

Research Article

Adaptive Two-Level Frame Aggregation for Fairness and Efficiency in IEEE 802.11n Wireless LANs

Minho Kim,¹ Eun-Chan Park,² and Chong-Ho Choi¹

¹*School of Electrical Engineering and Computer Science, Seoul National University, Seoul 151-742, Republic of Korea*

²*Department of Information and Communication Engineering, Dongguk University, Seoul 100-715, Republic of Korea*

Correspondence should be addressed to Eun-Chan Park; ecpark@dongguk.edu

Received 31 May 2013; Accepted 14 June 2013

Academic Editor: David Taniar

Copyright © 2015 Minho Kim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper deals with the problem of performance degradation in wireless local area networks (WLANs) based on IEEE 802.11n. When a wireless channel is shared by heterogeneous stations that have different data rates and packet sizes, each station occupies a different amount of airtime because the basic channel access mechanism of WLAN was originally designed to provide fair chance of channel access, regardless of packet size and data rate. This leads to the degradation of overall network throughput and airtime fairness among stations, which is known as performance anomaly. To resolve this problem, we firstly formulate an optimization problem for a generalized two-level frame aggregation whose objective is to maximize the achievable throughput under the constraint of airtime fairness. Then, we propose a frame size adaptation scheme that controls the number of packets in an aggregated frame. The proposed scheme is fully compatible with the IEEE 802.11 standard and works in a distributed manner, which neither modifies the channel access mechanism nor resorts to a centralized scheduling algorithm. The extensive simulation results confirm that the proposed scheme tightly regulates the airtime usage of each station to be almost the same and significantly improves the overall network throughput compared to other existing schemes.

1. Introduction

The explosive growth of mobile devices such as smart phones and tablet PCs accelerates the demand for wireless Internet access, and the wireless local area networks (WLANs) are widely deployed to provide wireless connectivity for mobile devices. To satisfy the increasing demand for higher throughput of WLANs, the IEEE 802.11n standard [1] introduces new physical (PHY) layer and medium access control (MAC) layer specifications. By using advanced PHY layer technologies such as multiple-input multiple-output (MIMO) antenna, orthogonal frequency division multiplexing (OFDM), adaptive channel coding, and channel bonding, the data rate in the PHY layer reaches up to 600 Mb/s with a 40 MHz channel bandwidth and 4×4 MIMO configuration. However, MAC-layer overheads such as the MAC header, contention time, and acknowledgement (ACK) limit the actual throughput. To reduce these overheads, the MAC layer in IEEE

802.11n introduces several enhancement mechanisms including frame aggregation, block ACK, and reverse direction.

In a typical and real WLAN environment, both the channel state and user application are quite different among stations. For example, a certain station may transmit small-size VoIP packets at a high data rate (e.g., 20-byte packets at 600 Mb/s), while another station may transmit large-size data packets at a low data rate (e.g., 1500-byte packets at 6.5 Mb/s) (hereafter, we refer to a *packet* as what MAC receives from the upper layer and a *frame* as what MAC transfers to the lower layer). In this extreme case, the channel occupation times to transmit a single packet differ by an order of magnitude even when considering the MAC/PHY headers. Therefore, when a wireless channel is shared among many heterogeneous stations of different data rates and packet sizes, it is imperative to share the channel in a fair and efficient manner. The basic channel access mechanism adopted in IEEE 802.11, that is, the distributed coordination function (DCF), was designed to provide long-term equal channel access opportunity for all

contending stations. In a multirate environment, however, the DCF can significantly degrade the overall network throughput, which is well known as *performance anomaly* [2]. This problem neutralizes many possible advantages of employing higher data rates and it is exacerbated when the difference among the data rates increases (the data rate supported in IEEE 802.11n ranges from 6.5 Mb/s to 600 Mb/s). Similarly, a performance anomaly problem also arises among stations that transmit packets of different sizes.

The performance anomaly problem can be resolved by enforcing fairness in terms of the channel access time, referred to as *airtime fairness*, or *temporal fairness* [3, 4]. There are several proposals for airtime fairness in the literature, and most of them adjust the channel access probability according to the data rate and packet size [5–9]. In [5], a transmitter adjusts its contention window size to be inversely proportional to the data rate under the assumption that the packet size is similar for all stations. The studies in [6–9] extend the basic idea in [5]. The authors in [9] provide the optimal channel access probability for each station to achieve airtime fairness and to maximize the aggregate throughput. The transmission opportunity (TXOP) mechanism in IEEE 802.11e [10] can be used to achieve airtime fairness. The amount of airtime consumed by each station can be brought to a comparable level by transmitting multiple back-to-back data packets within the same TXOP duration.

On the other hand, the IEEE 802.11n standard [1] introduces two types of frame aggregation schemes, an aggregated MAC-level service data unit (A-MSDU) and an aggregated MAC-level protocol data unit (A-MPDU), to increase the throughput by decreasing the MAC-layer overheads. The standard only specifies the frame format and maximum size of the aggregated frame; however, the optimal frame size and operation mode are not standardized and are left undetermined. Recently, several analytical studies have investigated the effect of A-MPDU and/or A-MSDU frame sizes on the achievable throughput and their optimal sizes to improve throughput [11–15]. We can consider a naive approach of combining frame aggregation with the TXOP mechanism to enforce airtime fairness. By allowing frame aggregation within the TXOP duration that is set to be the same for all stations, each station can share a fair amount of airtime, which is close to the TXOP duration as much as possible. This approach seems to be effective to enhance fairness; however, it still leaves several open problems. Firstly, the optimal value of TXOP duration has to be determined to maximize throughput under the constraint of fairness. Secondly, it is necessary to appropriately determine the type of aggregation schemes, that is, A-MSDU or A-MPDU, and to change it dynamically depending on the channel state and MAC overhead, for improving both fairness and throughput. Lastly, the limit on the number of subframes in a frame should be considered; otherwise some portion of TXOP duration would be unused and wasted, which leads to the degradation of fairness and throughput.

In this paper, we study the problem of sharing a wireless channel among heterogeneous stations that have different data rates and packet sizes. The objective of this paper is two-fold: (a) to strictly enforce airtime fairness regardless of the

data rate and packet size and (b) to maximize the network throughput by considering the channel error rate, data rate, and MAC overhead. For this purpose, we propose a dynamic frame size control scheme by employing a generalized two-level frame aggregation scheme that combines the A-MSDU and A-MPDU schemes. Compared with other existing studies, this study makes the following contributions.

- (i) In contrast to the approaches for airtime fairness in [5–9], the proposed mechanism is novel and effective; (a) it utilizes the frame aggregation scheme without modifying the basic channel access mechanism to fully comply with the IEEE 802.11 standard and (b) it introduces the notion of target airtime and controls the actual airtime tightly close to the target value.
- (ii) Unlike the frame aggregation schemes to increase throughput in [11–15], this work proposes a generalized solution; it (a) aims to improve both efficiency and fairness and (b) employs two-level aggregation to take advantage of both A-MSDU and A-MPDU.
- (iii) In determining the optimal size for frame aggregation, most current analytical models do not consider the effect of link adaptation and the optimal size cannot be represented in a closed form. In contrast, the analysis model in this study incorporates the link adaptation algorithm employed in 802.11n. In addition, it formulates the problem as an optimization problem and provides a closed-form solution for optimal frame aggregation.

Extensive simulation results confirm that the proposed scheme significantly outperforms other existing schemes in terms of the aggregate throughput and channel access delay while maintaining a high level of airtime fairness.

The rest of the paper is organized as follows. In Section 2, we describe the preliminaries of this paper. We formulate the problem as an optimization problem in Section 3 and provide an effective method to adjust frame size in Section 4. In Section 5, we validate the proposed scheme via simulations and compare its performance with other existing schemes. Finally, we conclude the paper in Section 6.

2. Preliminaries

In this section, we first present a motivating example that demonstrates the performance anomaly of the IEEE 802.11 DCF. We then introduce frame aggregation schemes in IEEE 802.11n, which will be used as basic tools for the proposed solution to resolve the performance anomaly problem.

2.1. Performance Anomaly in a Heterogeneous Network Environment. The IEEE 802.11 DCF was basically designed to provide a fair channel access opportunity for all contending stations, regardless of the data rate and packet size. Therefore, it results in a performance anomaly in a heterogeneous network environment where the stations have different packet sizes and data rates [2]. To illustrate this problem, we performed a preliminary simulation under a simple scenario where four stations (denoted as STA1~STA4) transmit data

TABLE 1: Performance anomaly of IEEE 802.11 DCF due to differences in packet size and data rate.

Configuration	STA1	STA2	STA3	STA4
Packet size (bytes)	250	1000	250	1000
Data rate (Mb/s)	13	13	65	65
Unit transmission time (ms)	0.312	0.774	0.150	0.242
Attempt probability	0.093	0.093	0.090	0.091
Throughput (Mb/s)	1.061	4.109	1.073	4.197
Airtime ratio	0.166	0.397	0.080	0.127

packets of different sizes at different data rates as described in Table 1. Here, we assume an ideal error-free channel. The *unit transmission time* in Table 1 is calculated as the time required to transmit a single data frame according to the IEEE 802.11n specification. It does not take into account the backoff time but includes the time for the PHY/MAC headers, short interframe space (SIFS), and ACK transmission. Note that the ACK is transmitted at the lowest data rate (with the most robust modulation and coding scheme) to enhance the transmission reliability. Throughout the paper, the *airtime ratio* of a station is calculated by accumulating all unit transmission times of successful frame transmissions for the entire simulation time and then dividing it by the simulation time. It is noted that the sum of the airtime ratios is less than one because the airtime does not account for the backoff time and the time consumed for unsuccessful transmissions. The *attempt probability* is measured as the probability that a station makes an attempt for frame transmission in an idle slot. From the results listed in Table 1, we observe the following.

