We propose a novel media access control (MAC) protocol, referred to as signaling-free max-min airtime fair (SMAF) MAC, to improve fairness and channel utilization in ad hoc networks based on IEEE 802.11 wireless local area networks (WLANs). We introduce busy time ratio (BTR) as a measure for max-min airtime fairness. Each node estimates its BTR and adjusts the transmission duration by means of frame aggregation and fragmentation, so that it can implicitly announce the BTR to neighbor nodes. Based on the announced BTR, each of the neighbor nodes controls its contention window. In this way, the SMAF MAC works in a distributed manner without the need to know the max-min fair share of airtime, and it does not require exchanging explicit control messages among nodes to attain fairness. Moreover, we successfully incorporate the hidden node detection and resolution mechanisms into the SMAF MAC to deal with the hidden node problem in ad hoc networks. The simulation results confirm that the SMAF MAC enhances airtime fairness without degrading channel utilization, and it effectively resolves several serious problems in ad hoc networks such as the starvation, performance anomaly, and hidden node problems.

1. Introduction

The recent explosive proliferation of mobile devices such as smartphones and tablet PCs has accelerated the demand for wireless Internet access. The wireless local area network (WLAN) based on IEEE 802.11 standard [1] is one of the most prevailing wireless communication technologies thanks to its ease of deployment and low installation cost. At the same time, Wi-Fi Alliance has recently released a standard for peer-to-peer communication based on WLAN (called Wi-Fi Direct) to support file sharing, multimedia streaming, content synchronization, and printing between Wi-Fi devices without the aid of an infrastructure device [2]. Furthermore, the recent development of the Internet of Things has facilitated and promoted the interconnectedness of many diverse devices and has driven the production of various applications [3, 4]. It is expected that new services based on ad hoc networks will emerge and increase rapidly. Thus, it is imperative in the operation of ad hoc networks to provide fair service and efficient channel utilization.

The current media access control (MAC) protocol in IEEE 802.11 WLANs provides fair services in terms of channel access opportunity, when all the nodes are located within the carrier-sensing range of each other and they are homogeneous in terms of transmission power, rate, and/or range, which is not a typical condition for an ad hoc network to satisfy. Therefore, the standard MAC protocol of IEEE 802.11, DCF (distributed coordination function), cannot provide fairness in ad hoc networks [5]. It is challenging to design a MAC protocol that works well in ad hoc networks and such attempts have hitherto not been very successful. This is due to several intrinsic characteristics of wireless channel (e.g., time-varying channel quality, transmission failure due to interference or channel error) and ad hoc networks (e.g., dynamic network topology, multihop transmission, and existence of hidden and exposed terminals) and the distributed nature of the MAC protocol (e.g., location-dependent contention, asymmetry in carrier sensing, and random backoff mechanism). The main objectives in designing a MAC protocol for ad hoc networks are to provide fair channel sharing among neighbor nodes, to maximize network capacity, and to minimize transmission delay and/or energy consumption of nodes. These objectives usually conflict with each other, so one has to find an acceptable compromise. This paper
primarily focuses on providing fairness without degrading overall channel utilization.

Extensive studies have been conducted to provide fairness in ad hoc networks [6–15]. The approach in [6] calculates the max-min fair share of a node by constructing a flow contention graph and controls the contention window size to provide max-min fairness. This approach can be modified to support proportional fairness by considering airtime usage [7]. In order to achieve both fairness and maximization of channel utilization, the distributed scheduling algorithm in [8] emulates the fair queuing mechanism by assigning start and finish tags for each packet. The weighted fair queuing mechanism can also be implemented by controlling interframe space of IEEE 802.11 WLANs [9]. The protocol described in [10] makes each node build a scheduling table by overhearing the packet priority (i.e., arrival time) of a neighbor node that is piggybacked in the control and/or data frames and then exploits the priority information in the backoff procedure to improve fairness. Also, the protocol in [11] aims to avoid the unfairness problem; the sender and/or receiver initiates information exchange among neighbor nodes before data transmission and establishes a neighbor table and a flow table. In [12], each node periodically broadcasts the estimate of its attainable throughput using a control frame, which is used to adjust the contention window of neighbor nodes. These mechanisms, which are based on the flow contention graph in [6, 7], require that each node should be aware of the local network topology of its neighbor nodes and that the flow contention information should be exchanged among neighbor nodes. Similarly, the tag-based fair scheduling mechanisms in [8–10] and the protocols in [11, 12] require that some information should be included in the frame header or in the control frame so that it can be overheard and exchanged among neighbor nodes.

Another approach available is topology-independent and works without information exchange [13–15]. The study in [13, 14] defines a fairness index that accounts for the fair share of channel occupied by each node and by its neighbor nodes and proposes to adjust the contention window size depending on the estimated value of the fairness index. In [15], each node measures the mean number of consecutive idle slots between two transmission attempts and controls the transmission opportunity based on the number of idle slots to attain fairness. This protocol is efficient in single hop networks but not suitable in ad hoc networks.

All these approaches cannot simultaneously and successfully handle various aspects of a network such as max-min fairness, network utilization, and problems originating from an ad hoc network with the hidden/exposed nodes and multihop transmission. Moreover, most of these studies, designed to provide max-min fairness in ad hoc networks, have to solve two major problems: (i) exchanging necessary information among neighbor nodes and/or (ii) computing the max-min fair share of a node. Firstly, the approach based on control messages may fail to ensure fairness among the nodes that are located within the carrier-sensing range, but out of the transmission range of each other, because they cannot obtain necessary information from the control messages. Furthermore, this approach may degrade the efficiency of channel usage due to signaling overheads, especially in dynamic environments where the control messages need to be frequently updated and exchanged (or propagated). Secondly, the calculation of the max-min fair share is not practical in ad hoc networks, because it is difficult to calculate the fair share correctly in a timely and distributed way, and the max-min fair share should be recalculated in response to the change in the network configuration such as the number of nodes, network topology, traffic load, and/or transmission rate. Moreover, the max-min throughput fairness may not be desirable when each node has a different transmission rate. Since a lower-rate node occupies a wireless channel longer than a higher-rate node to transmit a frame, the max-min throughput fairness results in the decrease of total network throughput, referred to as a performance anomaly in multirate networks [16].

This study is designed to resolve the drawbacks of the previous studies addressed above. In this paper, we propose a novel MAC protocol, called signaling-free max-min airtime fair (SMAF) MAC, in order to improve max-min fairness and channel efficiency in ad hoc networks. In the SMAF MAC, each node firstly estimates the ratio of its airtime usage with respect to the total busy time of the channel, defined as busy time ratio (BTR). This estimation can be simply performed by the carrier-sensing mechanism that is already implemented for channel access, and it does not require any information on network topology. Secondly, each node implicitly conveys its BTR to the neighbor nodes that are located within its carrier-sensing range and compete for the shared airtime. For this purpose, we adopt the idea in our previous work [17] that the transmission duration is adjusted to announce necessary information without resorting to any explicit control messages. Lastly, by sensing the transmission duration of neighbor nodes, each node estimates the BTR of neighbor nodes and compares it with its own BTR. Based on this simple comparison result, a node adjusts its contention window to improve fairness without the need to know its max-min fair share. The main contributions of the SMAF MAC compared to the previous studies are summarized as follows.

(i) The SMAF MAC announces the BTR information in a signaling-free manner by encoding it into the transmission duration of a data frame. Therefore, the SMAF MAC does not require any additional control frames or fields in the frame header, which induces a high overhead and decreases the effective capacity of the network. In addition, the SMAF MAC is effective in announcing the information to neighbor nodes even though they are located out of the transmission range of each other, and it is resistant to transmission collision or interference.

(ii) It operates in a distributed manner based on the BTR without requiring the explicit information of the max-min fair share that depends on the network configuration. The SMAF MAC is also effective in dealing with the problems of starvation and performance anomaly [5, 16], because each node adjusts its transmission attempt probability based on the BTR of neighbor
nodes so that the airtime can be shared fairly and efficiently among neighbor nodes.

(iii) The framework of the SMAF is extended to effectively handle the hidden node problem. It selectively utilizes the hidden node resolution mechanism, that is, ready-to-send and clear-to-send (RTS/CTS) exchange mechanism, based on the estimation of the achievable throughput. Also, it supports the virtual extension of CTS frame coverage, by which the hidden node problem can be further mitigated.

The rest of this paper is organized as follows. Section 2 presents the theoretical background on max-min airtime fairness in ad hoc networks. The basic operation of SMAF MAC and its implicit signaling mechanism based on frame aggregation and fragmentation are described in Sections 3 and 4, respectively. Section 5 discusses several issues related to the SMAF MAC in practical environments. Section 6 validates the performance of SMAF MAC via ns-2 simulations [18]. Finally, Section 7 concludes this paper.

2. Max-Min Airtime Fairness in Ad Hoc Networks

In this section, we first define max-min airtime fairness in ad hoc networks. We are primarily interested in ad hoc networks where each node is a saturated node (i.e., a node that always has backlogged packets to transmit) and its transmission affects the transmission/reception of some of the other nodes directly or indirectly. When there are two sets of nodes in a network where each node in a set affects only the communications between the nodes within this set, but not the communication between the nodes in the other set, we consider them as two independent ad hoc networks. Ad hoc networks consisting of saturated and unsaturated nodes will be discussed in Section 5. We also introduce several measures related to fairness and derive a condition to achieve fairness.

