7] or the aggregation size that the considered ARQ scheme currently allows [8], [9], [10], [11], [12], [13]. However, some performance modeling attempts include unsaturated traffic conditions. For example, Kuppa and Dattatreya [22] analyzed performance using embedded Markov chains assuming a predefined service time distribution. Similarly, in [23] and [24], the authors proposed a performance model for stations with unsaturated traffic and finite buffers but with the assumption that arrival times and service times have an exponential distribution. Assuming an exponentially distributed arrival times is usual practice in existing works because it can be repeated in real test-bed experiments as well, whereas assuming a predefined service time distribution is unfavorable. In reality, the service time is random and is determined by many factors, such as the number of active stations, transmission rate, random backoff process, packet size, and offered load. In another work [25], Kim et al. proposed an extension of the Markov chain in [17] and [18], where the transmission is triggered only when K packets arrive, and the arrivals during the service of an A-MPDU are ignored.

E. MOTIVATION

As discussed in Subsections II-C and II-D, the aggregation size can change for each A-MPDU depending on multiple factors, but the existing modeling approach considers fully saturated traffic conditions where the aggregation size is always fixed or controlled solely by ARQ [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. Since MPDU aggregation is a key element of modern WLAN technologies, correctly modeling its performance, including variable aggregation size, is still an open problem. So, in this work, we propose a new model that predicts the performance of aggregation-enabled WLANs more accurately for any given offered load and where the size of aggregation of an A-MPDU is dynamically determined by several factors, such as the offered traffic load, the number of stations, the transmission rate, and the random backoff process.

III. PROPOSED PERFORMANCE ANALYSIS MODEL

A. ASSUMPTIONS

We assume that the channel is ideal, the stations can hear each other perfectly, and all collided frames are undecodable at the listening stations upon collision. To further simplify the problem, we make the following assumptions:

1) A station always has at least one packet to send.
2) The size of an A-MPDU does not change after its service is started.

The rationale behind the Assumption 1 is to simplify the model by keeping all stations continuously competing for channel access. This makes it easy to derive the *transmission* and *conditional collision* probabilities. One can expect that the throughput produced by our model under Assumption 1 is higher than the true throughput for the given offered load. This is partially true, especially for small traffic loads. The reason is that Assumption 1 creates an extra packet whenever the station has no packet to send, therefore, the model displays greater throughput than the actual throughput, particularly at low offered loads. However, when the offered load grows, the probability of an empty queue decreases, consequently less extra packets are generated by Assumption 1, and therefore the produced throughput approaches the true throughput. Most importantly, unlike the previous modeling approach [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] where the aggregation size is always fixed, the aggregation size of A-MPDUs in our model is dynamically determined by multiple criteria, including offered traffic load, station count, transmission rate, and random backoff process. Furthermore, the proposed model can be used to generate "*Throughput versus Offered*"

load, "*Average aggregation size versus Offered load*", and other curves of interest.

Let us now address Assumption 2. The rationale behind this assumption is to simplify the model further, by decreasing the number of transitions between the states of different queue size. This will be more clear when we introduce the proposed three-dimensional Markov chain in the next subsection. Now, recall that, due to Assumption 1, a station always has a packet to send. In other words, after the service of the previous A-MPDU is finished, a station always has at least one packet in its queue. At this time moment, the station immediately composes an A-MPDU using currently available packets in its queue (even if there is only one) and invokes the backoff procedure for transmission, i.e., the service of the A-MPDU is started. During the service, new packets can arrive in the queue, but they are not included in the existing in-service A-MPDU. For example, a station finished a successful transmission (i.e., received a blockACK for the previous A-MPDU transmission), and after an idle DIFS period, at a time moment t_1, it immediately formed a new A-MPDU with the pending packets in queue (say two) and started the service of it. The A-MPDU collided twice and was eventually successfully delivered in the third attempt, and the station received blockACK at t_2. From t_1 till the third transmission attempt, several new packets arrived from upper layer, but due to Assumption 2 they were not appended to the existing in-service A-MPDU (of size two). The effect of Assumption 2 is that new packets cannot be appended to an in-service A-MPDU. The time interval $t_2 - t_1$ is usually referred to as the *service time* [26], [27] or the *access delay* [28], [29].

B. THREE-DIMENSIONAL MARKOV CHAIN

Due to Assumption 1, all stations always compete for channel access and consequently the statistical characteristics of each station are identical to each other. It is a common practice in the literature to assume that the stations have identical characteristics. For example, the models in [5], [6], [7], [8], [9], [10], [11], [12], and [13] assume a fully saturated traffic condition, so the stations always compete for channel access to send A-MPDUs of maximum size. In a result, in the long run, all stations experience the same transmission and collision probabilities. It helps to simplify the model or test new protocols. Therefore, in devising the performance model of aggregation-enabled 802.11n/ac WLANs, we can focus on the operations of a single station, as in the models of [5], [6], [7], [8], [9], [10], [11], [12], [13], [15], [16], [17], [19], [20], [21], [22], [23], and [24]. On the other hand, it is possible to model the performance of WLANs where the stations are not fully identical. For example, the work in [18], models the performance of 802.11g WLAN where stations have different offered loads. Another work in [30] takes into account individual frame error rates at each station to model the performance of 802.11g WLAN, considering rate adaptation algorithm. However, modeling the network with diverse stations comes at a cost – the model becomes more complex. In this work, we also take into account the existence of anomalous slots found by the authors in [16].

We take a discrete and integer timescale where t and $t + 1$ correspond to the beginning time moments of two consecutive slots, and the backoff counter of a station is decremented at the end of each slot. We propose a three-dimensional (3D) Markov chain to mathematically analyze the performance of aggregation-enabled 802.11n/ac WLANs for a given offered load. In addition to the stochastic process for a backoff counter, $b(t)$, and the stochastic process for a backoff stage, $s(t)$, which were introduced by Bianchi in [15], we introduce a new stochastic process $q(t)$, which represents the queue size $(1, \ldots, Q)$ of the station. Without loss of generality, we consider the queue size limit, Q, to be greater than or equal to the maximum aggregation size A. Unlike $b(t)$ and $s(t)$, due to Assumption 2, $q(t)$ can change only when the A-MPDU service is finished (either successfully or unsuccessfully). It does not mean that newly arriving packets during the service of an A-MPDU are ignored. But once the service is finished, we determine the number of newly arrived packets based on the offered traffic load and the average service time of the previous A-MPDU.

In the proposed Markov chain, each state is represented by (n, k, v), where

- n is the number of packets in a queue (i.e, queue size) right after the service of the previous A-MPDU finishes and $n \in [1, Q]$,
- k is the current backoff stage and $k \in [0, r]$,
- v is the value of the backoff counter at time t and $v \in [0, W_k)$.

Due to the lack of space and to avoid complexities in description, Figure 2 explicitly depicts only the outgoing transitions from the states with queue size n to the states with queue size m. The figure illustrates only the states of backoff stage 0 for queue size m. We expect that one can easily derive the outgoing transitions from all other queue sizes. l_n, given in the caption of Figure 2, is the number of packets aggregated as one A-MPDU (i.e., the aggregation size) when there are n packets in the queue. Thus, $l_n = \min(n, A)$ where A is the maximum size of aggregation and is 64 in the standard [1].

Let us examine the states with queue size n, which are enclosed by the colored circles inside a rectangle with the red and dashed line in Figure 2. There is a special 2D state $(n, 0^*)$ that looks different from the other 3D states. This state represents the anomalous slot that is accessible only by a station that has just completed a successful transmission. Therefore, the transmission of an A-MPDU from the state $(n, 0^*)$ is always successful.