- (i) DCF assures almost an equal chance of channel access for all contending stations, regardless of their data rate and packet size; that is, the difference in the attempt probability among stations is negligible. This implies that the number of packets transmitted by each station is almost the same, leading to *throughput fairness* among stations that have the same packet size. For example, the throughput of STA1 is comparable with that of STA3, even though the data rate of STA3 is five times higher than that of STA1.
- (ii) Performance anomaly also occurs due to the difference in the packet size. When stations have the same data rate, the throughput of a station is almost proportional to its packet size; for example, the throughput of STA2 is about four times higher than that of STA1.
- (iii) The airtime is mostly consumed by the station transmitting large packets at the low data rate (STA2); only a small fraction of airtime is consumed by the station transmitting small packets at the high data rate (STA3). Despite the high data rate, its advantage cannot be fully utilized under the DCF.

Consequently, the performance anomaly not only causes unfairness in airtime but also degrades the overall network throughput. Providing airtime fairness among stations,

which is shown to be equivalent to the proportional fairness [7, 16], can be an effective solution to this problem. It can be achieved by adjusting the attempt probability according to the packet size or data rate [5–9], as already addressed. This problem can also be mitigated by making use of the TXOP mechanism. In contrast to other existing approaches, we provide a novel approach to achieve airtime fairness based on the frame aggregation scheme introduced in IEEE 802.11n.

2.2. Frame Aggregation in IEEE 802.11n. The IEEE 802.11n standard introduces two types of frame aggregation schemes, A-MSDU and A-MPDU, to enhance the MAC efficiency by reducing MAC-layer overheads. The A-MSDU scheme aggregates multiple MSDUs into a single MPDU, whereas the A-MPDU scheme allows multiple MPDUs to be aggregated into a single PHY-layer service data unit (PSDU). Compared with A-MPDU, A-MSDU has less header overhead because it can transmit multiple packets with a single PHY/MAC header, whereas A-MPDU transmits them with separate MAC headers under a common PHY header. If the channel is error-free, A-MSDU outperforms A-MPDU and the throughput in both cases increases as the number of aggregated packets increases. However, A-MPDU is more effective than A-MSDU when the channel is error-prone. If some of the transmitted MPDU subframes are corrupted due to channel error, A-MPDU can selectively retransmit only the corrupted subframes, which is possible thanks to block ACK and the frame check sequence (FCS) field in the individual MAC header for each subframe. In contrast, A-MSDU has to unnecessarily retransmit the whole subframes even though only a subframe is corrupted. Therefore, when the bit error rate (BER) of a channel is not negligible, A-MPDU outperforms A-MSDU. Recently, a hybrid frame aggregation scheme known as *two-level aggregation*, which combines A-MSDU and A-MPDU, has been introduced [17, 18], and its performance has been evaluated via simulations. However, the effect of this scheme on throughput or fairness is not analyzed in depth. Figure 1 shows the frame structure of the two-level aggregation scheme. Let us denote n_1 and n_2 as the numbers of MSDU and MPDU subframes, respectively, in the two-level aggregation scheme. We can take advantage of both the A-MSDU and A-MPDU schemes by optimally determining the values of n_1 and n_2 , which is the starting point of this work.

Now, we generalize several aggregation schemes in terms of n_1 and n_2 as follows.

(1) *A-MSDU* ($n_1 > 1, n_2 = 1$). A PSDU consists of n_1 MSDU subframes each of which contains a packet.

(2) *A-MPDU* ($n_1 = 1, n_2 > 1$). A PSDU consists of n_2 MPDU subframes.

(3) *Two-Level Aggregation* ($n_1 > 1, n_2 > 1$). A PSDU consists of n_2 MPDU subframes, and each MPDU subframe consists of n_1 MSDU subframes.

The total numbers of packets transmitted in a single PSDU are n_1, n_2 , and $n_1 \cdot n_2$ in the cases of A-MSDU, A-MPDU, and two-level aggregation, respectively. Let us focus on the

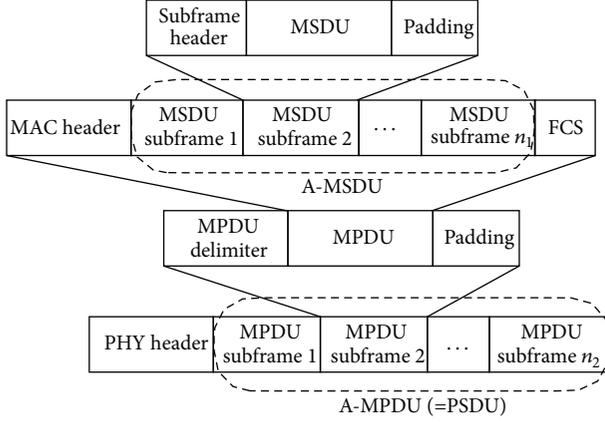


FIGURE 1: Frame structure of two-level frame aggregation scheme that combines A-MSDU with A-MPDU.

header overhead incurred during the aggregation process. Table 2 lists several notations of the PHY/MAC headers and timing overheads considered in this study. We define $L_{\text{oh}}(n_1, n_2)$ as the overhead to construct a single PSDU, which can be represented as

$$L_{\text{oh}}(n_1, n_2) = \begin{cases} n_1\alpha + \beta, & \text{for A-MSDU,} \\ n_2(\beta + \gamma), & \text{for A-MPDU,} \\ n_1n_2\alpha + n_2(\beta + \gamma), & \text{for Two-level,} \end{cases} \quad (1)$$

where $\alpha = L_{\text{sub}} + L_{\text{pad}}$, $\beta = L_{\text{mac}} + L_{\text{FCS}}$, and $\gamma = L_{\text{deli}} + L_{\text{pad}}$. It needs to be noted that there exist several limits on n_1 and n_2 . According to the IEEE 802.11n standard, the maximum lengths of A-MSDU and A-MPDU are 3839 and 65536 bytes, respectively. In addition, because of the length of the block ACK information field, at most 64 subframes can be aggregated in an A-MPDU frame. It is a challenging problem to determine the optimal values of n_1 and n_2 for maximizing throughput and assuring fairness, because they depend on various factors such as the BER, data rate, packet size, and header overhead. In the subsequent sections, we present an objective function for the optimal operation of frame aggregation by considering airtime fairness and achievable throughput. We then provide a simple and effective method to find the optimal solution of the objective function.

3. Problem Statement

In this section, we first describe how the data rate is adjusted by the fast link adaptation (FLA) algorithm in IEEE 802.11n, which is closely related to how the optimal size of frame aggregation is determined. Then, we formulate an optimization problem to derive the optimal frame size.

3.1. Fast Link Adaptation in IEEE 802.11n. The FLA algorithm is usually employed in IEEE 802.11n [1]. A transmitter requests for an appropriate modulation and coding scheme (MCS) to a receiver via the MCS request subfield. At the request, the receiver determines an appropriate MCS by

TABLE 2: Notations and values of PHY/MAC headers and timing parameters.

Notation	Description	Value
L_{sub}	Subframe header size (for A-MSDU)	14 bytes
L_{mac}	MAC header size	34 bytes
L_{pad}	Padding size	0~3 bytes
L_{deli}	MPDU delimiter size (for A-MPDU)	4 bytes
L_{FCS}	Frame check sequence (FCS) size	4 bytes
T_{phy}	PHY header transmission time	32 μs
T_{SIFS}	Short interframe space (SIFS)	16 μs
T_{DIFS}	Distributed interframe space (DIFS)	34 μs
T_{slot}	Slot time	9 μs

considering the throughput and/or target frame error rate and notifies it via the MCS feedback subfield. It is difficult to estimate the block error rate (BLER) of an L_B -byte block in a MIMO-OFDM system because the signal to noise ratio (SNR) levels among the subcarriers and/or spatial streams may differ due to frequency selectivity [19]. Therefore, a simple but accurate mapping method is needed to estimate the BLER as a function of the SNR for each subcarrier and spatial stream. The exponential effective SNR mapping (EESM) method [20] was proposed for this purpose and can be used for FLA in IEEE 802.11n [21, 22]. It translates an instantaneous set of postprocessing SNRs into a scalar link quality metric known as the effective SNR. This mapping procedure should be calibrated according to each MCS to estimate the BLER of an L_B -byte block by considering the relationship between the SNR and BLER for each MCS in an additive white Gaussian noise (AWGN) channel. Let us define γ_m^{eff} as the effective SNR with an MCS index of m , which can be calculated as in [21]. Then, the effective error rate of a PSDU consisting of several L_p -byte packets, P_e , can be expressed as

$$P_e(n_1, n_2) = 1 - \{1 - \text{BLER}(\gamma_m^{\text{eff}})\}^{L(n_1, n_2)/L_B}, \quad (2)$$

where $L(n_1, n_2)$ is

$$L(n_1, n_2) = \begin{cases} n_1(\alpha + L_p) + \beta, & \text{for A-MSDU,} \\ L_p + \beta, & \text{for A-MPDU,} \\ n_1(\alpha + L_p) + \beta, & \text{for Two-level.} \end{cases} \quad (3)$$

Here, $\text{BLER}(\gamma_m^{\text{eff}})$ in (2) is obtained by the SNR-BLER curve of the L_B -byte block in the AWGN channel. Hereafter, we assume that the receiver estimates $P_e(n_1, n_2)$ for each MCS according to (2) and then selects the MCS that gives the highest data rate among the MCSs that have the value of $P_e(n_1, n_2)$ below a target error rate η . From (2) and (3), we note the following points, which agree with intuition and provide insights into determining the optimal values of n_1 and n_2 .