The notion of max-min fairness was introduced in wired networks [19, 20]. However, it cannot be straightforwardly applied to ad hoc networks since the capacity of a wireless channel is time-varying and the channel resource is shared in a contention-based way and spatially reused. The channel resource can be considered from various aspects such as throughput, airtime, and/or transmission opportunity. In this paper, we consider airtime as the channel resource that should be shared among transmitting nodes in a fair manner, because airtime is suitable for dealing with time-varying capacity, spatial reuse of wireless channel, and the unfairness problem due to different transmission rates of nodes. Note that the max-min fairness in terms of airtime is equivalent to the proportional fairness in terms of throughput [7, 21, 22]. We consider an ad hoc network where node $i$ shares its channel resource with neighbor nodes. We call node $j$ a neighbor node of node $i$ if it can sense the transmission by node $i$. We denote the set of nodes consisting of node $i$ and its neighbor nodes as $S_i$. Let $A_i$ be the channel occupation time of node $i$ during the time interval of $T$ (for the time being, $T$ is assumed to be fixed for ease of explanation; however, it may vary and may be different from each node; details will be given in Section 3), and let $a_i = A_i/T$ be the normalized airtime of node $i$. Also, we define $B_i$ as the time duration in which node $i$ senses the channel busy due to the transmission of its neighbor nodes during the interval of $T$ and introduce $b_i$ as $b_i = B_i/T$. Here, $B_i$ can be measured by means of a clear channel assessment (CCA) mechanism defined IEEE 802.11 standard. When measuring $B_i$, node $i$ does not have to identify the transmitting node. Any unsuccessful transmission time, as well as successful transmission time, is taken into account in $A_i$ and $B_i$, because even an unsuccessful transmission by a certain node consumes the channel resource shared with its neighbor nodes and the neighbor nodes must defer their channel access during this transmission time. We will simply refer to $a_i$ and $b_i$ as the airtime of node $i$ and its neighbor nodes if no confusion arises, respectively. We define that the airtime vector $\vec{a} = (a_1, a_2, \ldots, a_N)$ is feasible if $a_i + b_i \leq 1$ for all $1 \leq i \leq N$, where $N$ is the total number of transmitting nodes in the network. Note that there can exist node $m$ such that $\sum_{k \in S_i} a_k > 1$; that is, the airtime can be spatially reused among at least two neighbor nodes of node $m$ that are located out of the carrier-sensing range of each other.

Definition 1 (max-min airtime fairness). An airtime vector $\vec{a}$ is max-min airtime fair if $\vec{a}$ is feasible and it is impossible to increase the airtime of a node without decreasing the airtime of the other node that has smaller airtime.

Note that the max-min airtime fairness in this definition is different from the conventional max-min fairness in wired networks unless all the nodes in the network share a common carrier-sensing coverage. Next, we introduce a measure for fairness. Jain’s fairness index [23] is a common measure for fairness in wired networks. However, it is not appropriate to evaluate the fairness in an ad hoc network because each node has different neighbor nodes and the number of neighbor nodes and their transmission rates may be different from one to another. Thus, we define new fairness indices for an ad hoc network.

Definition 2 (per-node and network-wide fairness indices). By modifying Jain’s fairness index, one defines the airtime fairness index for node $i$ as

$$J_i = \frac{(\sum_{k \in S_i} a_k)^2}{N_i \cdot \sum_{k \in S_i} a_k^2},$$

where $N_i$ is the number of transmitting nodes in $S_i$; that is, $N_i = |S_i|$. Also, one defines the network-wide airtime fairness index as

$$J = \frac{1}{N} \sum_{i=1}^{N} J_i.$$ 

Note from (1) and (2) that $J_i$ becomes one only when $a_i = a_j$ for all $i, j$ ($j \neq i$) $\in S_i$, and that $J = 1$ implies $J_i = 1$ for all $i$ since $J_i \leq 1$ for any $i$.

From these fairness indices, we will derive the condition for max-min airtime fairness in an ad hoc network.
Proposition 3 (condition for max-min airtime fairness). An airtime vector \( \overrightarrow{a} \) is max-min airtime fair if and only if \( \overrightarrow{a} \) is feasible and \( a_i (i = 1, 2, \ldots, N) \) have the maximum value under the constraint of \( \overrightarrow{f} = 1 \).

Proof. (i) (If part) since \( \overrightarrow{f} = 1 \), each node has the same airtime as its neighbor nodes; that is, \( a_i = a_j \) for all \( j \neq i \) = 1, 2, \ldots, N. Let \( a' \) be the maximum airtime that a node can achieve under the constraint of \( \overrightarrow{f} = 1 \). Then, \( a_i = a' \) for all \( i \). Assume that \( a_0 = (a', a', \ldots, a') \) is not max-min airtime fair. This assumption means that there exists another airtime vector \( \overrightarrow{a_0} = (a' + \delta, a' + \delta, \ldots, a' + \delta) \) (\( \delta > 0 \)). However, this contradicts the definition of \( a' \), confirming that \( \overrightarrow{a} \) is max-min airtime fair.

(ii) (Only if part) this is the case when each node in the network increases its airtime from zero until there exists a node that cannot increase its airtime any more. Let \( \overrightarrow{a} \) be the airtime value of each node at this point. In this case, if some nodes try to increase their airtime from \( \overrightarrow{a} \), there must exist a node that has to decrease its airtime from \( \overrightarrow{a} \), and consequently, the minimum airtime value of a node in the network decreases from \( \overrightarrow{a} \). Therefore, \( \overrightarrow{a} \) is the max-min airtime value that each node can attain, and \( \overrightarrow{f} = 1 \).

3. Signaling-Free Max-Min Airtime Fair (SMAF) MAC

In this section, we firstly introduce a new variable, busy time ratio (BTR), which plays a key role in enhancing max-min airtime fairness in the SMAF MAC, and present how it can be practically measured. Next, we propose a method to adjust the contention window to achieve fairness in a distributed way without explicit signaling messages.

3.1. Condition of the BTR for Max-Min Airtime Fairness and Its Estimation. Proposition 3 in Section 2 provides a clue for max-min airtime fairness, and we need to develop a practical mechanism to realize this objective in a distributed way. For this purpose, we define the BTR as follows.

Definition 4 (busy time ratio). The busy time ratio of node \( i \), \( \rho_i \), is defined as the fraction of the channel busy time due to the transmission by node \( i \) with respect to the total channel busy time measured by node \( i \); that is,

\[
\rho_i = \frac{A_i}{A_i + B_i} = \frac{a_i}{a_i + b_i}. \tag{3}
\]

We assume that the network is saturated; that is, there exist nodes that have a sufficient number of packets to transmit and try to fully utilize the available channel resource. This assumption is reasonable because there is no hindrance to fair channel sharing as long as the total demand for channel access of all the nodes can be met by the available resource (i.e., airtime). (The case involving unsaturated nodes will be handled in Section 5.3 and will be investigated via simulation in Section 6.3.)

Proposition 5 (condition of BTR for max-min airtime fairness). Airtime vector \( \overrightarrow{a} \) is approximately max-min airtime fair if and only if nodes \( i \) and \( j \) greedily access the channel to maximize \( \rho_i \) and \( \rho_j \) while satisfying the constraint of \( \rho_i = \rho_j \), for all \( j \in S_i \).

Proof. With the assumption of saturated network, the channel is occupied for most of the time by node \( i \) and/or its neighbor node \( j \) (\( j \in S_i \)); that is, \( a_i + b_i = a_j + b_j = 1 \), and \( \rho_i = a_i \) from (3). Therefore, the constraint of \( \rho_i = \rho_j \) results in \( a_i = a_j \), and \( \overrightarrow{f} = 1 \) from (1) and (2). Consequently, as given in Proposition 3, \( \overrightarrow{a} \) is max-min airtime fair if and only if \( \rho_i = a_i \) is maximized under the constraint of \( \rho_i = \rho_j \).

Now, we present how each node estimates \( \rho_i \) using only the measurable MAC layer statistics. Figure 1(a) illustrates the timeline from the viewpoint of node \( i \) where node \( i \) and its neighbor nodes compete and occupy the channel according to IEEE 802.11 DCF. The time interval \( T \) consists of backoff slots (i.e., idle slots), transmission time of node \( i \), channel busy time due to neighbor nodes’ transmission, and several overhead times (i.e., short interframe space (SIFS), acknowledgement (ACK) transmission time, and distributed interframe space (DIFS)). We define a transmission instant as the time period consisting of data frame transmission time, SIFS, ACK frame transmission time, and DIFS, as shown in Figure 1(a). For the sake of simplicity, we can abstractly draw transmission instants of node \( i \) and its neighbor nodes as black circles and white squares, respectively, as shown in Figure 1(b). We define the estimation period of node \( i \) as the time between two transmission instants of node \( i \). More specifically, it starts from the end of its current transmission instant and terminates at the end of its next transmission instant if it senses at least one transmission of its neighbor nodes between these two transmission instants. This case corresponds to the \( r \)th estimation period in Figure 1(b). If the transmission instants of node \( i \) happen to occur successively, for example, the \( (r + 1) \)th estimation period in Figure 1(b), the start time of the \( (r + 1) \)th estimation period is still the end of the \( r \)th estimation period, but its end time is extended to the end of the later transmission instant of node \( i \) until node \( i \) senses the channel busy due to the transmission by neighbor nodes. For example, if there is no transmission of neighbor nodes during the \( (r + 1) \)th period in Figure 1(b), then the \( (r + 1) \)th and \( (r + 2) \)th estimation periods should be merged into the \( (r + 1) \)th period. It is noteworthy that the estimation period may differ from one node to another and that it does not have to be synchronized for all the nodes. Therefore, there is no difficulty in a node’s ability to measure its own estimation period.