On the other hand, regular 3D states with the same queue size can be differentiated from each other by two components: a backoff stage (k) and a backoff counter (v). Thus, when focused on any specific queue size, its states can be expressed as a 2D process that is similar to those in [15] and [16]. Note that the biggest difference between our 3D process and the

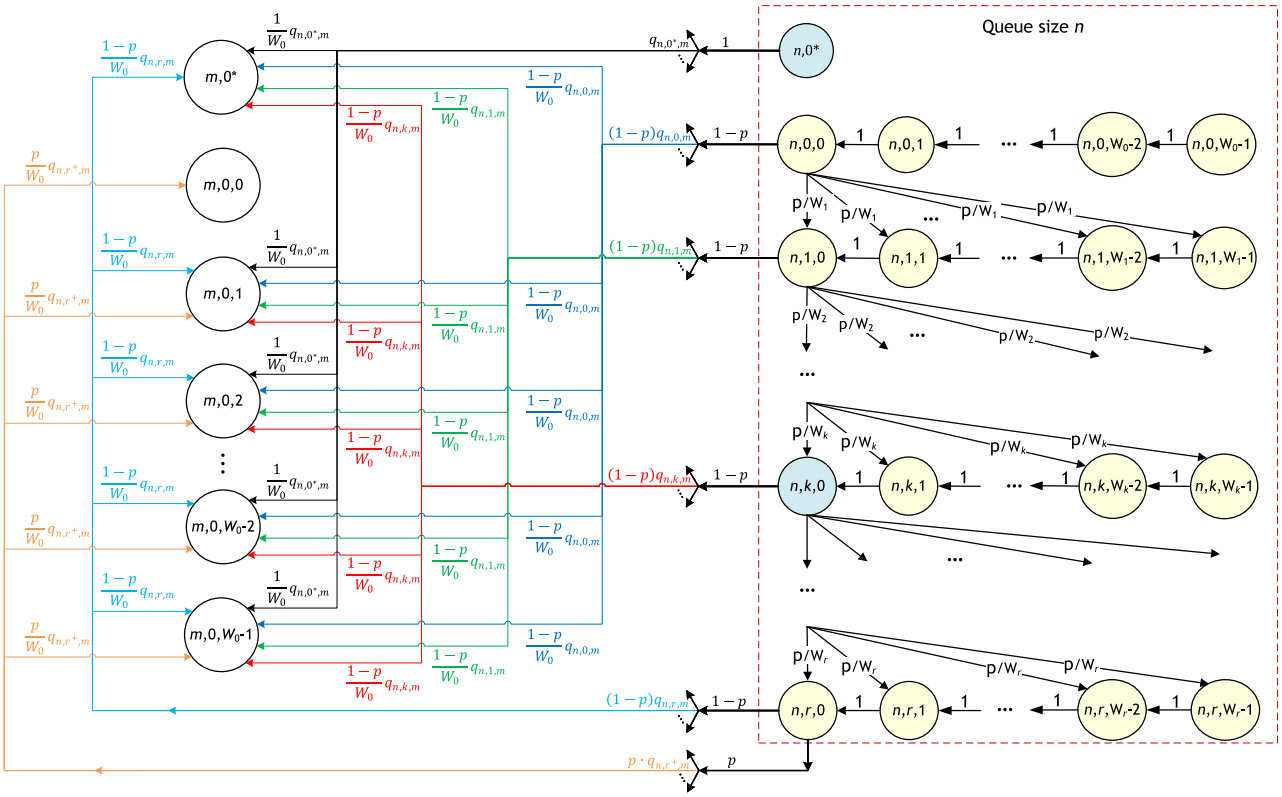


FIGURE 2. State transition diagram of outgoing transitions from the states with queue size n, where $n \in [1, Q]$, $k \in [0, r]$, and $m \in [\max(1, n - l_n), Q]$.

2D processes in [15] and [16] is manifested in the outgoing transitions to states with other queue sizes. Let us examine the outgoing transitions from the states with queue size n. As it was discussed in Subsection II-A, when the backoff counter reaches zero at backoff stages $k \in [0, r]$, i.e., in the states $(n, k, 0)$, the station transmits the A-MPDU.

- If the transmission is successful, the station transits to either $(m, 0^*)$ or one of the $(m, 0, v)$ states. The queue size m is determined by the number of new packet arrivals during the service time of the last A-MPDU. If the newly extracted random backoff counter v is 0, then the transition to the state $(m, 0^*)$ occurs; otherwise, i.e., $v \in [1, W_0)$, a transition occurs to the state $(m, 0, v)$.
- If the A-MPDU transmitted from the state $(n, k, 0)$ collides, the backoff stage is increased to $k + 1$ and a new backoff counter v is randomly selected from $[0, W_{k+1})$, i.e., transition happens to a state $(n, k + 1, v)$.
- Upon unsuccessful transmission from the state $(n, r, 0)$ (i.e., unsuccessful service and drop/discard of the A-MPDU), the station moves to one of the states $(m, 0, v)$, where $v \in [1, W_0)$. Note that the anomalous slot, that is, the state $(m, 0^*)$, is accessible only after a successful transmission with probability $1/W_0$. Therefore, after unsuccessful service, if a newly selected random backoff counter is 0, the station transits to the state $(m, 0, 0)$, not the state $(m, 0^*)$.
- Transmissions from the state $(n, 0^*)$ are always successful, since the state is accessible only by a station that has just finished a successful transmission. Therefore, after transmission, the state moves to $(m, 0^*)$ or one of the $(m, 0, v)$ states, where $v \in [1, W_0)$.

C. ONE-STEP TRANSITION PROBABILITIES

Let $P\{u|s\}$ denote the one-step transition probability from state s to state u. We examine the outgoing transition probabilities from the states with queue size n. Following the discussions of Subsections III-A and III-B, we can easily derive the transition probabilities of our chain in Figure 2. In Subsections III-C – III-D, since n and m represent the size of queue, $n \in [1, Q]$ and $m \in [1, Q]$ unless otherwise stated.

First, let us focus on the outgoing transitions from state $(n, 0^*)$. As discussed in Subsection III-B, all transmissions from this state are successful, since this anomalous slot is accessible only by a station that has just finished the successful transmission and has selected a zero backoff count with probability $1/W_0$. After the transmission, the station enters another anomalous slot state $(m, 0^*)$ if it again randomly selects zero as its new backoff counter value with probability $1/W_0$. With the same probability, the station chooses a non-zero $v \in [1, W_0)$ as a new backoff counter value and transits to the $(m, 0, v)$ state. The probability of such transitions is

as follows.

$$P\{(m,0^*) \mid (n,0^*)\} = P\{(m,0,v) \mid (n,0^*)\} = \frac{1}{W_0} \cdot q_{n,0^*,m}, \quad (1)$$

where $q_{n,0^*,m}$ is the probability that there are m packets in the queue following a successful transmission from the state $(n, 0^*)$.

Unlike the state $(n, 0^*)$, successful transmission from the state $(n, k, 0)$ occurs with probability $1 - p$, and then with the probability $1/W_0$ the state transitions to the anomalous slot or one of the regular slots among $v \in [1, W_0)$. The exact expression of the probabilities of those transitions is given as follows.

$$P\{(m,0^*) \mid (n,k,0)\} = P\{(m,0,v) \mid (n,k,0)\} = \frac{1-p}{W_0} \cdot q_{n,k,m}, \quad (2)$$

where $k \in [0, r]$ and $q_{n,k,m}$ is the probability that there are m packets in the queue following a successful transmission from the state $(n, k, 0)$.

With probability p, transmission from the state $(n, k, 0)$ encounters a collision, consequently, the station increases the backoff stage and enters the state $(n, k + 1, v)$.

$$P\{(n,k+1,v) \mid (n,k,0)\} = \frac{p}{W_{k+1}}, \quad (3)$$

where $v \in [0, W_{k+1})$ and $k \in [0, r)$.

