Ph.D. DISSERTATION

MAC Efficiency Improvement of IEEE 802.11 Wireless Local Area Networks

무선 랜 매체접근제어 (MAC) 효율화 기법

BY

JUNSU CHOI

FEBRUARY 2017

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Ph.D. DISSERTATION

MAC Efficiency Improvement of IEEE 802.11 Wireless Local Area Networks

무선 랜 매체접근제어 (MAC) 효율화 기법

BY

JUNSU CHOI

FEBRUARY 2017

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Abstract

In this dissertation, we develop medium access control (MAC) efficiency improvement schemes for IEEE 802.11 wireless local area networks.

In part I of this dissertation, we develop a contention window (CW) control scheme for practical IEEE 802.11 wireless local area networks (WLANs) that have node heterogeneity in terms of the traffic load, transmission rate, and packet size. We introduce activity probability, i.e., the probability that a node contends for medium access opportunities at a given time. We then newly develop a performance analysis model that enables analytic estimation on the contention status including the collision probability, collision time, back-off time, and throughput with comprehensive consideration of node heterogeneity. Based on the newly developed model, we derive the theoretically ideal contention status, and develop a CW control scheme that achieves the ideal contention status in an average sense. We perform extensive NS-3 simulations and real testbed experiments for evaluation of both the proposed performance analysis model and CW control scheme. The results show that the proposed model provides accurate prediction on the contention status, and the proposed CW control scheme achieves considerable throughput improvement compared to the existing schemes which do not comprehensively consider node heterogeneity.

In part II of this dissertation, we propose a sounding control scheme for IEEE 802.11ac multi-user multiple-input multiple-output (MU-MIMO). The proposed scheme comprehensively considers the long-term characteristics of a network environment including the downlink traffic loads and channel coherence times of wireless links, and jointly determines the sounding node set and sounding interval to maximize the long-term expected MU-MIMO throughput gain in consideration of sounding overhead. To this end, we analytically formulate an MU-MIMO throughput gain maximization problem considering the network environment and sounding overhead. We conduct MIMO

channel measurement in practical WLAN environments, and evaluate the performance of the proposed scheme by employing the real channel data traces. Simulation results verify that the proposed scheme adaptively determines the sounding node set and sounding interval according to the network environment, and outperforms the existing scheme which considers the channel coherence times only.

In part III of this dissertation, we develop an adaptive group ID (GID) control scheme to mitigate idle power consumption at nodes in IEEE 802.11ac wireless local area networks (WLANs) supporting multi-user multiple input multiple output (MU-MIMO). We analytically derive the expected idle power consumption at nodes sharing common GIDs, revealing that it has relations with their downlink (DL) traffic loads. Based on the analysis, we formulate an idle power consumption minimization problem, and develop an efficient algorithm to reduce the computational complexity. Simulation results reveal that idle power consumption becomes extremely severe when an access point (AP) has a large number of associated nodes. The proposed scheme assigns GIDs in consideration of DL traffic loads, thus considerably mitigating idle power consumption compared to random GID overloading.

keywords: Wireless local area network, carrier sense multiple access with collision avoidance, multi-user multiple-input multiple-output, channel sounding

student number: 2010-20908

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The demand for wireless connectivity has been tremendously increasing with the emergence of smart mobile devices such as smart phones and tablet PCs. Along with such a trend, the IEEE 802.11 wireless local area network (WLAN) has become very popular and widely deployed not only in personal areas but also in public areas thanks to its broadband wireless connectivity, low cost, and ease of use. In this dissertation, we develop medium access control (MAC) efficiency improvement schemes for IEEE 802.11 WLANs.

1.1 Activity Probability Based Performance Analysis and Contention Control for IEEE 802.11 WLANs

In most WLANs, nodes share medium access opportunities through distributed and contention based medium access control (MAC) called Distributed Coordination Function (DCF). In DCF, the contention level is adjusted using the contention window (CW), and hence, the efficiency of DCF heavily depends on the CW size, i.e., CW control has critical importance for WLAN performance. However, analytical derivation of the best CW size that achieves the maximum throughput needs to be preceded by performance analysis modeling in that it enables analytic estimation on the contention

status including collision probability, collision time, back-off time, and throughput. However, both performance analysis modeling and CW control have not been addressed considering node heterogeneity in terms of traffic loads, transmission rates, and packet sizes.