Let us define \( A_i[r, k] \) and \( B_i[r, l] \) as the durations of the \( k \)th transmission instant by node \( i \) and the \( l \)th transmission instant by node \( i \)’s neighbor nodes in the \( r \)th estimation period of node \( i \), denoted as \( T_i[r] \), respectively. Then, \( a_i[r] \) and \( b_i[r] \) can be represented as

\[
a_i[r] = \frac{\sum_{k=1}^{n_i} A_i[r, k]}{T_i[r]}, \tag{4}
\]

\[
b_i[r] = \frac{\sum_{l=1}^{n_i} B_i[r, l]}{T_i[r]},
\]
where \( n_i[r] \) and \( n_o[r] \) denote the numbers of transmission instants by node \( i \) and by its neighbor nodes in the \( r \)th estimation period, respectively. Node \( i \) estimates its BTR \( \rho_i \) at the end of the \( r \)th estimation period as

\[
\rho_i [r] = \frac{a_i [r]}{a_i [r] + b_i [r]} = \frac{\sum_{k=1}^{n_i[r]} A_i [r, k]}{\sum_{k=1}^{n_i[r]} A_i [r, k] + \sum_{l=1}^{n_o[r]} B_i [r, l]}.
\]

In order to reduce the estimation error and abnormal variations, node \( i \) computes the estimate of \( \rho_i \), denoted as \( \tilde{\rho}_i \) (hereafter, we denote the BTR of node \( i \) as \( \tilde{\rho}_i \), unless otherwise stated), by using an exponentially weighted moving average as

\[
\tilde{\rho}_i [r] = \omega \cdot \tilde{\rho}_i [r - 1] + (1 - \omega) \cdot \rho_i [r],
\]

where \( \omega (0 < \omega < 1) \) is a constant weight for the moving average.

This method of BTR estimation has the following several advantages: (i) it neither requires any changes in the standard of IEEE 802.11 nor incurs any additional overheads; (ii) each node estimates its own BTR in a distributed way without resorting to a globally synchronized estimation period; (iii) the BTR can be easily measured by the carrier-sensing mechanism requiring no information on channel idle time or collision probability.

3.2. Announcing the BTR Information and Adjusting the Contention Window. In order to achieve the max-min airtime fairness according to Proposition 5, it is essential for each node to inform its neighbor nodes of its BTR so that they can adjust their airtime based on this information. The announcement of BTR can be implemented by using frame headers or dedicate control frames; however, this approach is not desirable from the following aspects.

(i) **Efficiency.** The efficiency of channel usage may significantly decrease because of the additional signaling overheads, especially in high-speed networks where the frame headers or control frames are transmitted at a low rate, regardless of the transmission rate of the data frame. Moreover, this approach is not scalable with respect to the number of nodes because the number of control frames exchanged among nodes increases in proportion to the number of nodes.

(ii) **Effectiveness.** It is difficult to announce \( \tilde{\rho}_i \) to the neighbor nodes that are within the carrier-sensing range, but out of the transmission range of node \( i \).

(iii) **Reliability.** This approach cannot ensure reliable delivery of the BTR information because the frame header or control frame can collide with other control or data frames, or it cannot be correctly decoded due to interference or channel error.

To overcome these shortcomings of the conventional signaling mechanism, the SMAF MAC adopts a novel approach that does not rely on an explicit signaling mechanism. The key idea is to encode the value of \( \tilde{\rho}_i \) into the length of transmission duration. Each node adjusts its transmission duration according to the value of BTR and implicitly informs its neighbor nodes of its BTR by means of transmission duration. Since the BTR information is not delivered by a dedicate control frame, it can be estimated by all nodes.
within the carrier-sensing range and is less vulnerable to transmission failure of the control frame due to collision, interference, or channel error. Moreover, the signaling of BTR can be achieved at minimal cost because the transmission duration containing the BTR information is controlled by a frame fragmentation and aggregation scheme. Details will be given in Section 4. Consequently, the announcement of BTR can be achieved without resorting to any explicit signaling mechanism, and thus the proposed mechanism is called “signaling-free.”

In the SMAF MAC, node $i$ determines its transmission duration $T_{tx,i}$ in accordance with $\tilde{\rho}_i$ as

$$T_{tx,i} = T_{\min} + \lfloor \tilde{\rho}_i \cdot 100 \rfloor \Delta,$$  \hspace{1cm} (7)

where $T_{\min}$ is the minimal transmission duration that is a fixed design parameter, $\Delta$ ($>0$) is a basic unit for additional transmission duration, and $\lfloor \cdot \rfloor$ is the round-down operator. Each node implicitly announces its BTR as a percentage (%) by adjusting its transmission duration as shown in Figure 2. Then, the neighbor nodes can estimate $T_{tx,j}$ by using the clear channel assessment (CCA) mechanism (it is also possible for a neighbor node to determine $T_{tx,j}$ by decoding the fields of RATE and LENGTH in the physical layer convergence protocol (PLCP) header of IEEE 802.11 frame, each of which indicates the transmission rate and the size of the PHY-layer service data unit (PSDU), resp.; however, this approach is less reliable because the PLCP header may not be correctly decoded due to collision or interference), which is defined in the IEEE 802.11 standard to determine whether the channel is busy or not, and they can deduce $\tilde{\rho}_j$ from $T_{tx,j}$ by using (7). We provide a guideline for determining the values of $T_{\min}$ and $\Delta$. The value of $T_{tx,j}$ needs to be shorter than the channel coherence time (e.g., on the order of a few tens of milliseconds in a typical WLAN channel [24]), and $\Delta$ needs to be on the order of CCA time required for physical carrier sensing and slot time (e.g., $4 \mu s$–$9 \mu s$ in most IEEE 802.11 standards [1]) so that the channel noise level does not fluctuate severely during a frame transmission and the neighbor nodes can correctly estimate the channel state (e.g., busy or idle) by physical carrier sensing.

Next, we propose a method to achieve fairness based on the BTR information. Whenever node $i$ extracts $\tilde{\rho}_j$ of its neighbor node from the transmission duration, node $i$ adjusts its probability of transmission attempt $\tau_i$ so that $\tilde{\rho}_i$ becomes as close as $\tilde{\rho}_j$. In adjusting $\tau_i$, we adopt the method of additive increase and multiplicative decrease (AIMD), which is a well-known rate control algorithm to make the transmission rate of flows converge to a stable and fair equilibrium in a distributed manner [23, 25]. With this rationale, we adjust $\tau_i$ to achieve $\tilde{\rho}_i = \tilde{\rho}_j$ as

$$\tau_i \leftarrow \begin{cases} \tau_i + \lambda, & \text{if } \tilde{\rho}_i < \tilde{\rho}_j, \\ \mu \tau_i, & \text{otherwise}, \end{cases}$$  \hspace{1cm} (8)

where $\lambda (>0)$ and $\mu$ ($0 < \mu < 1$) are the design parameters for additive increase and multiplicative decrease, respectively. If node $i$ determines that its BTR is smaller than the BTR of its neighbor nodes (i.e., $\tilde{\rho}_i < \tilde{\rho}_j$), it additively increases its transmission attempt probability to acquire more airtime and get its fair. Otherwise, it multiplicatively decreases $\tau_i$ to yield the channel resource to its neighbor nodes. It is important to note that the value of $\tau_i$ is updated whenever node $i$ senses a transmission of a neighbor node and there is no need to identify the transmitter in updating $\tau_i$. Now, we rewrite the rule for adjusting $\tau_i$ in terms of contention window (CW). Because the transmission attempt is equivalent to the inverse of the average size of CW, that is, $\tau_i = 2/(\text{CW}_i + 1)$ [26], (8) becomes

$$\text{CW}_i \leftarrow \begin{cases} \max \left( \frac{2 (\text{CW}_i + 1)}{2 + \lambda (\text{CW}_i + 1)} - 1, \text{CW}_{\min} \right), & \text{if } \tilde{\rho}_i < \tilde{\rho}_j, \\ \min \left( \frac{\text{CW}_i + 1}{\mu} - 1, \text{CW}_{\max} \right), & \text{otherwise}, \end{cases}$$  \hspace{1cm} (9)

where $\text{CW}_{\min}$ and $\text{CW}_{\max}$ are the minimum and maximum values of CW, respectively. The SMAF MAC does not use the binary exponential backoff mechanism of IEEE 802.11 DCF, which blindly doubles the value of CW upon transmission failure, regardless of causes (e.g., collision, interference, and channel error), because the SMAF MAC controls the transmission attempt based on the BTR.