When a transmission from the state $(n, r, 0)$ collides, the station ceases the retransmission procedure, drops/discards the current A-MPDU, and transits to the $(m, 0, v)$ state with the following probability.

$$P\{(m,0,v) \mid (n,r,0)\} = \frac{p}{W_0} \cdot q_{n,r^+,m}, \quad (4)$$

where $v \in [0, W_0)$ and $q_{n,r^+,m}$ is the probability that the queue has m packets just after the drop of the A-MPDU due to an unsuccessful transmission from the backoff stage r. We will derive $q_{n,0^*,m}$, $q_{n,k,m}$, and $q_{n,r^+,m}$ in Subsection III-G.

Finally, we examine the state transitions representing the decrement of the backoff counter at the end of an idle slot as follows.

$$P\{(n,k,v-1) \mid (n,k,v)\} = 1, \quad (5)$$

where $v \in [1, W_k)$ and $k \in [0, r]$.

D. STEADY-STATE PROBABILITIES

Let π_s denote the limiting distribution of state s. We can easily derive the global balance equations of all feasible states from the state transition probability diagram in Figure 2 because in steady-state, the outgoing rate from a state is equal to the incoming rate to the state. Let us derive the balance equations for the states with queue size n. Since the states representing the anomalous slots have incoming one-step transitions only upon successful transmissions, their steady-state probabilities can be obtained as follows

$$\pi_{n,0^*} = \frac{1}{W_0} \sum_{m=1}^{Q} \Big(q_{m,0^*,n} \cdot \pi_{m,0^*} + (1-p) \sum_{k=0}^{r} (q_{m,k,n} \cdot \pi_{m,k,0}) \Big). \quad (6)$$

The state $(n, 0, 0)$ has the following incoming transitions: i) from the neighboring state $(n, 0, 1)$ at the end of an idle slot and ii) after unsuccessful transmission from state $(m, r, 0)$. Thus, its steady-state probability can be written as

$$\pi_{n,0,0} = \pi_{n,0,1} + \frac{p}{W_0} \cdot \sum_{m=1}^{Q} (q_{m,r^+,n} \cdot \pi_{m,r,0}). \quad (7)$$

In contrast, the state $(n, 0, W_0-1)$ has the following incoming transitions: i) after transmission from the anomalous slot state $(m, 0^*)$, ii) after successful transmission from state $(m, k, 0)$ where $k \in [0, r]$, and ii) after unsuccessful transmission from state $(m, r, 0)$. Thus, the steady-state probability is given as

$$\pi_{n,0,W_0-1} = \frac{1}{W_0} \sum_{m=1}^{Q} \Big(q_{m,0^*,n} \cdot \pi_{m,0^*} + (1-p) \sum_{k=0}^{r} (q_{m,k,n} \cdot \pi_{m,k,0}) + p \cdot q_{m,r^+,n} \cdot \pi_{m,r,0} \Big). \quad (8)$$

We can also easily derive the following relationship between the states $(n, 0, v)$ and $(n, 0, W_0 - 1)$ where $v \in [1, W_0)$.

$$\pi_{n,0,v} = (W_0 - v) \cdot \pi_{n,0,W_0-1}. \quad (9)$$

Since the backoff stage is incremented whenever the frame collides, the steady state probability of any $(n, k, 0)$ can be written in terms of $\pi_{n,0,0}$.

$$\pi_{n,k,0} = p^k \cdot \pi_{n,0,0}, \qquad k \in [0, r]. \quad (10)$$

The following relationship holds between the steady-state probabilities of states $(n, k, 0)$ and (n, k, v).

$$\pi_{n,k,v} = \frac{W_k - v}{W_k} \cdot \pi_{n,k,0}, \quad (11)$$

where $k \in [1, r]$ and $v \in [1, W_k)$.

Note that it may be intractable to get a closed-form expression of steady-state probabilities because the collision probability p depends on these steady-state probabilities being recursively expressed by p. On the other hand, the drop possibility of an A-MPDU is expected to be very small, because it needs to collide $r + 1$ times successively. Furthermore, just after such an infrequent drop event, the station selects the backoff counter 0 with probability $1/W_0$, and thus the second term on the right side of (7) has negligible values for reasonable values of p. To obtain a closed-form expression, we approximate (6) and (7) by removing the

second term on the right side of (7) and adding it to (6). Now, the equations for $\pi_{n,0^*}$ and $\pi_{n,0,0}$ are updated as follows.

$$\pi_{n,0^*} \cong \frac{1}{W_0}\sum_{m=1}^{Q}\Big(q_{m,0^*,n}\cdot\pi_{m,0^*} + (1-p)\sum_{k=0}^{r}(q_{m,k,n}\cdot\pi_{m,k,0}) + p\cdot q_{m,r^+,n}\cdot\pi_{m,r,0}\Big). \tag{12}$$

$$\pi_{n,0,0} \cong \pi_{n,0,1}. \tag{13}$$

Note that the expressions for the steady-state probabilities in (8) and (12) look identical, i.e., $\pi_{n,0,W_0-1} = \pi_{n,0^*}$. Then the expression in (9) can be re-written as $\pi_{n,0,v} = (W_0 - v)\cdot\pi_{n,0^*}$, for $v \in [1, W_0)$. Since $\pi_{n,0,0} = \pi_{n,0,1}$ from (13), consequently

$$\pi_{n,0,0} = (W_0 - 1)\cdot\pi_{n,0^*}. \tag{14}$$

From (10), (11), and (14), the followings hold for all (k, v) such that $k \in [1, r]$ and $v \in [0, W_k)$.

$$\pi_{n,k,v} = \frac{W_k - v}{W_k}\cdot p^k\cdot(W_0-1)\cdot\pi_{n,0^*}. \tag{15}$$

Let us substitute $\pi_{m,k,0}$ and $\pi_{m,r,0}$ in (12) with (15) and define

$$\Phi(m,n) := \frac{q_{m,0^*,n}}{W_0} + \frac{W_0-1}{W_0}\cdot\Big((1-p)\cdot\sum_{k=0}^{r}(p^k\cdot q_{m,k,n}) + p^{r+1}\cdot q_{m,r^+,n}\Big), \tag{16}$$

Now we re-express (12) as follows

$$\pi_{n,0^*} = \sum_{m=1}^{Q}\Phi(m,n)\cdot\pi_{m,0^*}. \tag{17}$$

and consequently

$$\pi_{n,0^*} = \frac{\sum_{m=1,m\neq n}^{Q}\Phi(m,n)\cdot\pi_{m,0^*}}{1-\Phi(n,n)}. \tag{18}$$

E. TRANSMISSION AND COLLISION PROBABILITIES

The sum of all feasible state probabilities is one. That is,

$$\sum_{n=1}^{Q}\Big(\pi_{n,0^*} + \sum_{k=0}^{r}\sum_{v=0}^{W_k-1}\pi_{n,k,v}\Big) = 1. \tag{19}$$

After applying (14) and (15) to (19),

$$\sum_{n=1}^{Q}\pi_{n,0^*} = \Big(\frac{W_0(W_0+1)}{2} + \frac{W_0-1}{2}\sum_{k=1}^{r}p^k(W_k+1)\Big)^{-1}. \tag{20}$$

Since the transmission can occur only at the states with zero backoff count, the transmission probability by a station in a random slot time, τ, can be obtained as follows

$$\tau = \sum_{n=1}^{Q}\Big(\pi_{n,0^*} + \sum_{k=0}^{r}\pi_{n,k,0}\Big) = \Big(1 + \frac{(W_0-1)(1-p^{r+1})}{1-p}\Big)\cdot\sum_{n=1}^{Q}\pi_{n,0^*}. \tag{21}$$

After applying (20) to (21),

$$\tau = \frac{2\Big(1-p+(W_0-1)(1-p^{r+1})\Big)}{(1-p)\Big(W_0(W_0+1)+(W_0-1)\sum_{k=1}^{r}p^k(W_k+1)\Big)}. \tag{22}$$

The collision probability is calculated as in [15] and [16].

$$p = 1-(1-\tau)^{N_{STA}-1}, \tag{23}$$

where N_{STA} denotes the number of stations in the network. Using (22) and (23), we can numerically find p and τ for a given N_{STA}.