One of the most widely referenced performance analysis models is developed in [1]. It adopts a Markov chain for back-off state modeling. That modeling enables analytical estimation on the contention status including the steady state back-off state probability distribution, collision probability, back-off time, and normalized throughput. Moreover, the optimal CW size is analytically derived based on the analytical estimation of the contention status. However, a single Markov chain describes all the nodes' back-off state probability distributions, which is inherently not able to reflect node heterogeneity. Therefore, the performance analysis modeling is based on simplifying assumptions that all nodes have the same saturated traffic load, and transmit the same sized packets using the same transmission rate, i.e., the performance analysis modeling and the optimized CW size are not valid in practical WLANs that have node heterogeneity.

There also have been many research efforts to develop performance analysis models considering non-saturated traffic condition or heterogeneous transmission rates [2–15]. However, they are also based on simplifying assumptions that nodes have the same traffic load, or the same transmission rate, i.e., node heterogeneity is not comprehensively considered. Moreover, those performance analysis models are not led to developing CW control schemes, i.e., CW control is not addressed with those models.

Run-time contention level estimation schemes have been proposed in [16, 17], which estimate the contention level by measuring the collision probability. Along with such studies, a CW control scheme has been proposed in [18], which adjusts the CW size based on the measured collision probability. However, the analysis of contention level estimation does not consider node heterogeneity either. Moreover, such an approach has limitations in that the CW size is solely determined by the collision prob-

ability. For example, a collision would result in different time loss according to node properties such as packet sizes and transmission rates, and hence, the CW size should be differently determined according to node properties for a given collision probability. Although there are other CW control schemes which adjust the CW size according to the run-time contention level [19, 20], they also have similar limitations as [18].

There have been research directions to develop new MAC that achieves better performance than DCF [21, 22]. They allow a node to determine the CW size for itself according to its transmission queue length. However, such approaches are not compliant with the WLAN standard which requires that all associated nodes shall use a common CW size announced by an AP [23]. Therefore, considering large scale adoption, those approaches are considered costly.

In this part of this dissertation, we first develop a new performance analysis model that considers node heterogeneity. Based on the newly developed model, we derive the theoretically ideal contention status, and then analytically derive a CW size that achieves the ideal contention status in an average sense.

1.2 Sounding Node Set and Sounding Interval Determination for IEEE 802.11ac MU-MIMO

The IEEE 802.11 wireless local are network (WLAN) has continuously evolved by introducing the latest communications technologies to meet the demand for larger capacity. Especially, IEEE 802.11ac [24] newly adopts downlink (DL) multi-user multiple-input multiple-output (MU-MIMO) technologies which are expected to remarkably increase capacity with the spatial multiplexing gain [25–27]. However, there are some issues to be dealt with for performance improvement through DL MU-MIMO.

For DL MU-MIMO, an AP needs to eliminate inter-user interference through precoding which requires channel state information (CSI) of wireless links from the AP to mobile nodes involved in MU-MIMO transmission. In IEEE 802.11ac, the AP per-

forms channel sounding to obtain the CSI. However, sounding is performed at the expense of precious time resources, and the time overhead increases with the frequency of sounding and the number of nodes involved in sounding. Accordingly, the AP should carefully determine which set of nodes to involve in sounding, and how frequently to perform sounding.