4. Transmission Duration Control in SMAF MAC

This section describes how a node in the SMAF MAC adjusts its transmission duration where the BTR information is encoded. Basically, the transmission duration is controlled by a frame fragmentation/aggregation (F/A) scheme, which is similar to the aggregated MAC protocol data unit (A-MPDU) in the IEEE 802.11n standard [27] and the aggregation with fragment retransmission (AFR) scheme proposed in [28]. The proposed frame F/A scheme precisely adjusts the transmission duration by supporting variable fragment size and combining the fragment scheme with the aggregation scheme. This enables neighbor nodes to accurately deduce the BTR information of a transmitting node, which is the primary objective of our frame F/A scheme. Similar to the other schemes, our frame F/A scheme can improve channel efficiency and transmission reliability by reducing channel overhead and frame error rate, which are its additional advantages.
Figure 3 shows the frame format and fragment subheader of the SMAF MAC. The SMAF MAC fragments and aggregates packets to form a MAC protocol data unit (MPDU) of arbitrary size. For this purpose, we slightly modify the MAC header to add a field called fragment number, which indicates the number of fragments (subframes) aggregated in an MPDU. Also, we define a fragment subheader consisting of the following four fields and frame check sequence (FCS), as shown in Figure 3.

1. Packet ID: the identifier of a packet to which a fragment belongs.
2. Start position: the position in a packet at which the fragment starts (in bytes).
3. Fragment size: the size of a fragment (in bytes).
4. Fragment ID: the identifier of a fragment.

Upon receiving an MPDU, a receiver reconstructs packets using the information in the fields of the fragment subheader. When packets are successfully constructed, they are delivered to the upper layer. When some of the fragments are corrupted, which can be determined by the individual FCS following each fragment, the receiver buffers noncorrupted fragments and requests the retransmission of corrupted fragments by sending a block ACK frame. In the ACK frame, we add a new field, Fragment ACK bitmap, similar to the block ACK in the IEEE 802.11n, with which a receiver can inform the sender of the fragments that are not successfully received. Thus, the sender can retransmit only the corrupted fragments indicated in Fragment ACK bitmap.

The SMAF MAC determines the size and number of fragments according to the frame F/A scheme explained above. Let us define $S_{PH}$ and $S_{MH}$ as the sizes of PHY and MAC headers in bytes, respectively. The PHY header is usually transmitted at the lowest rate of $R_{PH}$ (bit/sec), regardless of the transmission rate of data frame $R_i$, whereas the MAC header and fragment subheader are transmitted at the rate of $R_i$. Also, we define $S_{fb,i}$ (byte) as the size of frame body including fragment subheader and FCS (see Figure 3) that meets the requirement of node $i$’s transmission duration $T_{tx,i}$. Then, we can associate $T_{tx,i}$ with $S_{fb,i}$ as

$$T_{tx,i} = \frac{S_{PH}}{R_{PH}} + \frac{(S_{MH} + S_{fb,i})}{R_i}.$$  \hfill (10)

Once $T_{tx,i}$ is determined from (7) based on the value of BTR, node $i$ calculates $S_{fb,i}$ from (10) as

$$S_{fb,i} = \frac{T_{tx,i} \cdot R_i - S_{PH} \cdot R_{PH} - S_{MH}}{8}.$$  \hfill (11)

After determining $S_{fb,i}$, the SMAF MAC constructs an MPDU by means of frame aggregation and fragmentation. Let us define $S_{p,j}$ as the size of node $i$’s packet and $S_{fb}$ as the size of a fragment subheader and FCS. We also define $S_{frag}^{max}$ as the maximum size of a fragment. Here, we make two assumptions. Firstly, the sender has a sufficient number of packets to aggregate or fragment. Secondly, the value of $T_{min}$ in (7) is large enough to ensure $S_{fb,j}$ is the maximum size of a frame. Here, we consider two possible cases depending on the sizes of $S_{fb,i}$, $S_{p,j}$, and $S_{frag}^{max}$: the implementation of the frame F/A scheme is slightly different in these two cases.

Case 1 ($S_{fb,i} \geq S_{frag}^{max} > S_{p,j}$). If node $i$ has no remaining fragment belonging to the previous packet at the time of transmission, it aggregates several packets without fragmentation and an extra fragment to match the size of the frame body with the desired value of $S_{fb,i}$. Otherwise, the sender aggregates the leftover fragment of the previous packet and the next several packets and fragments until they fully occupy the frame body. In this case, the first and last fragments usually differ from the others in size.

Case 2 ($S_{fb,i} > S_{p,j} > S_{frag}^{max}$). In this case, all the packets are fragmented to comply with the maximum fragment size, and then, the fragments are aggregated up to the value computed in (11). The fragment size may differ in this case.

A larger $S_{frag}^{max}$ produces less fragments for a given $S_{fb,i}$, and thus, the overhead of the fragment subheader and FCS becomes smaller. However, a smaller $S_{frag}^{max}$ improves the reliability of fragment transmission and decreases retransmissions of corrupted fragments. Therefore, the value of $S_{frag}^{max}$ should be chosen carefully to balance efficiency and reliability in order to ensure the maximization of throughput. (It is analytically shown in [28] that the fragments ranging in
size from 200 to 500 bytes give near-optimal throughput performance for the typical configurations of IEEE 802.11 WLANs.)

The frame F/A scheme in the SMAF MAC seems to be similar to the A-MPDU [27] and AFR scheme [28]. However, our scheme is more flexible and elaborate than the other schemes based on the following points. Our scheme combines both the frame aggregation scheme and fragmentation scheme in a unified framework with minimal header overhead. Moreover, in contrast to the AFR, our scheme supports fragments of variable size so as to minimize the waste of wireless channel resources due to zero-padding. In addition to enhancing the channel efficiency and transmission reliability as the other fragmentation/aggregation schemes, our scheme plays the key role of implicit signaling for the BTR, with which the fairness can be significantly improved at minimal cost.

5. Several Issues in Applying SMAF MAC

5.1. Handling the Hidden Node Problem. It is essential to consider the hidden node problem in ad hoc networks, because hidden nodes result in a significant degradation of network performance. Let us denote node $h_i (\notin S_i)$ as a hidden node to node $i$; that is, node $h_i$ cannot sense the transmissions of node $i$. Since it can neither deduce $\hat{p}_i$ nor adjust its CW in accordance with (9), it may interfere with the transmissions of node $i$ to cause transmission failure and degradation of airtime fairness. In order to relieve this problem, we employ the well-known hidden node resolution mechanism, that is, RTS/CTS exchange. However, the RTS/CTS exchange involves large overhead, so it is imperative to activate the RTS/CTS only when hidden nodes are detected and its activation contributes to improving the channel efficiency despite the overhead. For this purpose, we extrapolate the SMAF MAC to increase its breadth of applications by incorporating hidden node detection and resolution mechanisms [29, 30] that were proposed by our research group.

5.1.1. Hidden Node Detection and Resolution. The frame loss in IEEE 802.11 WLAN occurs due to several causes such as collision, interference, and channel error. In [29], the frame losses were categorized into two types, entire and partial, depending on their cause, and a hidden node detection mechanism was proposed by using the features of frame aggregation and block ACK. A frame loss is denoted as entire loss (EL) if the whole frame including the PHY header is not successfully decoded. This loss is caused mainly because there is a transmission collision or a sender fails to sense the channel occupation by hidden nodes. A receiver does not transmit an ACK frame in this case. On the other hand, if a sender receives a block ACK indicating that some consecutive subframes are corrupted, this loss is denoted as partial loss (PL) and is mainly caused by interference from hidden nodes. We define $P_{EL,i}$ and $P_{PL,i}$ as the probabilities of EL and PL measured by sender $i$. The SMAF MAC can estimate $P_{EL,i}$ and $P_{PL,i}$ by applying the method proposed in [29]. As long as both values of $P_{EL,i}$ or $P_{PL,i}$ are smaller than predetermined thresholds of $\eta_{EL}$ or $\eta_{PL}$, respectively, the SMAF MAC does not use the RTS/CTS mechanism because the RTS/CTS mechanism incurs large overhead. If $P_{EL,i} > \eta_{EL}$ or $P_{PL,i} > \eta_{PL}$, sender $i$ concludes that there exist hidden nodes and decides whether to use the RTS/CTS exchange. It is worthwhile to note that even in this case of $P_{EL,i} > \eta_{EL}$ or $P_{PL,i} > \eta_{PL}$ the RTS/CTS mechanism is used only when its gain is expected to be larger than its loss.