F. THROUGHPUT

The throughput is calculated as the average amount of data payload successfully transmitted during an average system slot time. First, we derive the average system slot time, denoted by E_{slot}. Note that the slots are classified into *idle*, *success*, and *collision* slots. P_{idle} is the probability that the slot is detected as idle and can be found using $P_{idle} = (1-\tau)^{N_{STA}}$. The *idle* slot duration is σ, which is a standard slot duration in 802.11 [1].

Now, let P_{succ} denote the probability that a slot contains a successful transmission, i.e., *success* slot probability. Then

$$P_{succ} = N_{STA}\cdot\tau\cdot(1-\tau)^{N_{STA}-1}. \tag{24}$$

T_{succ} represents the average duration of *success* slot and is expressed as

$$T_{succ} = T_{RTS} + T_{CTS} + T_{pre} + \Big\lceil\frac{E_{aggr}(L_{hdr}+L_{pld})}{R\cdot T_{SYM}}\Big\rceil\cdot T_{SYM} + 3T_{SIFS} + T_{BACK} + T_{DIFS}, \tag{25}$$

where T_{RTS}, T_{CTS}, T_{pre}, T_{SYM}, T_{SIFS}, T_{BACK}, and T_{DIFS} represent the durations of an RTS frame, a CTS frame, a PHY preamble, and the header of A-MPDU, an OFDM symbol, SIFS, a blockACK frame, and DIFS, respectively; the outcome of ceiling operation represents the total number of OFDM symbols required to carry the A-MPDU bits, where E_{aggr} is the expected aggregation size, L_{hdr} is the total length of MAC, IP, and UDP layer headers, L_{pld} is the UDP payload length, and R is the PHY data transmission rate.

The slot contains a collision if multiple stations simultaneously transmit. Let P_{coll} be the probability of such a *collision*

slot. Then, $P_{coll} = 1 - P_{idle} - P_{succ}$ and its duration is $T_{coll} = T_{RTS} + T_{SIFS} + T_{CTS} + T_{DIFS}$.

The average system slot time is $E_{slot} = P_{idle}\sigma + P_{succ}T_{succ} + P_{coll}T_{coll}$, and the average amount of data payload within an A-MPDU is $E_{aggr} \cdot L_{pld}$. Since the total throughput of the system is the amount of data payload successfully transmitted during an average system slot time,

$$S = \frac{P_{succ} \cdot E_{aggr} \cdot L_{pld}}{P_{idle} \cdot \sigma + P_{succ} \cdot T_{succ} + P_{coll} \cdot T_{coll}}. \tag{26}$$

G. OFFERED LOAD AND SERVICE TIME RELATIONSHIP

As shown in Figure 2, the probabilities of a one-step transition between the queue sizes of n and m rely on the probabilities $q_{n,o*,m}$, $q_{n,k,m}$, and $q_{n,r^+,m}$. In this subsection, we explain how to derive them. Let Λ denote the total offered load in bits per second (bps) of the system of N_{STA} stations transmitting uplink traffic. Assuming that all stations have the same traffic characteristics, such as the packet arrival rate and the payload size of a packet, the average packet arrival rate at each station is $\lambda = (\Lambda/N_{STA})/L_{pld}$ packets per second. Let us assume that the packets arrive in a Poisson manner. Then, the general expression for the probability that i packets arrive during a time duration t is given as $\beta_i(t) = ((\lambda t)^i e^{-\lambda t})/i!$.

Since A is the maximum number of packets that can be contained within an A-MPDU, the size of an A-MPDU served in the states with queue size n can be expressed as $l_n = \min(n, A)$. Then, let $T_{n,k}$ be the average service duration of an A-MPDU with size l_n that is successfully transmitted at the backoff stage k, i.e., at the state $(n, k, 0)$. Note that the average service duration of an A-MPDU is the time interval from its generation until a blockACK for this A-MPDU is received. Thus,

$$q_{n,k,m} = \begin{cases} \beta_0(T_{n,k}) + \beta_1(T_{n,k}), & n \le A, m = 1 \\ \beta_m(T_{n,k}), & n \le A, m \in (1, Q) \\ 1 - \sum_{j=0}^{Q-1} \beta_j(T_{n,k}), & n \le A, m = Q \\ 0, & n > A, m < n - A \\ \beta_{m-n+A}(T_{n,k}), & n > A, m \in [n - A, Q) \\ 1 - \sum_{j=0}^{Q-1-n+A} \beta_j(T_{n,k}), & n > A, m = Q. \end{cases} \tag{27}$$

Let us first explain (27) for the cases when $n \le A$. In these cases, the aggregation size of A-MPDU l_n is n, and therefore no packet remains in the queue when the service of A-MPDU starts. According to assumption 1, at least one packet must be in a queue after the service finishes. Consequently, we take the probability of one packet in a queue as the sum of Poisson probabilities of one and no packet arrivals, as shown in the first equation of (27). The second equation is the probability of the cases where the queue has two or more packets after the service finishes, but is not yet full. The third equation is the probability of the case where the queue is full of new packets following the completion of a service.

Now, let us discuss the remaining three cases in (27) when $n > A$. In these cases, the aggregation size of the A-MPDU l_n is A and, therefore, the queue has $(n - A)$ packets remaining after the service of this A-MPDU starts. That is, a queue having less than $(n - A)$ packets after the service ends is impossible, and thus the probability of such a case is zero, as shown in the fourth equation. The next probability is for the cases where zero or more packets arrive during the service, but the queue is not yet full. The last probability is for the case when the queue is full of packets following a service.

Now, let us give the expression for $T_{n,k}$. In [31] and [32], Raptis et al. proposed a new approximation to estimate the per-stage average service time for the legacy 802.11 WLAN with saturated traffic. Note that if the A-MPDU is successfully transmitted at the backoff stage k, then the total service time of this A-MPDU includes: i) $(k + 1)$ random backoff periods, ii) k collisions, and iii) successful transmission at the stage k. During the backoff period, the target station waits for its backoff counter to expire, and the remaining $(N_{STA} - 1)$ stations compete for access. Let E^-_{slot} be the average duration of a slot time when the remaining stations, except the tagged station, compete for access. Then, since the average backoff counter at stage i is $(W_i - 1)/2$, the total average backoff waiting time of the tagged station is $E^-_{slot} \sum_{i=0}^{k} (W_i - 1)/2$. Let P^-_{idle}, P^-_{succ} and P^-_{coll} be the probabilities of idle, success, and collision slots in the situation that the remaining $(N_{STA} - 1)$ stations except the tagged station compete for medium access, then

$$E^-_{slot} = P^-_{idle}\sigma + P^-_{succ}T_{succ} + P^-_{coll}T_{coll}, \tag{28}$$

where $P^-_{idle} = (1 - \tau)^{N_{STA}-1}$, $P^-_{succ} = \tau \cdot (N_{STA} - 1) \cdot (1 - \tau)^{N_{STA}-2}$, and $P^-_{coll} = 1 - P^-_{idle} - P^-_{succ}$. The duration of a *success* slot for an A-MPDU with size l_n is denoted by $T_{succ}(l_n)$ and can be easily calculated using (25), by replacing E_{aggr} with l_n as follows.

$$\begin{aligned} T_{succ}(l_n) = {} & T_{RTS} + T_{CTS} + T_{pre} \\ & + \left\lceil \frac{l_n(L_{hdr} + L_{pld})}{R \cdot T_{SYM}} \right\rceil \cdot T_{SYM} \\ & + 3T_{SIFS} + T_{BACK} + T_{DIFS}. \end{aligned} \tag{29}$$

Finally, $T_{n,k}$ can be expressed as

$$T_{n,k} = \sum_{i=0}^{k} \frac{W_i - 1}{2} E^-_{slot} + k \cdot T_{coll} + T_{succ}(l_n). \tag{30}$$