- (i) For A-MSDU, $P_e(n_1, n_2)$ increases as n_1 increases; that is, A-MSDU becomes vulnerable to channel error as its size increases. There is a trade-off between the header overhead and the error rate. Although a large

value of n_1 contributes to increased throughput by reducing the header overhead, a more robust MCS is required to maintain an acceptable error rate, which results in the decrease of throughput.

- (ii) For A-MPDU, $P_e(n_1, n_2)$ is independent of n_2 ; that is, the error rate is independent of the number of aggregated MPDU subframes because of the selective repeat mechanism. This comes at the cost of a larger header overhead compared to A-MSDU (see (1)).

3.2. Optimization Problem for Two-Level Frame Aggregation.

From the viewpoint of airtime fairness, the basic concept of the proposed frame size adaptation scheme is to adjust the values of n_1 and n_2 such that the per-station airtime is regulated around the target reference value, T_{ref} . One of the most important issues is how to set T_{ref} properly. We discuss this later in Section 4.3.

Another objective of the proposed scheme is to maximize throughput. Now, let us derive the throughput that can be obtained with the two-level aggregation scheme under the constraint of airtime fairness. We only focus on the generalized two-level aggregation scheme because the single-level A-MSDU and A-MPDU schemes can be approximately considered as special cases of two-level aggregation. We emphasize that our objective is to find the optimal values of n_1 and n_2 , and not to derive an accurate analysis model of aggregate throughput. For this purpose, we only focus on the case where a station gets a chance to access the channel after competing with other stations and tries to maximize its own throughput by adjusting the values of n_1 and n_2 . In the proposed scheme, the optimal values of n_1 and n_2 are determined for each station in a distributed manner, depending on its data rate and packet size. According to the result in [11], it is shown that the optimal size of an A-MSDU maximizing the aggregate throughput is mainly affected by the BER but hardly affected by the number of contending stations. It is noteworthy that the proposed scheme is completely decoupled from the contention window control that deals with collisions among stations and that the standard binary exponential backoff mechanism is used to resolve collisions. We define $T_{\text{data}}(n_1, n_2)$ as the time required to transmit a single aggregate PHY-layer protocol data unit (PPDU) consisting of n_2 MPDU subframes and n_1 MSDU subframes per MPDU subframe. Then, $T_{\text{data}}(n_1, n_2)$ is represented as

$$T_{\text{data}}(n_1, n_2) = T_{\text{phy}} + \frac{8(n_1 n_2 L_p + L_{\text{oh}}(n_1, n_2))}{R_m}, \quad (4)$$

where T_{phy} is the time required to transmit the PHY header including the preamble and R_m is the data rate with MCS index m , which is selected by the FLA algorithm. Under the condition that there is no transmission failure due to channel error or collision, the maximum throughput $S(n_1, n_2)$ is expressed as

$$S(n_1, n_2) = \frac{8n_1 n_2 L_p}{T_{\text{data}}(n_1, n_2) + T_{\text{oh}}}, \quad (5)$$

where $T_{\text{oh}} = T_{\text{DIFS}} + T_{\text{BO}} + T_{\text{SIFS}} + T_{\text{back}}$. Here, T_{BO} and T_{back} are the average backoff time and block ACK transmission time, respectively. The average backoff time is approximated as $T_{\text{BO}} = T_{\text{slot}}(\text{CW}_{\text{min}}/2)$, where CW_{min} is the minimum contention window size. The maximum throughput depends mainly on the collision probability, but little on the values of n_1 and n_2 .

To find the appropriate values for n_1 and n_2 , we formulate the optimization problem as follows:

$$\begin{aligned} (n_1^*, n_2^*) &= \arg \max_{(n_1, n_2) \in \mathbb{N}^2} S(n_1, n_2) (1 - P_e(n_1, n_2)), \\ \text{subject to } & |T_{\text{ref}} - T_{\text{data}}(n_1, n_2)| < T_\epsilon, \\ & n_1 L_p + L_{\text{oh}}(n_1, 1) \leq L_{\text{max}}^{\text{msdu}}, \\ & n_1 n_2 L_p + L_{\text{oh}}(n_1, n_2) \leq L_{\text{max}}^{\text{mpdu}}, \\ & n_2 \leq N_{\text{max}}^{\text{mpdu}}. \end{aligned} \quad (6)$$

Here, T_ϵ is introduced to allow some tolerance between $T_{\text{data}}(n_1, n_2)$ and T_{ref} because it is difficult to make $T_{\text{data}}(n_1, n_2)$ exactly equal to T_{ref} with integer values of n_1 and n_2 . Also, in (6), $L_{\text{max}}^{\text{msdu}}$ (=3839 bytes) and $L_{\text{max}}^{\text{mpdu}}$ (=65536 bytes) are the maximum sizes of A-MSDU and A-MPDU frames, respectively, and $N_{\text{max}}^{\text{mpdu}}$ (=64) is the maximum number of MPDU subframes. These three constraints are imposed by the IEEE 802.11n standard, whereas the first one stems from airtime fairness. The problem in (6) is an integer programming problem, which is more difficult to solve than the case when the variables are real numbers. Each transmitter has to solve this optimization problem every time before constructing an aggregated frame, which may be too time-consuming to solve the problem in real-time. In addition, the transmitter cannot accurately estimate the frame error rate $P_e(n_1, n_2)$ for the currently transmitting frame. These difficulties demand a more efficient and effective method of finding the optimal solution, (n_1^*, n_2^*) , which will be provided in the subsequent section.

4. Adaptive Frame Aggregation for Fairness and Efficiency

In this section, we provide a simple and effective method to determine n_1^* and n_2^* . We first transform the optimization problem into an unconstrained, single-variable optimization problem and then provide a heuristic solution, which will be used to adjust the frame size.

4.1. Unconstrained Single-Variable Optimization Problem. To simplify the optimization problem, we replace n_1 and n_2 with x_1 and x_2 , respectively, which are positive real values without upper bounds so that $T_{\text{data}}(x_1, x_2)$ can be made exactly equal to T_{ref} . Then, we can release all the constraints and replace the objective function in (6) with

$$(x_1^*, x_2^*) = \arg \max_{(x_1, x_2) \in \mathbb{R}^2} K x_1 x_2 (1 - P_e(x_1, x_2)), \quad (7)$$

where constant K is

$$K = \frac{8L_p}{T_{\text{ref}} + T_{\text{oh}}}, \quad (8)$$

because $T_{\text{data}}(x_1, x_2) = T_{\text{ref}}$. We further simplify (7) by expressing x_2 in terms of x_1 . By solving the equation $T_{\text{data}}(x_1, x_2) = T_{\text{ref}}$ with respect to x_1 and x_2 , x_2 becomes

$$x_2 = \frac{C}{Ax_1 + B}, \quad (9)$$

where $A = \alpha + L_p$, $B = \beta + \gamma$, and $C = R_m(T_{\text{ref}} - T_{\text{phy}})/8$. It is noted that x_2 is mapped to a unique x_1 , as shown in (9).

Next, we consider the frame error rate. As the transmitter is only informed of the appropriate MCS index m for the previously transmitted frame, it cannot readily predict the frame error rate for the current frame. Let us define (x'_1, x'_2) as the numbers of A-MSDU and A-MPDU subframes used in the aggregation for the previously transmitted frame and define $P_e(x'_1, x'_2)$ as its frame error rate. To simplify the problem, we make the following assumptions: (a) the frame transmission time is sufficiently shorter than the channel coherence time so that the channel condition does not change significantly between two consecutive frames; (b) the FLA algorithm works properly so that the error rate is slightly lower than the target error rate η ($\ll 1$). A station may suffer from deep fading, and thus the frame error rate may exceed the target value even though the station uses the most robust MCS. This situation is out of control because any combination of (n_1, n_2) cannot guarantee the success of transmission. With these assumptions, we approximate $P_e(x'_1, x'_2) \approx \eta$ and relate $P_e(x_1, x_2)$ with $P_e(x'_1, x'_2)$ as follows:

$$\begin{aligned} P_e(x_1, x_2) &= 1 - \left\{1 - P_e(x'_1, x'_2)\right\}^{L(x_1, x_2)/L(x'_1, x'_2)} \\ &\approx \frac{L(x_1, x_2)}{L(x'_1, x'_2)} \eta, \end{aligned} \quad (10)$$

because $\eta \ll 1$. Then, $(1 - P_e(x_1, x_2))$ is represented as

$$\begin{aligned} 1 - P_e(x_1, x_2) &= 1 - \frac{x_1(\alpha + L_p) + \beta}{L(x'_1, x'_2)} \eta \\ &= D - A\eta' x_1, \end{aligned} \quad (11)$$

where $\eta' = \eta/L(x'_1, x'_2)$ and $D = 1 - \beta\eta'$.