In order to determine whether to use the RTS/CTS mechanism on detecting the presence of hidden nodes, node $i$ in the SMAF MAC roughly estimates the effective data rates to send a frame body that can be achieved with and without the RTS/CTS mechanism, each of which is defined as $R_{RTS}^i$ and $R_{basic}^i$ in bit/sec, respectively. To make the computation simple, we exclude the backoff time in calculating $R_{RTS}^i$ and $R_{basic}^i$. We assume that the probability of transmission failure due to collision is not significant as long as the value of CW is adjusted depending on the BTR as in (9); that is, CW tends to increase/decrease as the number of nodes sharing the channel resource increases/decreases. Also, we assume that the RTS/CTS mechanism can effectively deal with the entire and partial frame losses due to hidden nodes so that $P_{EL,i} = 0$ and $P_{PL,i} = 0$ when the RTS/CTS mechanism is employed. Under these assumptions (note that we do not intend to derive an accurate throughput model but only to determine whether to use the RTS/CTS mechanism or not), we can represent $R_{RTS}^i$ and $R_{basic}^i$ as

$$R_{RTS}^i = \frac{S_{fb} \cdot 8}{T_{rts/cts} + T_{basic}}$$

$$R_{basic}^i = (1 - P_{EL,i}) (1 - P_{PL,i}) \frac{S_{fb} \cdot 8}{T_{basic}}$$

In (12), $T_{rts/cts}$ is the overhead time due to the RTS/CTS exchange, and $T_{basic}$ is the total airtime required for node $i$ to transmit an $S_{fb}$-byte MPDU without RTS/CTS exchange, which includes several overhead times due to SIFS, DIFS, and ACK transmission. Finally, sender $i$ decides to use the RTS/CTS mechanism if $R_{RTS}^i > R_{basic}^i$.

5.1.2. Extending the Effective CTS Range. Consider the case where node $i$ transmits a data frame to node $j$ according to the RTS/CTS mechanism. Let us denote nodes $h_{i,1}$ and $h_{i,2}$ as hidden nodes of node $i$ and consider that $h_{i,1}$ is located within the transmission range of the CTS frame sent by node $j$ while $h_{i,2}$ is located out of this range, but within the carrier-sensing range of node $j$. Then, node $h_{i,1}$ can overhear the CTS frame and defers its transmission until the time indicated in the CTS frame. However, node $h_{i,2}$ can sense the transmission of the CTS frame but cannot correctly decode it and defers channel access for an extended interframe space (EIFS), which is generally not long enough for node $j$ to successfully receive a data frame from node $i$ without interference by the neighbor nodes.

To cope with this problem, we devise a scheme that can extend the effective CTS range and can be incorporated simply into the SMAF MAC. Here, we need to consider
two points. The first one is how node $h_{i,2}$ can identify the CTS frame by only sensing the transmission of a frame, and the second one is how long node $h_{i,2}$ should defer its channel access to avoid interference with node $j$. The first issue can be resolved by the fact that each control frame in IEEE 802.11 WLAN has a different size, similar to the idea in [30, 31]. Upon sensing a frame transmitted by node $j$, node $h_{i,2}$ can differentiate whether it is a data frame or a control frame according to the transmission duration. If node $j$ is the receiver, it may transmit either a CTS frame or an ACK frame. Due to the difference between their transmission durations, node $h_{i,2}$ can differentiate the CTS frame from the ACK frame. (The size of the CTS frame is 14 bytes per IEEE 802.11 standard, and that of the ACK frame is 30 bytes when the size of Fragment ACK bitmap is designed to be 16 bytes. The difference in transmission durations between these two control frames is about 17 $\mu$s when their transmission rate is 6.5 Mb/s. Considering that the time required for CCA to sense the channel in IEEE 802.11 is usually 4 $\mu$s, it is feasible to differentiate the CTS frame from the ACK frame based on the difference in transmission duration.) Next, we discuss the second issue. Let us define $T_{\text{def}}$ as the time during which node $h_{i,2}$ defers channel access after implicitly identifying a CTS frame. The transmission duration $T_{\text{tx},j}$ in the SMAF MAC can be used in determining the value of $T_{\text{def}}$. According to (7), $T_{\text{tx},j}$ ranges between $T_{\min}$ and $T_{\min} + 100 \cdot \Delta$. Thus, by setting $T_{\text{def}} = T_{\min} + 100 \cdot \Delta$, the upcoming data transmission by node $i$ can be protected from possible interference by node $h_{i,2}$. However, this conservative approach may lead to the wastage of channel resources if $\tilde{\rho}_i \ll 1$. Note that the hidden node problem occurs only when there exist at least two concurrent flows in a network. Therefore, as long as there exist hidden nodes that can result in mutual interference, the ideal value of $\tilde{\rho}_i$ can be at most 0.5 when the channel is shared according to the max-min fairness. Accordingly, we set the value of $T_{\text{def}}$ as $T_{\min} + 50 \cdot \Delta$. In this way, even though node $h_{i,2}$ is out of the transmission range of the CTS frame, it can avoid interfering with the transmission of node $i$.

5.2. Handling Transmission Collision and Overlap. Figure 4 depicts the cases of transmission collision and overlap, which may impede the correct deduction by a node of the BTR of its neighbor nodes. The dotted circle in Figure 4 indicates the carrier-sensing range of each node. The transmission collision occurs when two or more nodes that are in the carrier-sensing range of each other start to transmit data frames at the same time. On the other hand, the transmission overlap occurs when nodes are out of the carrier-sensing range of each other and their transmission durations overlap. Although the transmission collision and overlap are basically handled by adjusting CW in (9) or exchanging RTS/CTS as described in Section 5.1, we discuss how the SMAF MAC can effectively resolve this problem.

The transmission collision can be easily handled because the node can determine the BTR of a neighbor node whose transmission duration is the longest among the simultaneous transmissions. For example, as shown in Figure 4(a), intermediate node 2 can adjust its CW value based on $\max(\tilde{\rho}_i, \tilde{\rho}_j) = \tilde{\rho}_j (<1)$. Recall that when adjusting CW, a node does not need to identify the transmitting node.

Next, we focus on the transmission overlap. Let us define $T_{\text{tx},j}$ as the overlapped transmission duration of neighbor nodes, which is measured by node $i$ with the carrier-sensing mechanism, and define $\bar{\rho}_i$ as the estimated BTR based on $T_{\text{tx},j}$ from (7); that is,

$$\bar{\rho}_i = \frac{T_{\text{tx},j} - T_{\min}}{100 \cdot \Delta}.$$  \hspace{1cm} (13)

As long as $\bar{\rho}_i \leq 1$, node $i$ adjusts its CW according to (9). If $\bar{\rho}_i > 1$, that is, $T_{\text{tx},j} > T_{\min} + 100 \cdot \Delta$, node $i$ does not change its CW to avoid abnormal operation of SMAF MAC. The nodes in the SMAF MAC can handle the unfairness...
problems caused by the transmission overlaps in a distributed way according to the following steps.

1. The node that frequently observes transmission overlaps, for example, node 2 in Figure 4(b), is likely to starve, because it has to defer channel access until the channel becomes idle; node 2 is exposed to both nodes 1 and 3, a well-known problem often described as the starvation problem in ad hoc networks [5]. Then, $\hat{\rho}_2$ becomes small, which is announced to neighbor nodes 1 and 3.

2. The neighbor nodes find that $\hat{\rho}_2$ is smaller than their own BTRs, and they yield their transmission opportunities to starving node 2, by increasing the value of their CWs.

3. The increase of CW in the neighbor nodes, in turn, contributes to decreasing the probability of transmission overlap. Then, node 2 will more likely deduce $\hat{\rho}_1$ or $\hat{\rho}_3$ without transmission overlap. If node 2 finds that its BTR is smaller than those of neighbor nodes, it will increase its transmission attempt by decreasing its CW.

5.3. Handling Unsaturated Nodes. Until now, we have assumed that each node always has a sufficient number of packets to construct an aggregated frame of the frame body size $S_{b,j}$. In this subsection, we discuss how an unsaturated node operates in the SMAF MAC when it does not have a sufficient number of packets to construct an intended aggregated frame.

Let $T_{\text{us}}^{b,j}$ be the transmission duration of a frame in which unsaturated node $i$ transmits. Then, due to a lack of packets available for aggregation, $T_{\text{us}}^{b,j}$ becomes smaller than the desired value of $T_{\text{tx}}^{b,j}$ to correctly announce its BTR. We consider two cases: (i) $T_{\text{tx}}^{b,j} < T_{\text{min}}$ and (ii) $T_{\text{tx}}^{b,j} \geq T_{\text{min}}$. In the first case ($T_{\text{tx}}^{b,j} < T_{\text{min}}$), node $i$ transmits data frames by aggregating only the packets backlogged in its transmission buffer. Then, its neighbor node $j$ finds $\hat{\rho}_j$ to be smaller than zero from (13) when it detects the transmission duration of $T_{\text{tx}}^{b,j} (< T_{\text{min}})$. If $\hat{\rho}_j < 0$, node $j$ does not change its CW in order to avoid an abnormal operation and channel underutilization.