The probabilities $q_{n,0^*,m}$ and $q_{n,r^+,m}$ can be obtained as follows. $q_{n,0^*,m}$ is calculated using (27) but replacing $T_{n,k}$ with $T_{n,0^*}$, which is the sojourn time of state $(n, 0^*)$. Since transmitting an A-MPDU in the state $(n, 0^*)$ is always successful, it is obvious that $T_{n,0^*} = T_{succ}(l_n)$. Similarly, $q_{n,r^+,m}$ is also calculated by (27) but replacing $T_{n,k}$ with T_{n,r^+}, which is the average unsuccessful service time, that is, the average time

to drop/discard the A-MPDU due to the exhausted retry limit and can be obtained as follows.

$$T_{n,r^+} = \sum_{i=0}^{r} \frac{W_i - 1}{2} E_{slot}^- + (r+1) \cdot T_{coll} \tag{31}$$

H. AGGREGATION SIZE

To find the expected aggregation size, E_{aggr}, one needs to obtain the aggregation size distribution. The probability mass function (PMF) of the aggregation size l_n, for $l_n < A$, equals the sum of all steady-state probabilities of queue size n. However, for $l_n = A$, the PMF is calculated as the sum of all steady-state probabilities of all queue sizes greater than or equal to A, as given in the following systems of equations.

$$P(l_n) = \begin{cases} \pi_{n,0^*} + \sum_{k=0}^{r} \sum_{v=0}^{W_k-1} \pi_{n,k,v}, & l_n < A \\ \sum_{n=A}^{Q} \left(\pi_{n,0^*} + \sum_{k=0}^{r} \sum_{v=0}^{W_k-1} \pi_{n,k,v} \right), & l_n = A. \end{cases} \tag{32}$$

After some manipulations, we can rewrite the above expression for PMF as follows.

$$P(l_n) = \begin{cases} \left(\frac{W_0(W_0+1)}{2} + \frac{W_0-1}{2} \cdot \sum_{k=0}^{r} p^k (W_k+1) \right) \\ \qquad \cdot \pi_{n,0^*}, & l_n < A \\ \left(\frac{W_0(W_0+1)}{2} + \frac{W_0-1}{2} \cdot \sum_{k=0}^{r} p^k (W_k+1) \right) \\ \qquad \cdot \sum_{n=A}^{Q} \pi_{n,0^*}, & l_n = A. \end{cases} \tag{33}$$

Finally, we are in a position to derive the average aggregation size, E_{aggr}. It can be obtained using the PMF as follows.

$$E_{aggr} = \sum_{n=1}^{Q} l_n \cdot P(l_n). \tag{34}$$

Up to this point, we knew the values of τ and p only, for the given N_{STA}. Now, we can build a system of the following $(Q+1)$ linear equations to obtain the values of $\pi_{n,0^*}$ and E_{aggr} as follows.

- $(Q-1)$ independent linear equations of $\pi_{n,0^*}$ for $n \in [1, Q)$, from (18)
- the equation for the case of $n = Q$ as $\pi_{Q,0^*} = \frac{\tau(1-p)}{1-p+(W_0-1)(1-p^{r+1})} - \sum_{n=1}^{Q-1} \pi_{n,0^*}$, from (21)
- the last equation for E_{aggr} using (34).

Then, the system of equations can be solved numerically to find $\pi_{n,0^*}$ and E_{aggr} for given N_{STA}, A, Q, Λ, R, etc.

I. THROUGHPUT DERIVATION FOR EXISTING MODELING APPROACHES

As explained in Subsections II-D and II-E, most of the existing works model the condition of fully saturated traffic.

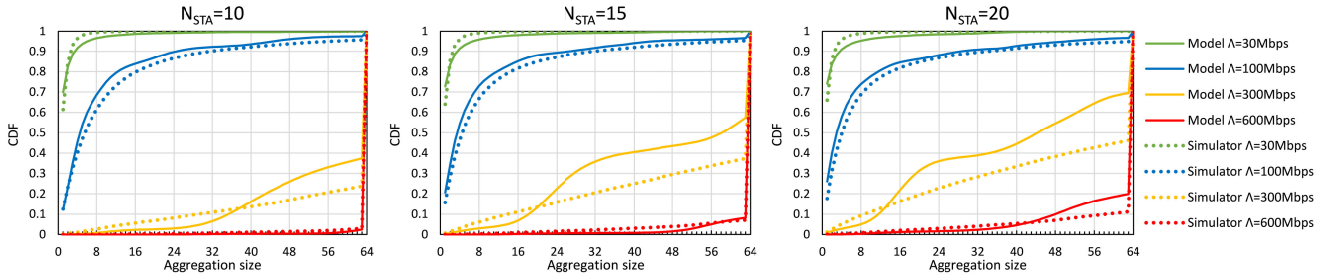
TABLE 1. System parameters.

Symbol	Name	Values
A	Maximum aggregation size (packets)	64
L_{hdr}	The sum length of MAC, IP, and UDP layer headers	64 bytes
N_{STA}	Number of stations	10, 15, 20
Q	Queue size limit (packets)	100 and 150
R	PHY data transmission rate	150 Mbps
r	Retry limit	6
T_{BACK}	blockACK duration	32 μs
T_{CTS}	CTS duration	44 μs
T_{DIFS}	DIFS duration	34 μs
T_{pre}	PHY preamble and header duration	36 μs
T_{RTS}	RTS duration	34 μs
T_{SIFS}	SIFS duration	16 μs
W_0	Minimum contention window size	16
W_{max}	Maximum contention window size	1024
σ	Idle/standard slot duration	9 μs
Λ	Total offered traffic load of the network	10, 20, .., 600 Mbps

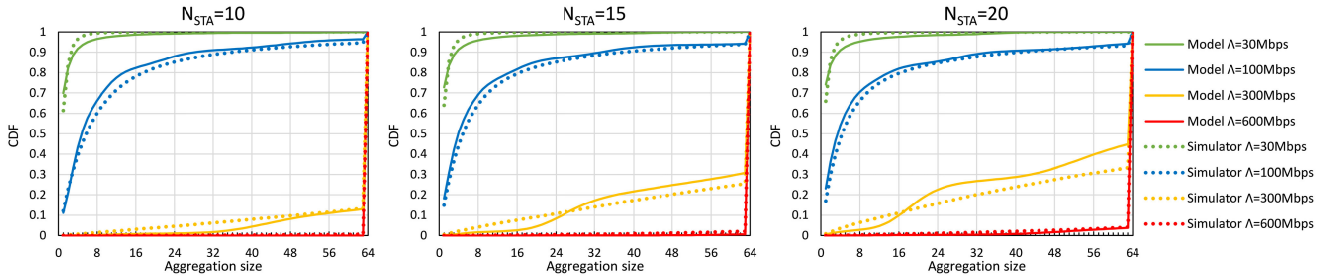
Those works can be divided into two categories. The works in the first category assume that all stations always have the necessary number of packets in their buffers to compose and send A-MPDU of maximum aggregation size [4], [5], [6], [7]. The models in the second category assume that the stations always have enough packets to compose and send the A-MPDU of the aggregation size that the considered ARQ scheme currently allows [8], [9], [10], [11], [12], [13]. However, this work focuses on the ideal channel condition only, and if such a condition is applied to the models in the second category, the impact of ARQ schemes on the aggregation size disappears and thus stations always transmit A-MPDUs of maximum aggregation size. Therefore, existing models in both categories can be considered as a single modeling approach where stations always transmit A-MPDUs of maximum aggregation size, A. The stations in the existing modeling approach always contend for channel access, therefore, the transmission and collision probabilities are the same as in the proposed model and thus can be easily obtained using (22) and (23), respectively. Furthermore, the probabilities and durations of *success*, *collision*, and *idle* slots are also calculated exactly the same way as in the proposed model discussed in III-F but except for the duration of the *success* slot. To obtain the duration of the *success* slot, we still use (25) but replace E_{aggr} with A. Finally, we can obtain the throughput of the existing modeling approach using (26).