Most studies on IEEE 802.11ac MU-MIMO have focused on performance evaluation as in [28–31], which show how much throughput improvement DL MU-MIMO would achieve. In [32], it is shown that the sounding overhead may severely degrade the performance of a WLAN without sounding control. A heuristic algorithm is also proposed to reduce the sounding overhead by considering channel coherence times of wireless links and selectively involving nodes in sounding. However, this algorithm has limitations in that it considers channel coherence times only. Let us assume that a node has a very light DL traffic load, and the corresponding wireless link has a short channel coherence time. Under the selective sounding scheme, the AP may frequently perform sounding to have accurate CSI for that node, which only ends up wasting a lot of precious time resources with little throughput improvement due to its light DL traffic load. Therefore, sounding control requires comprehensive consideration of the network environment including not only channel coherence times, but also nodes' DL traffic loads, DL signal-to-noise ratios (SNRs) and so on.

In this part of this dissertation, we mathematically formulate a sounding node set and interval determination problem in which the long-term expected MU-MIMO throughput gain is maximized considering the network environment and sounding overhead. To the best of our knowledge, the sounding node set and interval determination problem has not been mathematically formulated in the literature. We also develop a method to reduce the computational complexity of the proposed scheme. We conduct MIMO channel measurement in practical WLAN environments, and evaluate the performance of the proposed sounding control scheme through simulations employing the collected channel data.

1.3 Adaptive Group ID Control for Idle Power Consumption Mitigation in IEEE 802.11ac WLANs

In IEEE 802.11ac, the concept of a "node group" is newly introduced for multi-user multiple input multiple output (MU-MIMO) operation [24]. It allows an access point (AP) to simultaneously transmit multiple data streams only to nodes in the same node group, which means that node groups should be formed prior to MU-MIMO transmission. Each node group consists of up to four nodes, and different node groups are distinguished using group IDs (GIDs). While GIDs are originally introduced for the aforementioned MU-MIMO operation, IEEE 802.11ac also defines a built-in power saving mechanism which mitigates idle power consumption using GIDs.

In a legacy wireless local area network (WLAN), the medium access control (MAC) layer header of a packet contains the destination node address information. Therefore, a node identifies the actual destination node of an incoming packet only after decoding the packet and identifying the destination node information at the MAC layer, which means that a node should decode all incoming packets even though not all incoming packets are destined to itself, referred to as idle-listening. Idle-listening contributes to a large portion of power consumed at a WLAN node [33], and idle power consumption is undesirable for most WLAN devices such as smart phones, and tablet PCs considering that they are battery-powered.

In an IEEE 802.11ac WLAN, the physical (PHY) layer header contains the target GID of the packet. Therefore, nodes not belonging to the target node group could ignore the incoming packet right after decoding the PHY header and identifying the target GID, thus saving power. This power saving mechanism could mitigate idle power consumption at WLAN nodes, which leads to battery-power usage time increase. However, multiple node groups share common GIDs due to the limited number of available GIDs,[1] the power saving performance is severely degraded. Therefore, we need to

[1] In IEEE 802.11ac, only 62 GIDs are available while there can be much more node groups as detailed

develop a GID control scheme which mitigates the performance degradation of the built-in power saving mechanism.

Most studies on IEEE 802.11ac MU-MIMO have focused on the performance evaluation of MU-MIMO [28–31], and signaling overheads for managing GIDs [34]. However, GID control has never been considered in terms of power saving to the best of our knowledge. In this letter, we propose a GID control scheme, where node groups are efficiently handled with the limited number of available GIDs while mitigating idle power consumption at nodes sharing common GIDs. To this end, we analytically derive the expected idle power consumption induced from shared GIDs, and formulate an idle power consumption minimization problem.

later.

Chapter 2

Activity Probability Based Performance Analysis and Contention Control for IEEE 802.11 WLANs

2.1 DCF and Contention Window control

The IEEE 802.11 WLAN defines a MAC called Distributed Coordination Function (DCF) through which nodes share medium access opportunities in a distributed manner [35]. When a node has packets to transmit in its transmission queue, it first sets a back-off counter with a value randomly chosen between 0 and the CW, where the CW normally has a predefined minimum value. Whenever wireless medium is sensed idle for longer than Distributed Interframe Space (DIFS), the counter value decreases every slot time T_{slot}. On the other hand, while the medium is sensed busy, the counter value decrease is suspended until the medium becomes idle again.