Next, we consider the second case where $T_{\text{us}}^{b,j} \leq T_{\text{tx}}^{b,j} < T_{\text{tx}}^{b,j} (= T_{\text{min}} + [\hat{\rho}_j \cdot 100] \cdot \Delta)$. The approach in the first case is not applicable to this case. In contrast to the first case, the transmission with a duration of $T_{\text{us}}^{b,j} (< T_{\text{tx}}^{b,j})$ makes the neighbor nodes deduce the BTR of the unsaturated node $i$ unnecessarily smaller than the intended BTR. In order to prevent the dissemination of wrong information on $\hat{\rho}_i$, we propose two approaches. The first one is to preclude the unsaturated node from transmitting a frame until it accumulates sufficient packets to construct a frame that is $S_{b,j}$ in size. The second one is to fill the remaining space in the aggregated frame with zeros. There is a tradeoff between transmission delay and channel efficiency in these two approaches: the first one increases the transmission delay without wasting channel resources, while the second one can avoid extra transmission delay at the cost of channel wastage due to zero-padding. We propose the adoption of both approaches with a tunable design parameter, $\gamma$ ($>0$).

If $(T_{\text{tx}}^{b,j} - T_{\text{us}}^{b,j}) > \gamma$, the first approach is used to avoid unnecessary wastage of channel resources. Otherwise, the wastage due to zero-padding is insignificant and the second approach is used to avoid additional transmission delay.

5.4. Enhancing Channel Utilization. The basic operation of SMAF MAC improves the max-min airtime fairness by controlling the channel access probability so that all the neighbor nodes attain comparable BTRs. However, this approach may lead to the problem of channel underutilization when each node has quite a different number of neighbor nodes. For example, consider the configuration shown in Figure 4(b). Assume that there exists only one transmitting node within each of the left and center circles, while there are $m$ ($\geq 1$) nodes in the right circle. In this example, the node in the center circle should compete for the channel access with $m + 1$ neighbor nodes (one in the left circle and $m$ in the right circle), and its BTR will decrease as $m$ increases. Then, the SMAF MAC according to (8) and (9) will restrict the BTR of the node in the left circle from increasing. In this case, the channel for the left circle will be underutilized.

To cope with such a channel underutilization problem, we modify the SMAF MAC as follows. Let us define $\eta_i[r]$ as the ratio of total busy time due to the transmission by node $i$ and its neighbor nodes during the $r$th estimation period of node $i$; that is, $\eta_i[r] = a_i[r] + b_i[r]$. Similar to (6), the estimate of $\eta_i$, denoted as $\hat{\eta}_i$, is obtained as the weighted moving average as

$$\hat{\eta}_i[r] = w \cdot \hat{\eta}_i[r-1] + (1 - w) \cdot \eta_i[r].$$

Also, we introduce a parameter $\zeta$ ($0 < \zeta < 1$) that indicates channel underutilization. When $\hat{\eta}_i[r]$ is less than $\zeta$, node $i$ considers that the channel is underutilized and increases its channel access probability by the amount of $\lambda$, regardless of the BTR of neighbor nodes. Therefore, the control rule of CW is modified as

$$\text{CW}_i \leftarrow \begin{cases} \max \left(\frac{2(CW_i + 1)}{2 + \lambda(CW_i + 1)} - 1, CW_{\text{min}}\right), & \text{if } \hat{\eta}_i < \zeta, \\ \text{adjust } CW_i \text{ according to (9)}, & \text{otherwise}. \end{cases}$$

Consequently, the SMAF MAC can improve channel utilization while attaining acceptable max-min fairness.

6. Performance Evaluation

6.1. Simulation Configuration. We implemented the SMAF MAC in ns-2 simulator [18] and set the values of IEEE 802.11 PHY/MAC parameters with a channel bandwidth of 20 MHz in a 5 GHz frequency band and one spatial stream, according to the IEEE 802.11n standard [27]. The packet size was set to 1 Kbyte. The transmission rates of the data frame and control frame (e.g., RTS, CTS, and ACK frame) were set to 65 Mb/s and 6.5 Mb/s, respectively. In order to focus on the performance in a simple scenario, firstly we intentionally discounted the link adaptation mechanism and channel error.
in Sections 6.2–6.4. Later, we considered more realistic scenarios where each node had different transmission rates and transmission ranges with possible transmission failure due to channel error in Section 6.5. The transmission power and carrier-sensing threshold were set to 20 dBm and −82 dBm, respectively. We considered the path-loss model and the minimum receiver sensitivity specified in the TGn channel model [24] and the IEEE 802.11n standard, respectively. We considered several node topologies, including line and random topologies. We considered that each node always had a sufficient number of packets to transmit, except for the simulation with unsaturated nodes.

The parameters for the SMAF MAC were set as follows. The values of \( T_{\text{min}} \) and \( \Delta \) in calculating the transmission duration by (7) were set to 1.3 ms and 4 \( \mu \)s, respectively, by considering the typical channel coherence time and clear channel assessment time in IEEE 802.11n [24, 32]. The values of \( \lambda \) and \( \mu \), the AIMD parameters used to control the transmission attempt in (8), were set to 0.002 and 1/1.2, respectively. These values of \( \lambda \) and \( \mu \) result in the decrease and increase of CW by 3%–20% and 20%, respectively, for values of CW ranging between 32 and 256. The maximum fragment size, \( S_{\text{frag}}^{\text{max}} \), was set to 256 bytes, according to the analytic result in [28]. The thresholds used to determine the presence of hidden nodes, \( \eta_{\text{EL}} \) and \( \eta_{\text{PL}} \), were set to 0.1 and 0.05, respectively, as recommended in [29]. The values of \( \gamma \) and \( \zeta \), the design parameters used to deal with an unsaturated node and underutilization, were set to 50 \( \mu \)s and 0.5, respectively, by considering the typical configuration of ad hoc networks and parameters of the SMAF MAC. The value of \( \omega \), the coefficient for the weighted moving average in (6), was set to 0.8.

We investigate the performance of SMAF MAC in various aspects by introducing the following performance indices.

(i) **Airtime Fairness Index (J).** This is a network-wide measure for airtime fairness defined in (2). It ranges between \( 1/N \) (the worst case where only one node monopolizes the whole airtime) and one (the best case where all nodes have the same airtime).

(ii) **Channel Utilization (C ut).** Per-node channel utilization for node \( i \), denoted as \( C_{\text{ut}}^{i} \), is the ratio of busy time due to the transmissions of node \( i \) and its neighbor nodes with respect to the total airtime multiplied by the ratio of node \( i \)'s successful transmission time with respect to its transmission time. The network-wide channel utilization \( C_{\text{ut}} \) is obtained as the average value of \( C_{\text{ut}}^{i} \).

(iii) **Throughput (TH).** This is the sum of per-node throughput, that is, the total amount of packets transmitted successfully by all the nodes divided by the total simulation time. Note that the value of TH can be higher than the transmission rate of the data frame due to spatial reuse.

(iv) **Per-Flow Delay (D).** This is the average MAC-to-MAC delay per flow. The MAC-to-MAC delay of a packet is the time taken from the packet arrival in the sender’s MAC layer to its successful reception in the receiver’s MAC layer. This delay includes queuing delay, fragmentation/aggregation delay, backoff time, and retransmission delay.

6.2. Performance Validation of SMAF MAC in a Simple Scenario. Here, we focus on validating the performance of SMAF MAC in a simple network scenario. We compare several performance indices of SMAF MAC with those of the conventional MAC protocol of IEEE 802.11, named **BASIC**. (We have considered other MAC protocols besides DCF for possible performances comparison, but we have not found a suitable protocol that meets the various demands in the following simulation studies. Also, the implementation of other MAC protocols for the various simulation environments in this section requires too much modification so they may be quite different from the original ones and are not suitable for fair performances comparison.) For a fair comparison with SMAF and BASIC, the frame aggregation scheme, that is, A-MPDU, is employed in BASIC where several packets are aggregated to make the transmission time of the aggregated frame comparable to \( T_{\text{tx}} \) in the SMAF MAC. The hidden node detection and resolution mechanism proposed in Section 5.1 is not implemented in this simulation.

6.2.1. Improvement of Max-Min Airtime Fairness. First, we consider a simple scenario where there are three senders as in Figure 4(b). Let us denote node \( \text{C} \) as the node (node 2 in Figure 4(b)) located in the intersection of the carrier-sensing ranges of the left/right-side nodes, each of which is denoted as node \( L \), and node \( R \), respectively. In this configuration, node \( C \) is exposed to both node \( L \) and node \( R \) and suffers from the starvation problem, while node \( L \) is hidden from node \( R \) and vice versa.

Figure 5 compares per-node throughput and total throughput of BASIC with those of SMAF. In the case of BASIC, both node \( L \) and node \( R \) achieve similar throughput (approximately 50 Mb/s), which is close to the ideal value.