IV. PERFORMANCE EVALUATIONS

To demonstrate the performance accuracy of the proposed model, we developed a Python-based 802.11n/ac simulator with the capability of A-MPDU aggregation and support for Poisson traffic generation at each station [33]. Assumptions 1 and 2 are not applied to the simulator stations, i.e., the stations do not generate a dummy packet when the queue is empty and the size of in-service A-MPDU can increase if some new packets arrive during its service. In other words, the stations in the simulator mimic the real 802.11n/ac



(a) Queue size limit: Q = 100 packets



(b) Queue size limit: Q = 150 packets

FIGURE 3. CDF of aggregation size comparison for different total offered traffic load (Λ), number of stations (N_{STA}), and queue size limits (Q).

stations where these assumptions are not applied. Our intention in not including the Assumptions 1 and 2 was to check how similar is the performance of our proposed model to the performance of actual operations of WLAN. All other assumptions and system parameters in the simulator are the same as in the proposed model. More details of the simulator implementation can be found in our GitHub repository [33]. All system parameters and their values are summarized in Table 1.

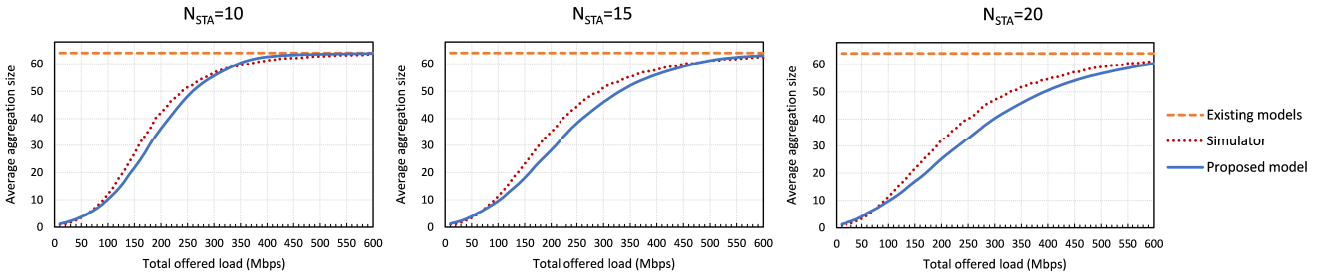
A series of simulations are run for N_{STA} (10, 15, and 20) stations with different total offered loads of Λ (10, 20, …, 600 Mbps) and queue size limits of Q (100 and 150), which compete to transmit uplink traffic for 30 seconds. For each $\{N_{STA}, \Lambda, Q\}$ configuration, the simulation is run with five different seed values, and the average performance is recorded. The values of the other main parameters are constant: $A = 64$ packets, $R = 150$ Mbps, and $L_{pld} = 6400$ bits. Corresponding simulation script is included in our GitHub repository [33]

A. AGGREGATION SIZE PERFORMANCE

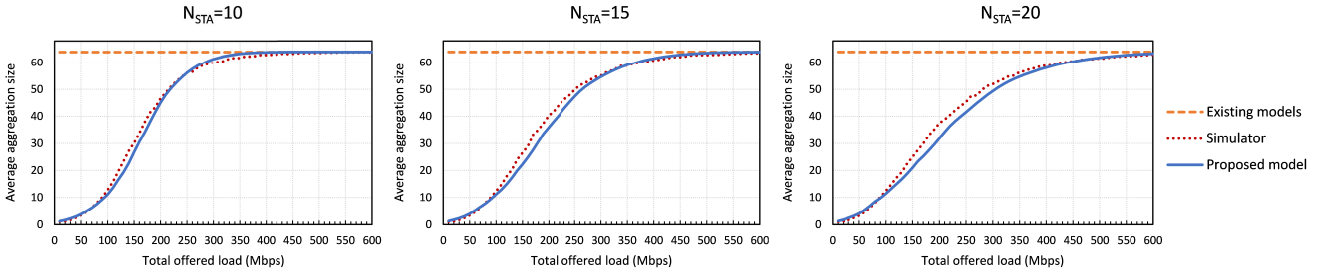
Figure 3 illustrates the cumulative distribution function (CDF) of aggregation size for different Q, N_{STA}, and Λ settings, where solid lines represent the theoretical CDF produced by the proposed model and the dotted lines represent the empirical CDF measured in the simulator. We could also include the CDF for the existing modeling approach, but it would include only value 1 at 64, since all A-MPDUs have the aggregation size of A.

At loads such as $\Lambda \leq 100$ Mbps, the proposed model and the simulator have a similar aggregation size distribution. However, there is a significant difference at an aggregation size of 1. At very small loads, the simulator stations do not always have packets to send; thus, they do not continuously compete for channel access. Moreover, newly arriving packets can be appended to in-service A-MPDU that is currently being served. In contrast, in the proposed model, due to Assumption 1, the station always has a packet to send, and thus they always compete for channel access. Moreover, because of Assumption 2, the in-service A-MPDU cannot include newly arriving packets. These two assumptions result in more A-MPDU transmissions with the aggregation size of 1. Thus, at $\Lambda = 30$ Mbps, about 70-75% of A-MPDUs have an aggregation size of 1 in the proposed model, while the corresponding proportion in the simulator is merely about 60-65%.

As Λ increases, the probability of not having a packet to send is understandably decreases, and the need for additional packets to satisfy Assumption 1 in the proposed model also decreases. However, due to the Assumption 2 in the proposed model, the number of A-MPDUs with a single packet can still be higher than in the simulator. That is the case for $\Lambda = 100$ Mbps in Figure 3a. For N_{STA}=10, the CDFs of the proposed model and simulator are very similar at an aggregation size of 1. That is because the offered load per station is high and the effect of assumptions 1 and 2 is minimal. For N_{STA}=15 and 20, however, about 21% and 26% of the A-MPDUs in the proposed model include a single packet, respectively. On the other hand, in the simulator, the CDF at aggregation size 1 is approximately 16% and 18% for N_{STA}=15 and 20, respectively. A very similar situation can be observed for $Q = 150$ packets in Figure 3b.



(a) Queue size limit: Q = 100 packets



(b) Queue size limit: Q = 150 packets

FIGURE 4. Average aggregation size performance comparison for different total offered traffic load (Λ), number of stations (N_{STA}), and queue size limits (Q).