Through the above procedure, we can convert the original problem in (6) to the following unconstrained single-variable optimization problem:

$$x_1^* = \arg \max_{x_1 \in \mathbb{R}} K C f(x_1), \quad (12)$$

where

$$f(x_1) = \frac{Dx_1 - A\eta' x_1^2}{Ax_1 + B}. \quad (13)$$

It is readily shown that $f'(x_1)$ is a strictly decreasing function of x_1 ; that is, $f(x_1)$ is concave. This confirms that a unique x_1^*

exists. We can find x_1^* from the derivative of $f(x_1)$ and x_2^* from (9) as follows:

$$x_1^* = \frac{B}{A} \left(\sqrt{1 + \frac{D}{B\eta'}} - 1 \right), \quad x_2^* = \frac{C}{Ax_1^* + B}. \quad (14)$$

4.2. Heuristic Solution for Frame Size Adaptation. Based on the results of x_1^* and x_2^* given in (14), we find the feasible set of (n_1, n_2) that are positive integer values. We assume that a station has a sufficient number of packets to aggregate. We first determine n_1 as the integer closest to n_1^* and then determine n_2 such that $T_{\text{data}}(n_1, n_2)$ becomes as close as possible to T_{ref} ; that is,

$$n_1 = \lceil x_1^* \rceil, \quad n_2 = \left\lceil \frac{C}{An_1 + B} \right\rceil, \quad (15)$$

where $\lceil x \rceil$ is the integer closest to x . In (15), n_2 is determined by using (9) instead of $\lceil x_2^* \rceil$ to minimize the difference between $T_{\text{data}}(n_1, n_2)$ and T_{ref} , which is due to the rounding error. However, the set of (n_1, n_2) obtained in this way cannot completely be free of rounding errors.

To compensate for the deviation of $T_{\text{data}}(n_1, n_2)$ from T_{ref} , we alternately use two sets of (n_1, n_2) , denoted as (n_1^u, n_2^u) and (n_1^l, n_2^l) , each of which makes the actual value of T_{data} comparable to T_{ref} but slightly larger and smaller than T_{ref} , respectively. We obtain these two sets as follows. If the set of (n_1, n_2) obtained from (15) results in $T_{\text{data}}(n_1, n_2) > T_{\text{ref}}$, then we set $(n_1^u, n_2^u) = (n_1, n_2)$. To find (n_1^l, n_2^l) , we increase n_1^l by one from n_1^u while recalculating n_2^l as

$$n_1^l = n_1^u + 1, \quad n_2^l = \left\lceil \frac{C}{An_1^l + B} \right\rceil. \quad (16)$$

Recall that the header overhead is smaller for the MSDU subframe than for the MPDU subframe. These changes in n_1 and n_2 (i.e., the increase in n_1 and the decrease in n_2) reduce the total header overhead and decrease $T_{\text{data}}(n_1^l, n_2^l)$. If $T_{\text{data}}(n_1^l, n_2^l)$ remains larger than T_{ref} even with the set in (16), we keep increasing n_1^l by one at a time until $T_{\text{data}}(n_1^l, n_2^l)$ becomes less than T_{ref} . Conversely, if the set of (n_1, n_2) obtained from (15) makes $T_{\text{data}}(n_1, n_2)$ smaller than T_{ref} , we set $(n_1^l, n_2^l) = (n_1, n_2)$. In a similar way, we obtain (n_1^u, n_2^u) by decreasing n_1^u by one from n_1^l and recalculating n_2^u as

$$n_1^u = n_1^l - 1, \quad n_2^u = \left\lceil \frac{C}{An_1^u + B} \right\rceil. \quad (17)$$

Once a transmitter selects (n_1^u, n_2^u) and (n_1^l, n_2^l) , it performs frame aggregation with either (n_1^u, n_2^u) or (n_1^l, n_2^l) . Here, we introduce a weighting factor w ($0 < w < 1$) to alternate between these two sets. When a transmitter gets a chance to access the channel (i.e., the backoff counter reaches zero), it generates a random number X uniformly distributed between zero and one. If X is less than w , the transmitter performs frame aggregation with $(n_1, n_2) = (n_1^u, n_2^u)$. Otherwise, it performs frame aggregation with

$(n_1, n_2) = (n_1^l, n_2^l)$. To make the average value of T_{data} as close as possible to T_{ref} , we set the value of w such that

$$T_{\text{ref}} = wT_{\text{data}}(n_1^u, n_2^u) + (1-w)T_{\text{data}}(n_1^l, n_2^l); \quad (18)$$

that is,

$$w = \frac{T_{\text{ref}} - T_{\text{data}}(n_1^l, n_2^l)}{T_{\text{data}}(n_1^u, n_2^u) - T_{\text{data}}(n_1^l, n_2^l)}. \quad (19)$$

To validate the effectiveness of this method, we compared the results obtained by the proposed and total enumeration methods via simulations. In the total enumeration method, (n_1^u, n_2^u) and (n_1^l, n_2^l) were obtained by solving the original optimization problem in (6) under the assumption that the transmitter knows the actual value of $P_e(n_1, n_2)$. Note that (n_1^u, n_2^u) and (n_1^l, n_2^l) are the optimal values of (n_1^*, n_2^*) obtained by solving (6) with the first constraint $(T_{\text{data}}(n_1^u, n_2^u) - T_{\text{ref}}) < T_\epsilon$ and $(T_{\text{ref}} - T_{\text{data}}(n_1^l, n_2^l)) < T_\epsilon$, respectively. We first obtained the candidate sets of (n_1, n_2) that satisfy the four constraints in (6). Here, T_ϵ was initially set to 0.1 ms. If there was no candidate set, we relaxed the constraint on T_ϵ by doubling its value until at least one set of (n_1, n_2) satisfied the constraints. We then calculated the objective function for all candidate sets of (n_1, n_2) and determined the optimal set that maximizes the objective function. This method gives an ideal solution that can be used to evaluate the effectiveness of the proposed method.

Table 3 lists the sets of (n_1^u, n_2^u) and (n_1^l, n_2^l) obtained by these two methods, along with T_{data} and throughput (indicated as *thput*). We set the packet size L_p as 500 bytes and T_{ref} as 3 ms. The guideline to set T_{ref} will be given in Section 4.3 and its effect on throughput and fairness will be investigated in Section 5.4. The results in Table 3 show that particularly when the data rate is low, (n_1^u, n_2^u) and (n_1^l, n_2^l) obtained by the proposed method somewhat differ from those obtained by the total enumeration method, and the difference between T_{data} and T_{ref} is greater with the proposed method than with the total enumeration method. The reason for this difference is that when the data rate is low, the number of candidate sets of (n_1, n_2) is small, and thus the rounding error and approximation error become relatively large. In some low data rate cases, however, the throughput is larger for the proposed method than for the total enumeration method, at the expense of a larger deviation of T_{data} from T_{ref} . Recall that such a deviation can be compensated for by alternately adopting (n_1^u, n_2^u) and (n_1^l, n_2^l) when constructing aggregated frames; this will be confirmed via simulations in the next section. Despite these differences between the proposed and total enumeration methods, the throughput difference between the two methods is negligible for most cases. Therefore, we conclude that the proposed method is suitable for adjusting frame size, as it yields performance comparable to that of the total enumeration method with the reduced complexity.

4.3. Comments on Several Issues. We discuss the following issues related to the proposed frame size adaptation scheme.

4.3.1. Guidelines to Set T_{ref} . When assuring airtime fairness, T_{ref} is used as the target reference value. If this value is too small, only a few packets can be aggregated. This decreases the possibility of reducing the header and timing overheads, which results in the loss of opportunity to enhance the throughput. However, a large value of T_{ref} may have several drawbacks. We need to carefully consider the following points in setting T_{ref} .

(i) *Link Adaptation and Channel Coherence Time.* Let us consider the MCS feedback delay, T_{MFD} , which is defined as the time interval between the start of two consecutive data frames for a station. As the number of stations in the network increases, T_{MFD} necessarily increases because the DCF gives fair chance to access the channel for all stations. When T_{MFD} exceeds the channel coherence time, the link adaptation algorithm may work poorly because the currently estimated channel condition cannot reflect the channel condition for the next transmission. Consequently, the increase of T_{MFD} results in the increase of $P_e(n_1, n_2)$, leading to the degradation of the throughput. In the proposed scheme, not only the number of stations but also the value of T_{ref} affects T_{MFD} . Therefore, it is necessary to appropriately set T_{ref} by considering the channel coherence time (it is possible that a receiver may estimate the channel coherence time and report it to a transmitter periodically [23]; however, this is beyond the scope of our study; we consider that the coherence time of a typical WLAN channel is on the order of a few tens of milliseconds [24]), so that the channel condition is kept nearly constant during T_{MFD} .

(ii) *Delay and Short-Term Fairness.* The value of T_{ref} affects short-term fairness, as well as delays in channel access, queuing, and retransmission. As T_{ref} increases, a station occupies the shared channel for a longer time, during which other stations have to defer channel access. Thus, the increase of T_{ref} increases the channel access delay between two consecutive aggregated frames as well as the queuing delay and jitter, which are undesirable for real-time service or streaming service. Moreover, the increase of T_{ref} also increases the retransmission delay for corrupted subframes because they cannot be retransmitted until the station occupies the channel again. All of these effects contribute to the degradation of short-term fairness among stations.

(iii) *Feasibility of Frame Aggregation.* The value of T_{ref} is related with the feasibility of n_1 and n_2 . If T_{ref} is large, many packets need to be aggregated to assure airtime fairness. However, because of limits on n_1 and n_2 , excessively large values of T_{ref} make n_1 and n_2 fail to comply with this constraint.