![Figure 5: Throughputs comparison for the starvation scenario in Figure 4(b).](image-url)
that can be achieved with the transmission rate of 65 Mb/s and the A-MPDU scheme. However, node_1 completely starves; that is, its throughput is only about 2.8 Mb/s, which is about 18 times smaller than those of node_2 and node_3. Moreover, $\bar{T}$ is 0.698 which is quite close to 0.667, corresponding to the case where node_1 and node_3 are isolated. The mechanism of this unfairness can be explained as follows. Since node_1 and node_3 can access the channel without any contention, it has much less chance to access the channel and a much higher probability of transmission collision, and thus, it has significantly lower throughput compared to node_2 and node_3. However, SMAP effectively solves the fairness problem. As shown in Figure 5, the throughput of node_2 is remarkably increased (by about seven times compared to that of BASIC), and it is larger than half of the throughput of node_1 or node_3. According to (8), each node in SMAP adjusts its transmission attempt probability so that its airtime increases up to the max-min fair share. The value of $\bar{p}$ for each node can go up to 0.5 under an ideal circumstance where the transmission time of node_1 perfectly coincides with that of node_3. However, this ideal circumstance rarely exists because node_1 and node_3 are hidden from each other and they transmit their frames independently. Therefore, a fair and feasible value of $\bar{p}$ for node_1 is between 1/2 and 1/3, which may result in the underutilization of the channel to enforce the max-min fairness. However, the SMAP MAC does not degrade channel utilization because node_1 and node_3 are allowed to consume more airtime without decreasing the BTR of node_1 as long as it is less than $\zeta$ (= 0.5) for both nodes (see (15)). As a result, the airtime can be spatially reused by node_2 and node_3, and channel sharing achieved by SMAP becomes much fairer (i.e., $\bar{T} = 0.945$) while maintaining high channel utilization (i.e., $\bar{C}_\text{out} = 0.902$). The results in Figure 5 confirm a serious starvation problem in BASIC and notable improvement of fairness in SMAP.

6.2.2. Effect of the Number of Nodes. Next, we investigate how the number of nodes affects the performance of BASIC and SMAP. We consider the same topology in Figure 4(b). Let us denote $R_{\text{in}}$ as the intersection area of the left and right circles in Figure 4(b) and $R_{\text{out}}$ as the area within the center circle excluding $R_{\text{in}}$. We randomly placed $N_{\text{in}}$ and $N_{\text{out}}$ pairs of transmitters and receivers in the areas of $R_{\text{in}}$ and $R_{\text{out}}$, respectively. The hidden node problem may occur depending on the configuration of nodes. We performed twenty simulations with random placements of nodes and represented the average values of $\bar{T}$, $\bar{C}_\text{out}$, and $\bar{D}$ in Figure 6, where $N_{\text{in}}$ ranges from 1 to 10 and $N_{\text{out}}$ is fixed to 2.

Figure 6(a) shows three different values of $\bar{T}$: two are obtained from simulations for BASIC and SMAP, and the other one, labeled as fair monopoly, is added to represent a case of extreme unfairness, which is calculated under the assumption that the whole channel resource is completely and evenly monopolized by two nodes in $R_{\text{out}}$ so that no resource is available for the remaining $N_{\text{in}}$ nodes. SMAP maintains $\bar{T}$ close to one, regardless of $N_{\text{in}}$. However, $\bar{T}$ of BASIC decreases as $N_{\text{in}}$ increases, and it is slightly larger than that of fair monopoly, implying severe unfairness. Figure 6(b) shows the total throughput ($\bar{C}_\text{out}$) of BASIC and SMAP, along with the throughput achieved by the $N_{\text{in}}$ nodes, denoted as $\bar{TH}_{\text{in}}$. Due to the monopolization of channel resources by the nodes in $R_{\text{out}}$, $\bar{TH}_{\text{in}}$ of BASIC is at most 11 Mb/s for the whole range of $N_{\text{in}}$. In the case of SMAP, $\bar{TH}_{\text{in}}$ increases from 19 Mb/s to 44 Mb/s as $N_{\text{in}}$ increases from 1 to 10, which is larger than that of BASIC by 3.9–6.6 times. However, the total throughput of SMAP is smaller than that of BASIC by 20%–37%, which is inevitable to ensure max-min fairness in the SMAP MAC. Next, we compare the channel utilization and delay in Figure 6(c). In the case of BASIC, $\bar{C}_\text{out}$ decreases from 0.91 to 0.73 as $N_{\text{in}}$ increases, whereas it is higher than 0.85 for all values of $N_{\text{in}}$ in the case of SMAP. Moreover, SMAP maintains $\bar{D}$ smaller than that of BASIC by 3.5–4.0 times, and the delay of SMAP increases more slowly than that of BASIC as $N_{\text{in}}$ increases.

We repeated the simulation under the configuration where $N_{\text{in}}$ is fixed to two, but $N_{\text{out}}$ increases from 2 to 10. As shown in Figure 7(a), BASIC still attains airtime fairness at a satisfactory level; that is, $\bar{T} > 0.94$ for $2 < N_{\text{out}} < 10$. In contrast to the result in Figure 6(a), $\bar{T}$ of BASIC is lower than that of fair monopoly, which means that there is a serious unfairness problem among the $N_{\text{out}}$ nodes in $R_{\text{out}}$, as well as the nodes in $R_{\text{in}}$ and $R_{\text{out}}$. Furthermore, the starvation problem in BASIC becomes exacerbated as $N_{\text{out}}$ increases; that is, $\bar{TH}_{\text{in}}$ is only 0.3 Mb/s when $N_{\text{out}} = 10$. Unlike BASIC, $\bar{TH}_{\text{in}}$ of SMAP is 4.7 Mb/s when $N_{\text{out}} = 10$, 15 times higher than that of BASIC. Figure 7(c) shows that $\bar{C}_\text{out}$ of SMAP is almost immune to the increase of $N_{\text{out}}$ and is at least 0.84, but that of BASIC decreases as $N_{\text{out}}$ increases and drops below 0.75 when $N_{\text{out}} = 10$. The superior performance of SMAP in terms of $\bar{D}$ is still maintained for various values of $N_{\text{out}}$. In summary, these simulation results in Figures 6 and 7 clearly validate the airtime fairness of the SMAP MAC as well as its high channel utilization and small MAC-to-MAC delay.

6.2.3. Effect of $\zeta$ on the Performance of SMAP MAC. The parameter $\zeta$ was introduced to improve channel utilization of SMAP MAC when the nodes are not evenly distributed. Figure 8 shows several performance indices of SMAP MAC when the number of node $m$ ranges from 3 to 10 for several values of $\zeta$ (= 0.2, 0.5, and 0.8) under the configuration discussed in Section 5.4. Figure 8(a) shows that as $m$ increases from 3 to 10, that is, the degree of asymmetry in node distribution increases, $\bar{T}$ decreases from 0.995 and 0.989 to 0.971 and 0.936 when $\zeta = 0.2$ and 0.5, respectively. On the other hand, as the value of $\zeta$ increases, the node that perceives channel underutilization accesses the channel more aggressively; and thus, the value of $\bar{T}$ decreases. When $\zeta = 0.8$, $\bar{T}$ is decreased by about 7% and 3% compared to the cases of $\zeta = 0.2$ and 0.5, respectively. However, $\bar{T}$ still remains higher than 0.91 for the whole range of $m$ even when $\zeta = 0.8$. On the other hand, as shown in Figures 8(b) and 8(c), the increase of $\zeta$ contributes to throughput and channel utilization. When $\zeta = 0.8$, the total throughput is increased by up to about 41%
Figure 6: Performance indices for various values of \( N_{in} \).

Figure 7: Performance indices for various values of \( N_{out} \).
and 21%, compared to when $\zeta = 0.2$ and 0.5, respectively. Also, the increase of $\zeta$ from 0.2 and 0.5 to 0.8 improves the channel utilization by up to 6% and 4%, respectively. From the results in Figure 8, we can conclude that the introduction of $\zeta$ is effective in improving the efficiency of channel usage while maintaining acceptable airtime fairness.

6.3. Max-Min Airtime Fairness in a Dynamic Scenario with Unsaturated Nodes. We further validate the performance of SMAF MAC in two dynamic scenarios: one where the number of nodes changes with time and the other where there exists an unsaturated node whose traffic load also changes with time. We consider the same topology used in Section 6.2.

In the first scenario, $N_{\text{out}}$ was set to two and $N_{\text{in}}$ was initially set to one and then increased by one every 5 sec to a value of five and then decreased from five to one every 5 sec. Figure 9(a) compares $\bar{f}$ of BASE and SMAF. In the case of SMAF, $\bar{f}$ is hardly affected by the change of $N_{\text{in}}$; it ranges mostly between 0.94 and 1.0, regardless of the changes in $N_{\text{in}}$. However, $\bar{f}$ of BASIC considerably decreases/increases in response to the increase/decrease of $N_{\text{in}}$; for example, $\bar{f}$ drops below 0.4 during $t = [15, 30]$ s where $N_{\text{in}} \geq 4$. The result of TH shows the opposite behavior compared to $\bar{f}$. As shown in Figure 9(b), TH of SMAF changes as $N_{\text{in}}$ changes, to attain max-min airtime fairness. On the other hand, TH of BASIC is little affected by the change in $N_{\text{in}}$, because the nodes in $R_{\text{in}}$ almost starve, regardless of its value, and the throughput is mostly achieved by the nodes in $R_{\text{out}}$. These results reconfirm the tradeoff between fairness and throughput.