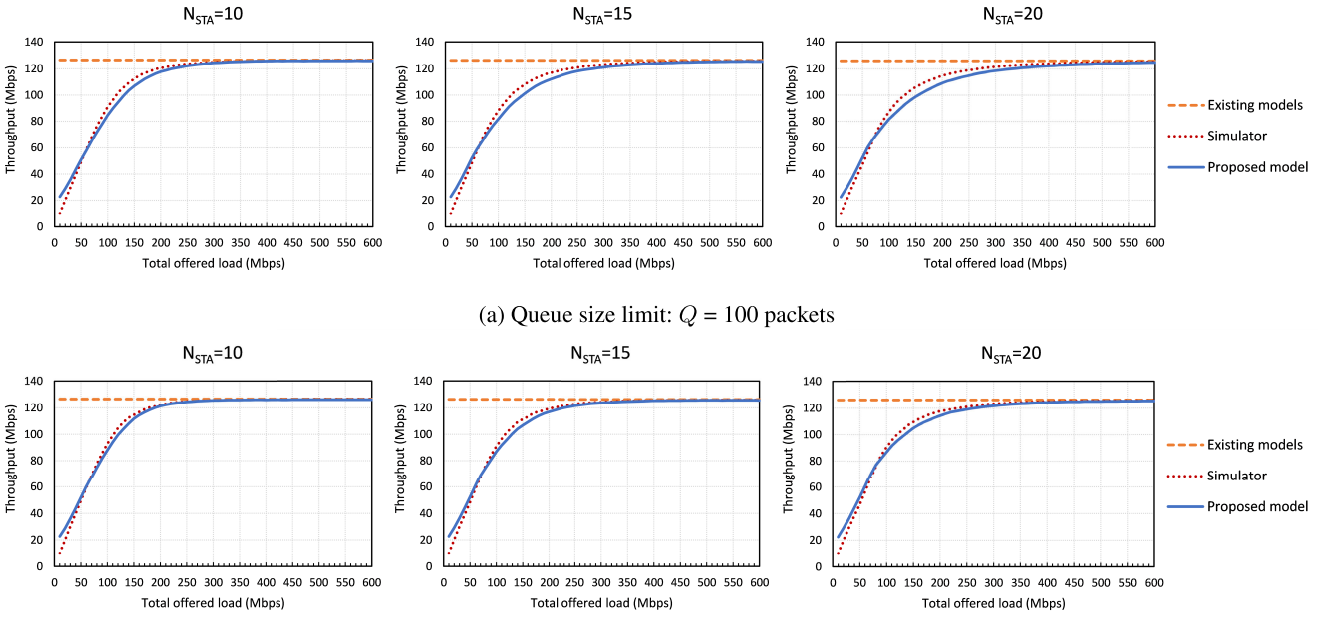
At high offered loads such as $\Lambda = 300$ Mbps, simulator stations transmit A-MPDUs with larger sizes; especially most A-MPDUs have aggregation size of A. The main reason is that Assumption 2 is not applied to the simulator, and thus in-service A-MPDU can include new packets that arrive before its successful transmission starts. The aggregation size of in-service A-MPDU can grow from 1 to A before its successful transmission. This means that the simulator stations generally have more space for packets compared to the proposed model. In other words, the effective system size of a station, which is defined as the maximum allowable number of packets within an in-service A-MPDU and the queue, can be as high as $Q+A$.

At very high loads, such as $\Lambda = 600$ Mbps, the effective system size of stations in both the proposed model and the simulator becomes very similar and close to $Q+A$. Therefore, most of the time, A-MPDUs include A packets in both the simulator and the proposed model. Moreover, with the larger Q, the proposed model and the simulator have much more similar effective system size and thus more similar performance.

Next, let us compare the average aggregation size. Figure 4 illustrates the average aggregation size produced by the proposed model and the simulator for different Q, N_{STA}, and Λ settings. It also includes the average aggregation size used in the existing modeling approach [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], represented by dashed yellow lines. Since all stations transmit A-MPDUs of maximum aggregation size in the existing modeling approach, the average aggregation size is fixed at 64.

At low traffic loads such as $\Lambda \leq 100$ Mbps, the average aggregation sizes of the proposed model and the simulator are very similar. At very small loads, for most of the time, the stations have no packets in their queues after the service of the previous A-MPDU. In the proposed model, if the queue is empty, an additional (dummy) packet is generated to keep the station continuously competing for medium access. However, in the simulator, the stations compete only when there is a real packet in the queue. In general, at low traffic loads, the stations transmit A-MPDUs of a single or just a few packets in both the simulator and the proposed model, and consequently, they produce very similar average aggregation sizes. The main difference is that, during a certain time interval, the stations in the proposed model transmit more A-MPDUs than in the simulator.

At higher offered loads, such as $\Lambda > 100$ Mbps, the stations transmit A-MPDU of larger sizes in the simulator than in the proposed model. The main reason is that assumption 2 is not applied to the simulator. As it is discussed in the previous paragraphs, in the proposed model, due to assumption 2 the size of the A-MPDU is fixed at the beginning of the service and newly arriving packets cannot be included in this A-MPDU. However, in the simulator, the in-service A-MPDU can include new packets that arrive before its transmission, which means the simulator stations have more space for newly arriving packets compared to the proposed model. It results in bigger effective system size of stations in the simulator, which leads to bigger aggregation sizes compared to the proposed model.



(a) Queue size limit: $Q = 100$ packets

(b) Queue size limit: $Q = 150$ packets

FIGURE 5. Average network throughput performance for different total offered traffic load (Λ), number of stations (N_{STA}), and queue size limits (Q).

At very high traffic loads, the average aggregation size performance of both the simulator and the proposed model becomes very similar. The reason is that due to a very high packet arrival rate, the effective system sizes of the proposed model and the simulator become very similar and thus most A-MPDUs have size A. When the queue size limit is increased from $Q = 100$ to 150 packets, the effective system size also increases. Consequently, we can see from Figure 4b that for a bigger Q, the average aggregation size increases faster as Λ continues increasing.

B. THROUGHPUT PERFORMANCE

Figure 5 depicts the average network throughput performance produced by the proposed model and the simulator for different Q, N_{STA}, and Λ. It also includes the average network throughput of the existing modeling approach that is derived according to the explanation in Subsection III-I.

At lower offered loads, the proposed model shows higher throughput than the simulator. As discussed in the previous subsection, for low loads, the average aggregation size is similar in the model and the simulator, but the number of A-MPDU transmissions is much higher in the proposed model due to additional packets generated by Assumption 1. Because of such additional packets, the proposed model produces higher throughput than the simulator.

As the offered load continues to increase, the simulator starts producing a higher throughput than the proposed model. The first reason is that the amount of additional packets generated by Assumption 1 starts to decrease in the proposed model because the probability that the station has no packet becomes smaller. The another reason is that in-service A-MPDUs in the simulator can include new packets and thus increase its aggregation size from 1 to A. Consequently, the effective system size of the simulator stations gets bigger. In the proposed model, Assumption 2 fixes the aggregation size of the A-MPDU when the service starts, so newly arriving packets are not appended to this A-MPDU but placed in the queue, leading to a smaller effective system size, a smaller average aggregation size, and consequently smaller throughput.

When the load further increases, due to a very high packet arrival rates, the effect of assumptions 1 and 2 diminishes, leading to very similar effective system sizes in simulator and the proposed model. As a result, the average aggregation sizes become very similar and consequently leading to similar throughput performance. At some value of the offered load, the throughput of the simulator and the proposed model become very similar to the throughput of the existing modeling approach, which means fully saturated condition is reached. Similarly to the case of average aggregation size performance, for larger Q, the throughput increases faster as the offered load continues increasing, eventually reaching the throughput of the existing modeling approach.

V. CONCLUSION

The conventional performance models of aggregation-enabled IEEE 802.11n/ac WLANs assume fully saturated traffic, so stations always contend to transmit A-MPDUs of maximum or other fixed aggregation size. However, in reality, the stations transmit A-MPDUs, of different aggregation sizes

depending on several factors such as offered load, number of stations, random backoff process, etc. In this work, we have proposed a new 3D Markov chain to model the performance of aggregation-enabled WLANs with variable aggregation size.

The performance accuracy of the proposed model has been evaluated for different offered loads, numbers of nodes, and queue size limits compared to the simulator and the existing modeling approach. At low and very high offered loads, the average aggregation size of A-MPDU in the model is very similar to that in the simulator. In contrast, the accuracy of throughput performance gets better as the offered load increases. Moreover, evaluations have shown that the larger queue size limit leads to higher accuracy in the aggregation size and throughput performance. As a result, the proposed model improves the overall accuracy in predicting the performance of aggregation-enabled WLANs where the aggregation size varies depending on dynamic factors such as offered traffic load, random backoff time, etc.

In our future work, we plan to extend the proposed model further so that it can accurately predict the performance of aggregation-enabled WLANs with fully unsaturated traffic sources.

REFERENCES

[1] *IEEE Standard for Information Technology—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput*, IEEE Standard 802.11n-2009, 2009, pp. 1–565.

[2] *IEEE 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 4: Enhancements for Very High Throughput for Operation in Bands Below 6 GHz*, IEEE Standard 802.11ac-2013, 2013, pp. 1–425.

[3] E. Perahia and R. Stacey, *Next Generation Wireless LANs: 802.11n and 802.11ac*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[4] Y. Lin and V. W. S. Wong, "WSN01-1: Frame aggregation and optimal frame size adaptation for IEEE 802.11n WLANs," in *Proc. IEEE GLOBECOM*, Nov. 2006, pp. 1–6.