Taking all of these points into consideration, we consider that the appropriate range of T_{ref} is on the order of several milliseconds, which agrees well with the default value of TXOP in IEEE 802.11e.

4.3.2. Compliance with the IEEE 802.11 Standard. The values of n_1 and n_2 obtained from (15) should satisfy the constraints of $L_{\text{max}}^{\text{msdu}}$, $L_{\text{max}}^{\text{mpdu}}$, and $N_{\text{max}}^{\text{mpdu}}$ imposed by the IEEE 802.11n

TABLE 3: Comparison between proposed method and total enumeration method in terms of upper and lower sets of (n_1, n_2) , frame transmission time, and throughput for various data rates ($L_p = 500$ bytes and $T_{\text{ref}} = 3$ ms).

(a) Upper set of (n_1, n_2)						
Data rate (Mb/s)	(n_1^u, n_2^u)	Proposed		(n_1^u, n_2^u)	Total enumeration	
		T_{data} (ms)	Throughput (Mb/s)		T_{data} (ms)	Throughput (Mb/s)
6.5	(3, 2)	3.97	5.83	(5, 1)	3.28	5.83
13	(2, 5)	3.36	11.57	(1, 9)	3.16	11.20
26	(2, 9)	3.03	23.10	(2, 9)	3.03	23.20
39	(2, 14)	3.14	34.68	(4, 7)	3.07	35.04
52	(2, 18)	3.03	46.21	(2, 18)	3.03	46.39
65	(2, 23)	3.10	57.78	(2, 23)	3.10	58.01
78	(2, 27)	3.03	69.31	(4, 14)	3.07	70.07
104	(2, 36)	3.03	92.41	(2, 36)	3.03	93.11
117	(2, 41)	3.07	103.99	(3, 28)	3.10	105.16
130	(2, 45)	3.03	115.52	(3, 31)	3.09	116.84

(b) Lower set of (n_1, n_2)						
Data rate (Mb/s)	(n_1^l, n_2^l)	Proposed		(n_1^l, n_2^l)	Total enumeration	
		T_{data} (ms)	Throughput (Mb/s)		T_{data} (ms)	Throughput (Mb/s)
6.5	(4, 1)	2.64	5.78	(1, 4)	2.82	5.59
13	(3, 3)	2.99	11.61	(3, 3)	2.99	11.68
26	(3, 6)	2.99	23.21	(3, 6)	2.99	23.35
39	(3, 9)	2.99	34.82	(3, 9)	2.99	35.03
52	(3, 12)	2.99	46.43	(3, 12)	2.99	46.71
65	(3, 15)	2.99	58.03	(3, 15)	2.99	58.39
78	(3, 18)	2.99	69.64	(3, 18)	2.99	70.06
104	(3, 24)	2.99	92.85	(3, 24)	2.99	93.42
117	(3, 27)	2.99	104.46	(3, 27)	2.99	105.10
130	(3, 30)	2.99	116.07	(3, 30)	2.99	116.77

standard. Let us define the maximum numbers of n_1 and n_2 that comply with these limits as N_1^{max} and N_2^{max} , respectively. We need to deal with the case where $x_1^* > N_1^{\text{max}}$ and/or $x_2^* > N_2^{\text{max}}$.

Case 1 ($x_1^* > N_1^{\text{max}}$ and $x_2^* < N_2^{\text{max}}$). In this case, we set

$$n_1 = N_1^{\text{max}}, \quad n_2 = \left\lceil \frac{C}{AN_1^{\text{max}} + B} \right\rceil. \quad (20)$$

This configuration may incur more header overhead due to the decrease in n_1 and increase in n_2 . On the other hand, the decrease in n_1 contributes to reduce the frame error rate.

Case 2 ($x_2^* > N_2^{\text{max}}$ and $x_1^* < N_1^{\text{max}}$). As in the first case, we set

$$n_2 = N_2^{\text{max}}, \quad n_1 = \left\lceil \frac{C}{AN_2^{\text{max}}} - \frac{B}{A} \right\rceil. \quad (21)$$

In this case, n_2 is decreased from n_2^* while n_1 is increased from n_1^* . Contrary to the first case, the header overhead decreases whereas the error rate increases.

Case 3 ($x_1^* > N_2^{\text{max}}$ and $x_2^* > N_2^{\text{max}}$). In this case, there is no choice but to set $n_1 = N_1^{\text{max}}$ and $n_2 = N_2^{\text{max}}$.

These three cases mainly occur when either T_{ref} is very large or the packet size is small but the data rate is high. As long as we carefully determine T_{ref} according to the guidelines above, we can avoid such cases.

4.3.3. Dealing with Unsaturated Cases. Thus far, we have assumed that each station has a sufficient number of packets in the transmission queue so that it can arbitrarily control the number of packets to be aggregated. However, we need to consider the case of unsaturated traffic condition where the number of backlogged packets for frame aggregation (denoted as N_{pkt}) is insufficient. The proposed scheme can easily cope with this case. If $N_{\text{pkt}} < x_1^* \cdot x_2^*$, the station performs frame aggregation with n_1 and n_2 such that

$$n_1 = \min(N_{\text{pkt}}, \lceil x_1^* \rceil), \quad n_2 = \left\lceil \frac{N_{\text{pkt}}}{n_1} \right\rceil, \quad (22)$$

where $\lceil x \rceil$ is the largest integer that does not exceed x . This can avoid unnecessary delay resulting from the waiting

time of packet arrivals for optimal frame aggregation, but it makes T_{data} less than T_{ref} . The unused portion of airtime by unsaturated stations can be well used by other saturated stations. Therefore, this policy does not debase the overall network efficiency and maintains airtime fairness among the saturated stations.

5. Performance Evaluation

In this section, we evaluate the performance of the proposed frame size adaptation scheme via simulations and compare its performance with several existing schemes.

5.1. Simulation Configuration. We implemented MAC/PHY layers with MATLAB according to the system model of IEEE 802.11n operating at the 5 GHz frequency band. The system parameters related to IEEE 802.11n can be found in Table 2 and [1]. TGN channel models were implemented by MATLAB [25] and used to generate MIMO channel matrices for the models defined in [24]. Simulations were conducted for channel model B, which considers residential homes or small offices. The velocity of the moving environment was assumed to be 1.2 km/h, which corresponds to a coherence time of 15 ms approximately. The transmit and the background noise powers were set to 15 dBm and -87 dBm, respectively. FLA was employed as a link adaptation algorithm, and the target error rate η was set to 0.01. The data rate was in the range of 6.5 to 130 Mb/s with a 2×2 MIMO configuration for a channel bandwidth of 20 MHz and a guard interval of 800 ns. User datagram protocol (UDP) packets were generated by constant bit rate (CBR) sources so that the transmitter always had a sufficient number of packets for frame aggregation. We measured the performance in the aspects of per-station throughput, total throughput, and fairness index. The fairness index (FI) was calculated as

$$\text{FI} = \frac{\left(\sum_{i=1}^{N_{\text{sta}}} T_i\right)^2}{N_{\text{sta}} \sum_{i=1}^{N_{\text{sta}}} T_i^2}, \quad (23)$$

where N_{sta} is the number of stations and T_i is the airtime consumed by the i th station for successful frame transmission. Note that FI in (23) is the modified version of Jain's fairness index [26] and it has the ideal value of one when all stations consume the same airtime.

5.2. Validation of Proposed Frame Size Adaptation Scheme. First, we observe how the proposed scheme resolves the unfairness problem that occurred under the same condition described in Section 2.1. We consider two versions of two-level frame size adaptation schemes based on the proposed and total enumeration methods (see Section 4.2), which are denoted as FA2-H and FA2-T, respectively. The FA2-T scheme can be considered as the ideal and optimal scheme that maximizes the aggregate throughput while assuring airtime fairness, while FA2-H scheme is the practical and feasible version of FA2-T with the reduced complexity. The scheme denoted as DCF is a baseline scheme

TABLE 4: Performances of DCF, FA2-H, and FA2-T schemes under the heterogeneous network condition in Section 2.1.

Performance measure		DCF	FA2-H	FA2-T
Airtime ratio	STA1	0.166	0.211	0.210
	STA2	0.397	0.221	0.216
	STA3	0.080	0.210	0.216
	STA4	0.127	0.221	0.219
Fairness index		0.7137	0.9994	0.9998
Throughput (Mb/s)	STA1	1.06	2.24	2.32
	STA2	4.11	2.56	2.51
	STA3	1.07	11.40	11.78
	STA4	4.20	12.51	12.62
Aggregate throughput (Mb/s)		10.44	28.71	29.23

that only employs DCF without any mechanism to enhance fairness or efficiency.

Table 4 compares the airtime ratio and throughput for DCF, FA2-H, and FA2-T schemes. Both FA2-Hand FA2-T schemes provide the acceptable level of airtime fairness; all the stations have the comparable airtime ratio, for example, 0.210~0.221 for FA2-H and 0.210~0.219 for FA2-T, and the values of FI are not smaller than 0.999 for both schemes and their difference is negligible. However, in the case of DCF, the difference in the airtime ratio among stations is quite large; for example, it ranges between 0.080 (STA3) and 0.397 (STA2), and the value of FI is dropped below 0.714. Also, it is important to note that the sum of airtime ratio is about 0.77 in the case of DCF but it is increased to about 0.86 in the cases of FA2-H and FA2-T. This is because the frame aggregation effectively reduces the overhead time for backoff and ACK transmission and contributes to the increase of efficiency.