In the second scenario, we set $N_{\text{out}}$ and $N_{\text{in}}$ to two and six, respectively. There is one unsaturated node among the six $N_{\text{in}}$ nodes, whereas the other nodes are saturated. Let us denote node$_{\text{in,us}}$ as the unsaturated node among $N_{\text{in}}$ nodes and node$_{\text{in,j}}$ and node$_{\text{out,j}}$ as saturated node $i$ ($i = 1, 2, \ldots, 5$) and node $j$ ($j = 1, 2$) in $R_{\text{in}}$ and $R_{\text{out}}$, respectively. The offered traffic load of node$_{\text{in,us}}$ changes with time; that is, it increased from 1 Mb/s to 5 Mb/s every 5 sec and then decreased from 5 Mb/s to 1 Mb/s every 5 sec. Figure 10(a) shows per-node throughput of node$_{\text{in,us}}$, along with those of node$_{\text{in,1}}$, node$_{\text{out,1}}$, node$_{\text{in,2}}$, and node$_{\text{in,2}}$ as typical of the $N_{\text{in}}$ nodes. As shown in Figure 10(a), node$_{\text{in,us}}$ uses sufficient channel resources to transmit its packets so that its throughput is almost identical to the offered traffic load. The other saturated nodes share the remaining resources in a fair way: (i) the throughput of node$_{\text{out,1}}$ is comparable to that of node$_{\text{out,2}}$ and there is also no significant difference between the throughputs of node$_{\text{in,1}}$ and node$_{\text{in,2}}$; (ii) the per-node throughput of node$_{\text{out,j}}$ is notably higher than that of node$_{\text{in,j}}$ because SMAF tries to utilize the available channel resources as much as possible while maintaining fairness among the nodes. Figure 10(b) shows TH and $\bar{f}$ of SMAF; TH ranges between 57 Mb/s and 63 Mb/s and $\bar{f}$ is mostly higher than 0.94 for the total simulation time. Both TH and $\bar{f}$ are almost immune to the presence of an unsaturated node with varying traffic load.

6.4. Performances Comparison under Line Topology. This simulation focuses on the performance of SMAF MAC when

![Figure 8: Performance indices for several values of $\zeta$ and $m$.](image-url)
hidden nodes and exposed nodes coexist under the line topology as illustrated in Figure 11. In addition to BASIC and SMAF, we consider two additional schemes, called HDR and SMAF+. In HDR, the hidden node detection and resolution mechanisms proposed in [29] are incorporated into BASIC. Moreover, SMAF+ is an enhanced version of SMAF that employs the selective RTS/CTS exchange mechanism and the extended CTS range described in Section 5.1.

We consider three representative patterns of flows as shown in Figure 11. The transmission rate of all the nodes was set to 26 Mb/s. Table 1 lists several performance indices for these three cases. In the case of Line (i), node TX2 is expected to suffer from the starvation problem because it is exposed to both nodes TX1 and TX3, similar to the case in Figure 4(b). The performances of BASIC and HDR are quite similar in all aspects of TH, J, and CWT, and the throughput of flow 2 is much lower than that of the other flows. On the other hand, both SMAF and SMAF+ increase the throughput of flow 2 by about five times, compared to those of BASIC and HDR. Moreover, there is little difference in the performance indices between SMAF and SMAF+, which implies the selective RTS/CTS mechanism was not activated most of the time in SMAF+, because the expected gain was estimated to be lower than the loss.

In the case of Line (ii), node TX2 is exposed to node TX3, but it is hidden from node TX1, and thus, the transmission of node TX2 can be disrupted by the transmission of node TX1. This configuration exacerbates the starvation problem in BASIC; that is, the throughput of flow 2 is almost zero, because the transmission failure of node TX2 due to the interference from the hidden node leads to the increase of CW according to the binary exponential backoff mechanism of DCF, which further deprives node TX2 of transmission opportunities. Also, the transmission failure due to interference decreases
Table 1: Performance indices for the three flows in Figure 11.

<table>
<thead>
<tr>
<th>Topology Scheme</th>
<th>Throughput (Mb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow 1</td>
<td>Flow 2</td>
</tr>
<tr>
<td><strong>Line (i)</strong></td>
<td><strong>BASIC</strong></td>
</tr>
<tr>
<td></td>
<td><strong>HDR</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SMAF</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SMAF+</strong></td>
</tr>
<tr>
<td><strong>Line (ii)</strong></td>
<td><strong>BASIC</strong></td>
</tr>
<tr>
<td></td>
<td><strong>HDR</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SMAF</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SMAF+</strong></td>
</tr>
<tr>
<td><strong>Line (iii)</strong></td>
<td><strong>BASIC</strong></td>
</tr>
<tr>
<td></td>
<td><strong>HDR</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SMAF</strong></td>
</tr>
<tr>
<td></td>
<td><strong>SMAF+</strong></td>
</tr>
</tbody>
</table>

**Figure 11:** Three patterns of flows in a line topology.

Line (ii) (or Line (iii)), upon sensing the CTS frame sent by node RX2, node TX1 (or TX3) defers its channel access, and thus, it precludes interference with the transmission by node TX2. The throughput of flow 2 in SMAF+ is increased by more than two orders of magnitude compared to those of the other schemes, and \( \bar{J} \) of SMAF+ is larger than 0.997. The extended CTS range, together with the hidden node detection/resolution mechanism in Section 5.1, significantly contributes to the superior performance of the SMAF MAC even in the presence of hidden/exposed nodes.

6.5. Performances Comparison under Random Topologies and Different Data Transmission Rates. Finally, we investigate and compare the performances of four MAC mechanisms in a more realistic scenario where (i) 20 pairs of senders and receivers are randomly located in a square area of 500 m × 500 m; (ii) the transmission rate of each flow is selected from 6.5, 13, 52, and 65 Mb/s depending on the distance between the sender and receiver; and (iii) the bit error rate of the channel is set to 5 × 10^{-5}. This scenario causes several potential problems including hidden node, starvation, and performance anomaly problems, which result in a significant degradation of network performance.

Figure 12 shows the cumulative distributions of several performance indices obtained from 30 instances of simulations. We can observe from Figure 12(a) that SMAF and SMAF+ ensure the airtime fairness quite strictly, regardless of the random placement of nodes and the difference in transmission rates. For example, even the 10-percentile value of \( \bar{J} \) in SMAF and SMAF+ is larger than the 90-percentile value of \( \bar{J} \) in BASE and HDR by at least 1.7 and 1.6 times, respectively. The superior performance of SMAF and SMAF+ in terms of channel utilization can also be confirmed from Figure 12(b). In the cases of BASIC and HDR, the median values of \( \bar{C}_{ut} \) are only 0.22 and 0.26, respectively; that is, the channel resource is largely wasted by unsuccessful transmissions and long backoff time. Compared to BASIC, the median value of \( \bar{C}_{ut} \) is increased by about 2.2 and 2.8 times in SMAF and SMAF+, respectively. Next, we compare the aggregate throughputs for several mechanisms in Figure 12(c). Compared to BASIC...
The deviation of TH is much smaller in SMAF and SMAF+. HDR shows the best performance in terms of the median value of TH, though it is achieved at the significant expense of fairness as shown in Figure 12(a). The 10-percentile value of TH in SMAF and SMAF+ is larger than that in BASE and HDR by up to about 1.8 and 1.5 times, respectively. This is because the SMAF MAC effectively deals with the performance anomaly problem in the multirate WLANs, which decreases the aggregate throughput [16]. Figure 12(d) shows the outstanding performance of SMAF+ in terms of per-flow delay. Even the 90-percentile value of $D$ in SMAF+ is lower than the median values of $D$ in BASIC and HDR by two orders of magnitude. Consequently, as shown in Figure 12, SMAF+ provides excellent performance in terms of airtime fairness, channel utilization, and delay, even though its throughput is somewhat lower than those of the other mechanisms. The results in Figure 12 show that the channel error and noise do not have a significant effect on the performances of SMAF and SMAF+. It is important to note that the lower throughput of SMAF+ does not mean poor channel utilization but implies that the channel is shared by low-rate nodes and high-rate nodes in a fair way according to the policy of max-min airtime fairness.

7. Conclusion

In this paper, we propose the SMAF MAC to improve max-min airtime fairness as well as channel utilization in ad hoc networks based on IEEE 802.11 WLAN. In the SMAF MAC, each node estimates the BTR and implicitly announces it to its neighbor nodes and then adjusts its contention window based on the BTR of neighbor nodes to maintain fairness without degrading channel utilization. The BTR is encoded into the transmission duration by means of frame fragmentation and aggregation schemes. It is announced without requiring any signaling messages, and its coverage can be extended beyond the transmission range. Moreover, we successfully incorporated the hidden node detection and resolution mechanisms into the SMAF MAC to minimize performance degradation due to hidden nodes. The simulation results confirmed the outstanding performance of SMAF MAC in terms of airtime fairness, channel utilization, and delay in various configurations.
configurations. We strongly believe that the SMAF MAC can be extrapolated to practical applications beyond those described herein to provide fairness even when nodes have different quality-of-service requirements, different priorities, and/or a different number of flows, all of which are topics for future research.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Research Foundation (NRF) grant funded by the Korea government (MSIP) (no. NRF-2013R1A2A2A01009883) and the Dongguk University Research Fund of 2015.

References