Once the counter value reaches zero, the node transmits the packet at the head of its transmission queue. If the transmission is successful, a receiver transmits an acknowledgement (ACK) after Short Interframe Space (SIFS). However, if the back-off processes of multiple nodes expire at the same slot, they transmit packets simultaneously, which results in an unsuccessful transmission. If a transmission is not successful, it is not acknowledged, and a transmitter doubles the CW as long as it is smaller than a pre-

defined maximum value. Then, the transmitter goes through the back-off process again with the increased CW. The CW returns to the minimum value after every successful transmission.

Although DCF facilitates fare sharing of medium access opportunities without central scheduling, its efficiency heavily depends on the CW size. For example, when the CW size is too small compared to the contention level, a large portion of time resources are wasted by collisions, thus severely degrading the network performance. On the other hand, if the CW size is too large compared to the contention level, a large portion of time resources are wasted by back-off, i.e., wireless medium is underutilized, which again leads to severe network performance degradation. Therefore, the CW size should be adaptively determined according to the contention level.

However, due to the distributed nature of DCF, it is extremely difficult to predict nodes' behavior in a practical WLAN where nodes have heterogeneity in terms of traffic loads, transmission rates, and packet sizes. It makes contention level estimation challenging, and hence, CW control has not been addressed with comprehensive consideration of node heterogeneity. Considering node heterogeneity is even more critical in today's WLANs with high node density, longer frames, and various transmission rates of IEEE 802.11n and 802.11ac [24, 36].

2.2 Activity Probability-Based Performance Analysis Model

2.2.1 System Description

We consider an IEEE 802.11 WLAN in which an AP has N associated nodes, and they use a common CW size notified by the AP [23]. The AP and all associated nodes can carrier-sense each other, i.e., there is no hidden node. Nodes have different link conditions according to their random location within the AP service range, thus having heterogeneous transmission rates depending on the link conditions. The n-th node's average transmission rate is denoted as r_n $(n = 1, 2, \ldots, N)$. Nodes also have hetero-

geneous MAC payload sizes and rate sources, where the n-th node's average MAC payload size and average traffic load are denoted as L_n and λ_n, respectively. Therefore, the average packet enqueque rate at the n-th node's transmission queue is given by λ_n/L_n.

DIFS and SIFS are denoted as T_{DIFS} and T_{SIFS} respectively, and the slot time is denoted as T_{slot}. When a node transmits a packet, the time duration of the PHY preamble and the MAC header length are denoted as T_{PHY} and L_{MAC} respectively. The MAC layer cyclic redundancy check (CRC) code length is denoted as L_{CRC}. Lastly, the average acknowledgement transmission time for the n-th node is denoted as $T_{\text{ACK},n}$.

In an infrastructure mode WLAN, nodes transmit packets only to the associated AP, and hence, the AP has the knowledge of node properties including average transmission rate, packet size, and traffic load by monitoring their medium access statistics. Based on the knowledge, the AP determines the CW size considering node properties, and announces it to associated nodes. In a practical WLAN, those node properties vary over time. When the AP detects any change in the node properties, it redetermines the CW size with the updated information of node properties, thus leading to adaptive CW control in a time-varying network.

This framework can be extended to a multi-AP scenario in which an AP subset and their associated nodes that can carrier-sense each other use a common CW size. Each AP determines the CW size considering node properties of its associated nodes as well as the other APs' ones within its carrier sensing range. It leads APs that carrier-sense each other to have the same node property information, thus again leading to the same CW size calculated at those APs. An example of such a multi-AP scenario could be an enterprise WLAN in which multiple APs are connected to and controlled by an AP controller. The AP controller combines node property information, and determines the CW size for an AP subset and their associated nodes within the carrier sensing range of each other. All the APs in the AP subset and their associated nodes use the common CW size notified by the AP controller.