Next, we observe the per-station throughput and aggregate throughput from Table 4. In the case of DCF, stations with the same packet size (e.g., STA1 and STA3) have similar throughput, although they have different data rates. However, in the cases of FA2-H and FA2-T, the packet size hardly affects the throughput and it is differentiated mostly depending on the data rate; for example, STA1 and STA2 have the same data rate but different packet size, and they have similar throughput. From the viewpoint of aggregate throughput, FA2-H and FA2-T remarkably outperform DCF. Compared to DCF, the aggregate throughput is increased by 2.75 and 2.80 times when FA2-H and FA2-T are employed, respectively. These results are mainly due to the overhead reduction by the frame aggregation. There is no notable difference in the aggregate throughput between FA2-H and FA2-T.

5.3. Validation of Frame Transmission Time by the Two-Level Frame Aggregation. The superior performance of FA2-H shown in Section 5.2 results from the adaptive two-level frame aggregation so that the data transmission time (T_{data}) of each station is regulated close to the target value (T_{ref}). The simulation in this subsection is performed to observe how FA2-H adjusts T_{data} close to T_{ref} , regardless of data rate and packet size. For this purpose, we consider

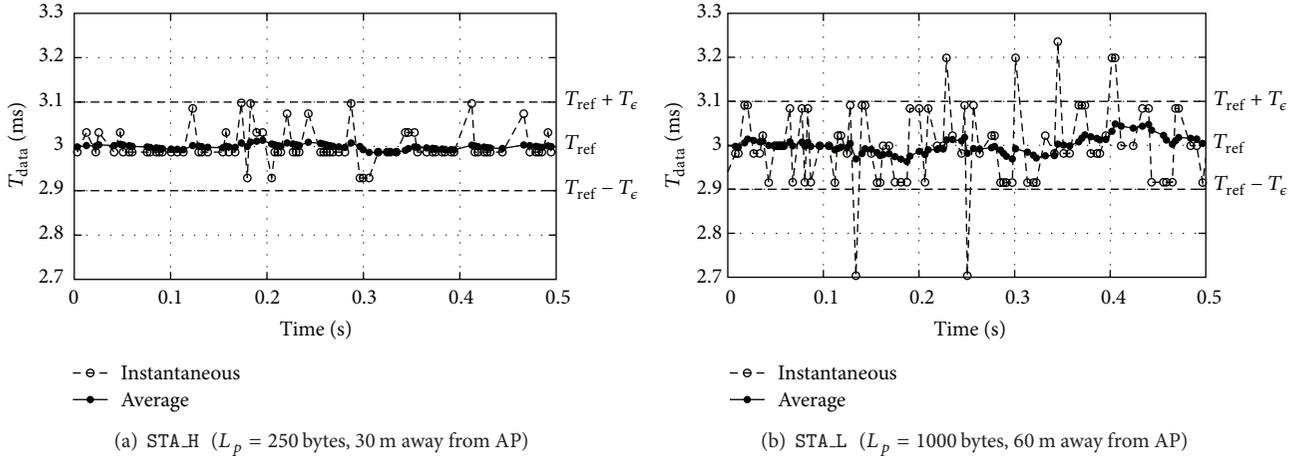


FIGURE 2: Variation of frame transmission time by the proposed two-level frame aggregation.

two representative stations STA_H and STA_L that have different data rate and packet size; STA_H and STA_L were, respectively, located away from the AP by 30 m and 60 m so that STA_H is allowed to have a higher data rate than STA_L, and their packet sizes were set to 250 bytes and 1000 bytes, respectively. In this configuration, STA_H can aggregate more number of packets than STA_L, due to smaller packet size and higher data rate. Note that, unlike the configuration in Section 5.2 where the data rate was fixed, the data rate of station in this configuration was adjusted by the FLA algorithm depending on the time-varying channel quality. The channel was modeled by considering the log-distance path-loss model with the path-loss exponent of 3.5 and Rayleigh fading model as described in [24, 25].

Figure 2 shows how FA2-H adjusts T_{data} for STA_H and STA_L. The instantaneous values indicate the actual T_{data} at each transmission instant, whereas the average values are obtained by an exponentially weighted moving average. In the case of STA_H, the instantaneous value of T_{data} is more tightly regulated around T_{ref} than in the case of STA_L, because FA2-H provides better granularity of control for STA_H thanks to the smaller packet size and higher data rate. However, the average value of T_{data} can be made very close to T_{ref} for both stations due to the compensation mechanism with two sets of (n_1^u, n_2^u) and (n_1^l, n_2^l) (see (18)). This simulation result confirms that the proposed scheme can maintain the average airtime of each station close to T_{ref} , regardless of the data rate and packet size, even when the data rate is changed by the FLA algorithm.

5.4. Effect of Target Airtime on the Performance of Two-Level Frame Aggregation. The objective of this simulation is to evaluate the effect of target airtime (T_{ref}) on the performance of the proposed scheme. The value of T_{ref} is one of key factors affecting the performance. In setting the value of T_{ref} , there exists a trade-off between the decrease of channel access overhead and the increase of frame error rate. The

simulations were carried out under a realistic heterogeneous network configuration where N_{sta} (=10, 20, and 30) stations were randomly located within a circle area of 60 m radius and transmitted data frames to the AP located in the center of circle. The FLA algorithm was implemented to adjust the data rate. The packet size of each station was randomly selected between 200 and 1000 bytes. We investigate the performances of FA2-H and FA2-T in terms of throughput and fairness.

Figure 3(a) shows the aggregate throughput of FA2-H and FA2-T for various values of T_{ref} ranging between 0.5 ms and 5 ms. When $T_{\text{ref}} < 2$ ms, the throughput increases as T_{ref} increases. However, the throughput decreases when T_{ref} exceeds 3 ms. If the channel is free from transmission error, the throughput may increase with respect to the increase of T_{ref} due to the decrease of transmission overheads. However, in the case of error-prone channel, the throughput is not proportional to T_{ref} as shown in Figure 3(a). The reason is that a larger value of T_{ref} increases the channel feedback delay of the link adaptation, which results in the increase of frame error rate and the decrease of throughput accordingly. In addition, we can see that N_{sta} also affects the throughput of FA2. For the entire range of T_{ref} , the throughput with a small value of N_{sta} is always higher than that with a large value of N_{sta} , and their difference is almost constant. The decrease of throughput with respect to the increase of N_{sta} results from the increase of collision probability. Another interesting result is that the optimal value of T_{ref} that maximizes the throughput decreases as N_{sta} increases. This result also stems from the increase of delay in the channel access and/or channel feedback according to the increase of N_{sta} .

Figure 3(b) shows the fairness index for various values of T_{ref} . Unlike the aggregate throughput, FI is nearly immune to the change of T_{ref} . Although FI is slightly decreased as N_{sta} increases, its value is almost close to the ideal value of one for most cases. Both FA2-H and FA2-T provide a high level of airtime fairness, regardless of N_{sta} and T_{ref} . By considering both throughput and fairness shown in Figures 3(a) and 3(b), we can conclude that the appropriate value of T_{ref} is about

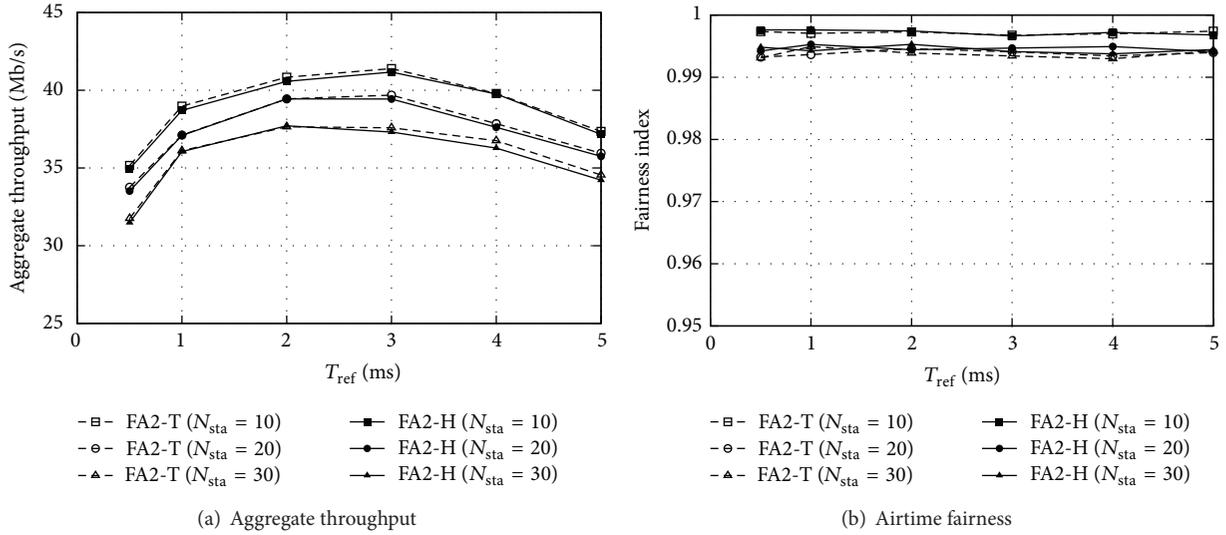


FIGURE 3: Effect of target airtime on the throughput and fairness in the two-level frame size adaptation schemes.