[5] T. Li, Q. Ni, D. Malone, D. Leith, Y. Xiao, and T. Turletti, "Aggregation with fragment retransmission for very high-speed WLANs," *IEEE/ACM Trans. Netw.*, vol. 17, no. 2, pp. 591–604, Apr. 2009.

[6] N. Hajlaoui, I. Jabri, and M. B. Jemaa, "Analytical study of frame aggregation in error-prone channels," in *Proc. 9th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jul. 2013, pp. 237–242.

[7] S. Seytnazarov, D. G. Jeong, and W. S. Jeon, "Parallel PPDU transmission mechanism for wideband wireless LANs," *IEEE Access*, vol. 8, pp. 198714–198729, 2020.

[8] N. Hajlaoui, I. Jabri, and M. B. Jemaa, "An accurate two dimensional Markov chain model for IEEE 802.11n DCF," *Wireless Netw.*, vol. 24, no. 4, pp. 1019–1031, May 2018.

[9] I. Jabri, K. Mansour, I. Al-Oqily, and T. Ezzedine, "Enhanced characterization and modeling of A-MPDU aggregation for IEEE 802.11n WLANs," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 1, p. e4384, 2022.

[10] K.-T. Feng, Y.-Z. Huang, and J.-S. Lin, "Design of MAC-defined aggregated ARQ schemes for IEEE 802.11n networks," *Wireless Netw.*, vol. 17, no. 3, pp. 685–699, Apr. 2011.

[11] J.-S. Lin, K.-T. Feng, Y.-Z. Huang, and L.-C. Wang, "Novel design and analysis of aggregated ARQ protocols for IEEE 802.11n networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 3, pp. 556–570, Mar. 2013.

[12] S. Seytnazarov, J.-G. Choi, and Y.-T. Kim, "Enhanced mathematical modeling of aggregation-enabled WLANs with compressed Block-ACK," *IEEE Trans. Mobile Comput.*, vol. 18, no. 6, pp. 1260–1273, Jun. 2019.

[13] K. Mansour, I. Jabri, and T. Ezzedine, "Revisiting the IEEE 802.11n A-MPDU retransmission scheme," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 1097–1100, Jun. 2019.

[14] (2017). *Atheros ath9k Driver*. Accessed: Mar. 16, 2023. [Online]. Available: https://wireless.wiki.kernel.org/en/users/drivers/ath9k

[15] 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.

[16] I. Tinnirello, G. Bianchi, and Y. Xiao, "Refinements on IEEE 802.11 distributed coordination function modeling approaches," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1055–1067, Mar. 2010.

[17] K. Duffy, D. Malone, and D. J. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 715–717, Aug. 2005.

[18] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions," *IEEE/ACM Trans. Netw.*, vol. 15, no. 1, pp. 159–172, Feb. 2007.

[19] K. Kosek-Szott, "A comprehensive analysis of IEEE 802.11 DCF heterogeneous traffic sources," *Ad Hoc Netw.*, vol. 16, pp. 165–181, May 2014.

[20] R. P. Liu, G. J. Sutton, and I. B. Collings, "A new queueing model for QoS analysis of IEEE 802.11 DCF with finite buffer and load," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2664–2675, Aug. 2010.

[21] G. J. Sutton, R. P. Liu, and I. B. Collings, "Modelling IEEE 802.11 DCF heterogeneous networks with Rayleigh fading and capture," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3336–3348, Aug. 2013.

[22] S. Kuppa and G. R. Dattatreya, "Modeling and analysis of frame aggregation in unsaturated WLANs with finite buffer stations," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, Jun. 2006, pp. 967–972.

[23] B. Bellalta, "A queuing model for the non-continuous frame assembly scheme in finite buffers," in *Analytical and Stochastic Modeling Techniques and Applications*, K. Al-Begain, D. Fiems, G. Horváth, Eds. Berlin, Germany: Springer, 2009, pp. 219–233.

[24] B. Bellalta, A. Faridi, D. Staehle, J. Barcelo, A. Vinel, and M. Oliver, "Performance analysis of CSMA/CA protocols with multi-packet transmission," *Comput. Netw.*, vol. 57, no. 14, pp. 2675–2688, Oct. 2013.

[25] B. S. Kim, H. Y. Hwang, and D. K. Sung, "Effect of frame aggregation on the throughput performance of IEEE 802.11n," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2008, pp. 1740–1744.

[26] A. Abdrabou and W. Zhuang, "Service time approximation in IEEE 802.11 single-hop ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 305–313, Jan. 2008.

[27] X. Ma, X. Chen, and H. H. Refai, "Unsaturated performance of IEEE 802.11 broadcast service in vehicle-to-vehicle networks," in *Proc. IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 1957–1961.

[28] D. Xu, T. Sakurai, and H. L. Vu, "An access delay model for IEEE 802.11e EDCA," *IEEE Trans. Mobile Comput.*, vol. 8, no. 2, pp. 261–275, Feb. 2009.

[29] T. Sakurai and H. Vu, "MAC access delay of IEEE 802.11 DCF," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1702–1710, May 2007.

[30] I. Kim and Y.-T. Kim, "Realistic modeling of IEEE 802.11 WLAN considering rate adaptation and multi-rate retry," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1496–1504, Nov. 2011.

[31] P. Raptis, V. Vitsas, K. Paparrizos, P. Chatzimisios, and A. C. Boucouvalas, "Packet delay distribution of the IEEE 802.11 distributed coordination function," in *Proc. 6th IEEE Int. Symp. World Wireless Mobile Multimedia Netw.*, Jun. 2005, pp. 299–304.

[32] P. Raptis, V. Vitsas, and K. Paparrizos, "Packet delay metrics for IEEE 802.11 distributed coordination function," *Mobile Netw. Appl.*, vol. 14, no. 6, pp. 772–781, Dec. 2009.

[33] S. Seytnazarov. (2023). *802.11n/ac Wi-Fi Simulator*. [Online]. Available: https://github.com/shinnazar/802.11n-ac-Wi-Fi-simulator

SHINNAZAR SEYTNAZAROV (Member, IEEE) received the M.S. and Ph.D. degrees in information and communication engineering from Yeungnam University, South Korea, in 2014 and 2018, respectively. From 2019 to 2021, he was a Postdoctoral Research Fellow with the Department of Computer Science and Engineering, Seoul National University, South Korea. Since 2021, he has been an Assistant Professor with the Faculty of Computer Science and Engineering, Innopolis University, Russia. His current research interests include performance evaluation, resource management, and QoS in wireless networks and network security.

DONG GEUN JEONG (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, South Korea, in 1983, 1985, and 1993, respectively.

From 1986 to 1990, he was a Researcher with the Research and Development Center, DACOM, South Korea. From 1994 to 1997, he was with the Research and Development Center, Shinsegi Telecomm Inc., South Korea, where he conducted and led research on advanced cellular mobile networks. In 1997, he joined the Hankuk University of Foreign Studies, South Korea, as a Faculty Member, where he is currently a Professor with the Department of Electronics Engineering. His current research interests include resource management for wireless and mobile networks, the Internet of Things, and wireless technologies for industry.

Dr. Jeong served on the editorial board for the *Journal of Communications and Networks*, from 2002 to 2007.

WHA SOOK JEON (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer engineering from Seoul National University, Seoul, South Korea, in 1983, 1985, and 1989, respectively.

From 1989 to 1999, she was with the Department of Computer Engineering, Hansung University, Seoul. In 1999, she joined Seoul National University, as a Faculty Member, where she is currently a Professor with the School of Electrical Engineering and Computer Science. Her current research interests include resource management for wireless and mobile networks, high-speed networks, communication protocols, network performance evaluation, and the Internet of Things.

Dr. Jeon served on the editorial board for *Journal of Communications and Networks*, from 2002 to 2017.