2.2.2 Activity Probability-Based Throughput Estimation

In DCF, nodes share medium access opportunities through contention, and hence, how wireless medium is shared among nodes heavily depends on how often each node participates in contention. To model the activeness in contention, we introduce activity probabilities for nodes. The activity probability of a node is the probability that it has packets to transmit, and hence, is active in contention at a given time. We denote the n-th node's activity probability as p_n.

When a node has packets to transmit, it participates in contention, and wins medium access opportunities. The n-th node's expected throughput then is described by

$$\psi_n = p_n \times \frac{L_n}{T_{\text{exp},n}} + (1 - p_n) \times 0 = \frac{p_n L_n}{T_{\text{exp},n}}, \quad n = 1, 2, \ldots, N, \tag{2.1}$$

where $T_{\text{exp},n}$ is the n-th node's expected time interval between two consecutive successful medium accesses when it is active in contention as illustrated in Fig. 2.1. When the n-th node is active in contention, some other nodes could be also active. The medium access opportunities are then shared among all those active nodes, where back-off processes of several nodes could expire at the same time slot, and transmissions of active nodes may collide at times. Therefore, $T_{\text{exp},n}$ is a function of how frequently the other $N - 1$ nodes contend for medium access opportunities.

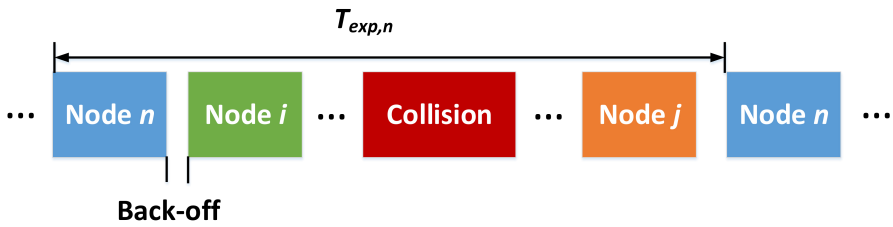


Figure 2.1: Expected access interval.

Given that the n-th node is active in contention, there are 2^{N-1} possible cases depending on whether each of the other $N-1$ nodes is also active or not. For example, the n-th node solely accesses the medium if none of the other nodes is active. On the other hand, if some other nodes are also active, the active nodes contend for medium access opportunities. Therefore, $T_{\text{exp},n}$ is presented as the expectation over all the 2^{N-1} cases.

$$T_{\text{exp},n} = \sum_{A \in \Lambda_n} T_{\text{avg},n}(A) P_n(A), \tag{2.2}$$

where set Λ_n includes 2^{N-1} elements, representing all possible combinations of active nodes, $T_{\text{avg},n}(A)$ is the average time interval between two consecutive successful medium accesses of the n-th node when nodes in set A are also active, and $P_n(A)$ is the probability that nodes in set A are also active. As shown in the following equation, $P_n(A)$ is a function of the activity probabilities of the other $N-1$ nodes.

$$P_n(A) = \prod_{j \in A} p_j \prod_{j \in (A^C - \{n\})} (1 - p_j). \tag{2.3}$$

Deriving $T_{\text{avg},n}(A)$ requires estimation on both the average number of collisions and average back-off time when nodes in set A as well as the n-th node are active. When a node has a packet in its transmission queue, it first sets its back-off counter with a value randomly chosen between $[0, W(m)]$, where $W(m)$ is the CW given by

$$W(m) = \min\left(2^m (W_{\text{min}} + 1) - 1, \quad W_{\text{max}}\right), \tag{2.4}$$

where m is the number of consecutive collisions and M is the maximum back-off stage, which satisfies $2^M (W_{\text{min}} + 1) - 1 = W_{\text{max}}$. The derivation of the average number of collisions and average back-off time is then preceded by deriving the transmission and collision probabilities considering the exponential increase of $W(m)$.

When k nodes have the saturated traffic load, i.e., they are always active in contention, it is shown that the transmission probability $p_t(k)$ of a node is given by [1]

$$p_t(k) = \frac{2}{1 + (W_{\text{min}} + 1) + p_c(k)(W_{\text{min}} + 1) \sum_{m=0}^{M-1} (2p_c(k))^m}, \tag{2.5}$$

where the collision probability $p_c(k)$ is given by

$$p_c(k) = 1 - (1 - p_t(k))^{k-1}. \tag{2.6}$$

The two equations can be solved by numerical methods, and always have a unique joint solution [1].