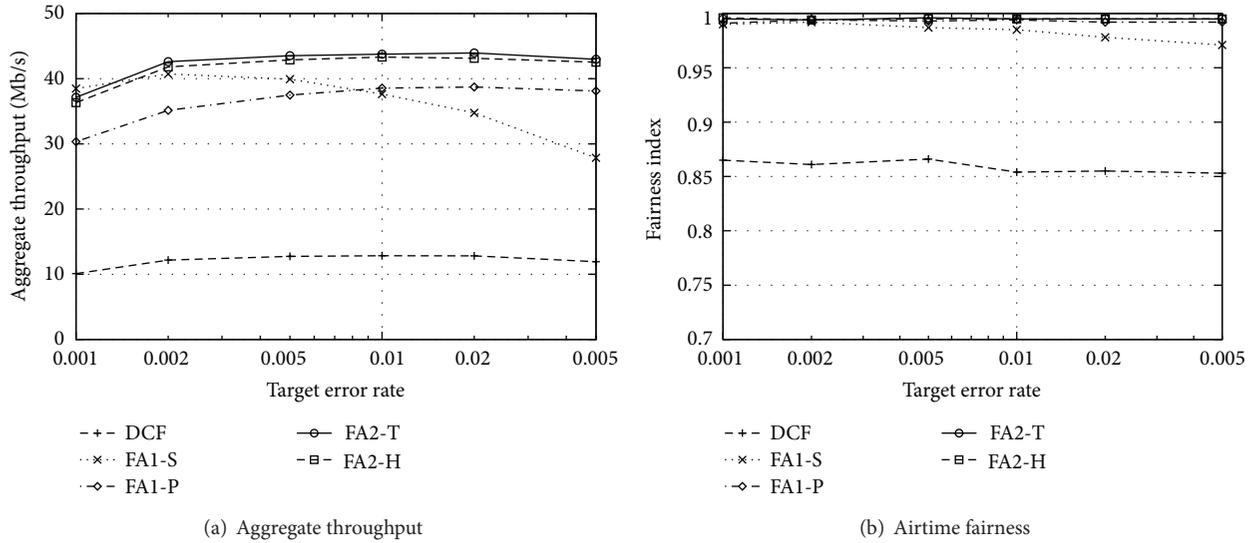


FIGURE 4: Effect of target frame error rate on the throughput and fairness for several aggregation schemes.

3 ms in our simulation environment and we set T_{ref} to 3 ms in the following simulations. Moreover, all of the performance indices of FA2-H are nearly equal to those of FA2-T for all the cases, meaning that the proposed heuristic method is highly effective.

5.5. Two-Level Aggregation versus Single-Level Aggregation.

In this simulation, we compare the performance of two-level aggregation scheme with that of single-level aggregation scheme. For this purpose, we consider two additional comparative mechanisms that employ the single-level aggregation, FA1-S and FA1-P, each of which aggregates MSDU and MPDU subframes up to the given target airtime ($T_{ref} = 3$ ms) as much as possible. We remove the

restriction on the maximum size of A-MSDU in FA1-S for fair comparison. Compared to FA1-S, FA1-P is more robust to the channel error due to the individual ACK of each subframe and selective retransmission, but it has more header overheads. Conversely, FA1-S is vulnerable to channel error but has less header overhead. We fixed the packet size to 500 bytes for all the stations and $N_{sta} (=10)$ stations were randomly distributed and their data rates were adjusted by the FLA algorithm, as described in the previous subsections.

Figure 4(a) shows the aggregate throughput of several frame aggregation schemes where the target error rate η ranges from 0.001 to 0.05. When $\eta = 0.001$, FA1-S outperforms FA1-P because FA1-S has less header overhead than FA1-P and the probability of retransmission

is negligible. As the value of η increases, there exist both positive and negative effects on the throughput. The gain of large η is that the aggressive MCS can be used so that stations can transmit frames at the higher data rate, while its loss is that the frame error rate increases so that the corrupted frame should be retransmitted. In the case of FA1-S, as shown in Figure 4(a), the loss of large η overwhelms its gain; the throughput of FA1-S notably decreases as η increases, because the transmission failure of only an MSDU subframe causes the retransmission of the whole subframes. However, in the case of FA1-P, the throughput slightly increases as η increases; FA1-P is quite less sensitive to the increase of η compared to FA1-S because FA1-P can selectively retransmit only corrupted subframes. In the cases of FA2-T and FA2-H, the advantage of large η is mostly balanced with its disadvantage so that the throughput does not change largely depending on the value of η . The throughput of FA2-H is almost equal to that of FA2-T, and it is larger than that of DCF by about 3.5 times and larger than those of FA1-S and FA1-P up to about 54% and 22%, respectively.

Figure 4(b) shows the effect of η on the fairness. Except for FA1-S, the FI is little affected by the change in the η ; the FI of DCF is about 0.85~0.86 and the FI of FA1-P, FA2-T, and FA2-H is at least 0.99 for the entire range of η . In the case of FA1-S, the FI is slightly decreased from 0.990 to 0.971 as η increases from 0.001 to 0.05, which results from the increase of transmission failure due to the aggressive MCS. Recall that the unsuccessful transmission time is excluded in calculating FI, as shown in (23). The results in Figure 4 confirm the outstanding performance of the proposed two-level aggregation scheme over the single-level aggregation scheme in terms of efficiency and fairness.

5.6. Performance Comparison with respect to the Number of Stations. Lastly, we compare the performance of the proposed scheme with several existing schemes in terms of throughput, fairness, and delay when N_{sta} changes. Here, we consider another existing scheme, denoted as STFM [9], for the comparative study. This scheme adjusts the channel access probability of the legacy WLAN system to achieve airtime fairness and to maximize the aggregate throughput. It has the same objective but different approach as the proposed scheme. We consider only FA1-P as the single-level aggregation scheme. Also, we introduce another performance index, MAC-to-MAC delay, which is defined as the time interval between the instant when a packet is delivered to the MAC layer in the sender and the instant when it is successfully received and delivered to the upper layer from the MAC layer in the receiver.

Figures 5(a), 5(b), and 5(c) represent the aggregate throughput, fairness index, and the average MAC-to-MAC delay, respectively, for various values of N_{sta} ranging from 1 to 30. The value of η was set to 0.01 and the other configurations were the same with those in Section 5.4. First, we can see from Figure 5(a) that the throughput gain of STFM over DCF is marginal even though it greatly improves airtime fairness (see Figure 5(b)). This is because STFM controls the channel access

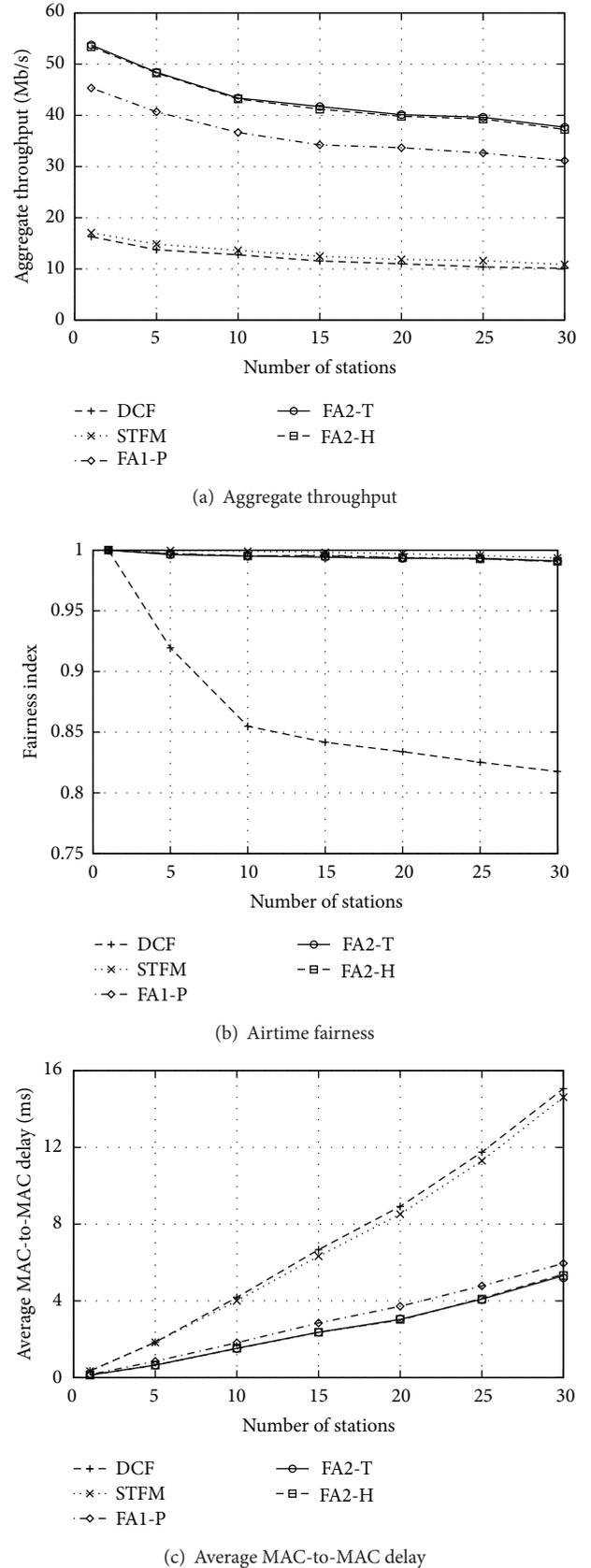


FIGURE 5: Performance comparison of several schemes with respect to the number of stations.

probability to assure fairness, but it cannot take advantage of frame aggregation. However, the throughput enhancement achieved by FA2-H over STFM and FA1-P reaches up to 240% and 22%, respectively. As N_{sta} increases, the aggregate throughput decreases for all schemes, due to the increase of collision probability. However, it is worthwhile noting that the relative throughput gain of FA2-H over other schemes is maintained nearly constant. In terms of airtime fairness, with the exception of DCF, the FI values are nearly close to one regardless of N_{sta} , as shown in Figure 5(b). However, the fairness of DCF remarkably deteriorates as N_{sta} increases; that is, the FI decreased from 0.92 to 0.82 as N_{sta} increases from 5 to 30. Next, we compare the MAC-to-MAC delay of several schemes in Figure 5(c). Similar to the results of throughput, the delay of FA1-P, FA2-T, and FA2-H is significantly smaller than those of DCF and STFM. For the entire range of N_{sta} , the average delay of FA2-H is smaller than that of STFM and FA1-P by approximately 60~65% and 12~23%, respectively. Compared to FA1-P where the single-level aggregation is used, FA2-H further decreases the delay, because the two-level aggregation controls the frame size to maintain minimal overheads depending on the channel state. The delay is nearly proportional to the value of N_{sta} for all schemes due to the contention among stations. However, the relative gain of FA2-H over the others still maintains. In terms of all the performance indices, the performance of FA2-H is almost equal to that of FA2-T.

6. Conclusion

In this paper, we proposed a novel frame size adaptation scheme by employing the generalized two-level frame aggregation scheme in IEEE 802.11n WLANs, to achieve airtime fairness and enhance overall network throughput. By dynamically controlling the size of aggregated frame, the proposed scheme tightly regulates the airtime of each station around the target value and thus effectively resolves the performance anomaly problem that arises due to the difference in data rates and packet sizes among stations. At the same time, the proposed scheme adjusts the frame size so that the throughput can be maximized by considering the channel error rate, data rate, and MAC overhead. We formulated the problem of frame aggregation as an optimization problem and proposed the simple and effective method to reduce the computational complexity. The simulation results under various configurations confirm that the proposed scheme significantly outperforms other existing schemes in terms of the aggregate throughput and channel access delay while maintaining a high level of airtime fairness.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

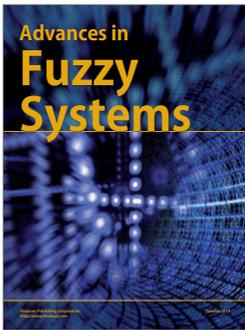
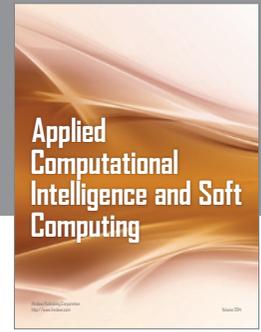
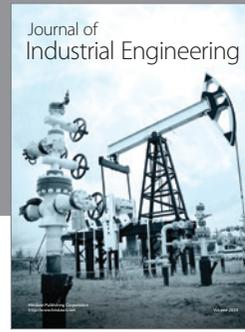
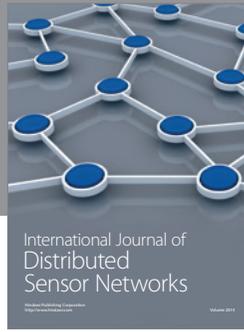
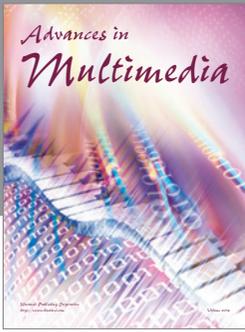
This work was supported by Basic Science Research Program through the National Research Foundation of Korea

(NRF), funded by the Ministry of Education, Science, and Technology under Grants NRF-2012-0002019 and NRF-2013RIA2A2A01009883.

References

- [1] IEEE 802.11, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 5: enhancements for higher throughput," IEEE Std.802.11n-2009, October 2009.
- [2] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proceedings of the IEEE INFOCOM*, pp. 836–843, April 2003.
- [3] G. Tan and J. Guttag, "Time-based fairness improves performance in multi-rate WLANs," in *Proceedings of USENIX*, pp. 269–282, June 2004.
- [4] I. Tinnirello and S. Choi, "Temporal fairness provisioning in multi-rate contention-based 802.11e WLANs," in *Proceedings of the 6th IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WOWMOM '05)*, pp. 220–230, June 2005.
- [5] H. Kim, S. Yun, T. Kang, and S. Bahk, "Resolving 802.11 performance anomalies through QoS differentiation," *IEEE Communications Letters*, vol. 9, no. 7, pp. 655–657, 2005.
- [6] C.-T. Chou, K. G. Shin, and N. S. Shankar, "Contention-based airtime usage control in multirate IEEE 802.11 wireless LANs," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1179–1192, 2006.
- [7] A. Banchs, P. Serrano, and H. Oliver, "Proportional fair throughput allocation in multirate IEEE 802.11e wireless LANs," *Wireless Networks*, vol. 13, no. 5, pp. 649–662, 2007.
- [8] T. Joshi, A. Mukherjee, Y. Yoo, and D. P. Agrawal, "Airtime fairness for IEEE 802.11 multirate networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 4, pp. 513–527, 2008.
- [9] H. Lee and C.-H. Choi, "Achieving airtime fairness and maximum throughput in IEEE 802.11 under various transmission durations," *IEICE Transactions on Communications*, vol. 94, no. 11, pp. 3098–3106, 2011.
- [10] IEEE 802.11, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 8: medium access control (MAC) quality of service enhancements," IEEE Std. 802.11e-2005, November 2005.
- [11] Y. Lin and V. W. S. Wong, "WSN01-I: frame aggregation and optimal frame size adaptation for IEEE 802.11n WLANs," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '06)*, pp. 1–6, San Francisco, Calif, USA, November-December 2006.
- [12] B. Ginzburg and A. Kesselman, "Performance analysis of A-MPDU and A-MSDU aggregation in IEEE 802.11n," in *Proceedings of the IEEE Sarnoff Symposium (SARNOFF '07)*, May 2007.
- [13] B. S. Kim, H. Y. Hwang, and D. K. Sung, "Effect of frame aggregation on the throughput performance of IEEE 802.11n," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '08)*, pp. 1740–1744, April 2008.
- [14] X. He, F. Y. Li, and J. Lin, "Link adaptation with combined optimal frame size and rate selection in error-prone 802.11n networks," in *Proceedings of the IEEE International Symposium on Wireless Communication Systems (ISWCS '08)*, pp. 733–737, October 2008.
- [15] T. Li, Q. Ni, D. Malone, D. Leith, Y. Xiao, and T. Turetletti, "Aggregation with fragment retransmission for very high-speed

- WLANs,” *IEEE/ACM Transactions on Networking*, vol. 17, no. 2, pp. 591–604, 2009.
- [16] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, “Opportunistic media access for multirate ad hoc networks,” in *Proceedings of the 8th ACM Annual International Conference on Mobile Computing and Networking*, pp. 24–35, September 2002.
- [17] D. Skordoulis, Q. Ni, H.-H. Chen, A. P. Stephens, C. Liu, and A. Jamalipour, “IEEE 802.11n MAC frame aggregation mechanisms for next-generation high-throughput WLANs,” *IEEE Wireless Communications*, vol. 15, no. 1, pp. 40–47, 2008.
- [18] C.-Y. Wang and H.-Y. Wei, “IEEE 802.11n MAC enhancement and performance evaluation,” *Mobile Networks and Applications*, vol. 14, no. 6, pp. 760–771, 2009.
- [19] M. Lampe, H. Rohling, and W. Zirwas, “Misunderstandings about link adaptation for frequency selective fading channels,” in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '02)*, vol. 2, pp. 710–714, IEEE, September 2002.
- [20] S. Tsai and A. Soong, “Effective-SNR mapping for modeling frame error rates in multiplestate channels,” 3GPP2-C30-20030429-010, April 2003.
- [21] G. Martorell, F. Riera-Palou, and G. Femenias, “Cross-layer fast link adaptation for MIMO-OFDM based WLANs,” *Wireless Personal Communications*, vol. 56, no. 3, pp. 599–609, 2011.
- [22] T. L. Jensen, S. Kant, J. Wehinger, and B. H. Fleury, “Fast link adaptation for MIMO OFDM,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 8, pp. 3766–3778, 2010.
- [23] C. Yeo, *Improving IEEE 802.11n link adaptation using coherence time estimation [M.S. thesis]*, Seoul National University, Seoul, Korea, 2010.
- [24] V. Erceg, L. Schumacher, P. Kyritsi et al., “TGn channel models,” *IEEE 802.11 document 03/940r4*, May 2004.
- [25] L. Schumacher and B. Dijkstra, “Description of a MATLAB implementation of the indoor MIMO WLAN channel model proposed by the IEEE 802.11 TGn channel model special committee,” Implementation note version 3.1, January 2004.
- [26] D.-M. Chiu and R. Jain, “Analysis of the increase and decrease algorithms for congestion avoidance in computer networks,” *Computer Networks and ISDN Systems*, vol. 17, no. 1, pp. 1–14, 1989.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