As shown in (2.2), the proposed performance analysis model considers node heterogeneity by taking the expectation over all possible combinations of active nodes. For a given active node combination, $A \in \Lambda_n$, only the n-th node and nodes in set A are active in contention, i.e., the number of active nodes becomes $|A|+1$, where $|A|$ is the cardinality of set A. Therefore, the transmission and collision probabilities for that active node combination are respectively presented by $p_t(|A|+1)$ and $p_c(|A|+1)$.

Once the transmission and collision probabilities are calculated, the number of slots during back-off could be approximately modeled as a geometric random variable which describes the number of trials before obtaining one success. The average back-off time then becomes the slot time T_{slot} multiplied by the mean of the geometric random variable:

$$T_{\text{BO}}(|A|+1) = \frac{T_{\text{slot}}(1 - p_t(|A|+1))}{p_t(|A|+1)}. \tag{2.7}$$

Given that at least one of the active nodes transmits a packet, the probability that the transmission would be successful is given by

$$p_s(|A|+1) = \frac{(|A|+1)\cdot p_t(|A|+1)(1-p_t(|A|+1))^{(|A|+1)-1}}{1-(1-p_t(|A|+1))^{|A|+1}}, \tag{2.8}$$

where the denominator is the probability that at least one of the $|A|+1$ nodes transmits a packet and the numerator is the probability that only one of the $|A|+1$ nodes transmits a packet. Then, the average number of consecutive collisions per successful transmission also becomes the mean of a geometric random variable:

$$N_c(|A|+1) = \frac{1 - p_s(|A|+1)}{p_s(|A|+1)}. \tag{2.9}$$

Let us denote the time duration of the n-th node's data packet transmission session as

$$T_{d,n} = T_{\text{DIFS}} + T_{\text{PHY}} + \frac{L_{\text{MAC}}+L_n+L_{\text{CRC}}}{r_n} + T_{\text{SIFS}} + T_{\text{ACK},n}, \quad n = 1, 2, \ldots, N. \tag{2.10}$$

When transmissions collide, they are not acknowledged by ACKs. Therefore, we also define

$$T_{d|c,n} = T_{d,n} - T_{\text{SIFS}} - T_{\text{ACK},n}, \qquad n = 1, 2, \ldots, N, \tag{2.11}$$

to be the time duration of the n-th node's data packet transmission session when the transmission is involved with a collision.

From all the above analysis, $T_{\text{avg},n}(A)$ is presented by

$$T_{\text{avg},n}(A) = T_{\text{BO}}(|A|+1) + (|A|+1) N_c (|A|+1) T_c(A \cup \{n\}) + T_{d,n} + \textstyle\sum_{j \in A} T_{d,j}, \tag{2.12}$$

where $T_c(A \cup \{n\})$ is the worst case collision time when the n-th node and nodes in set A are active. When transmissions collide, the collision time is determined by the node which has the longest transmission time among all the nodes involved in the collision. Therefore, $T_c(A \cup \{n\})$ is given by

$$T_c(A \cup \{n\}) = \max_{j \in A \cup \{n\}} T_{d|c,j}. \tag{2.13}$$

Now, we calculate $T_{\text{exp},n}$ by substituting (2.3) and (2.12) for $P_n(A)$ and $T_{\text{avg},n(A)}$ in (2.2). Therefore, in order to eventually achieve the n-th node's expected throughput defined in (2.1), we need to obtain nodes' activity probabilities, p_n $(n = 1, 2, \ldots, N)$.

2.2.3 Determination of Activity Probabilities

In Section 2.2, we defined the activity probability as the probability that a node contends for medium access opportunities at a given time. A node contends to win the medium access opportunity only when it has packets to transmit in its transmission queue. Therefore, the activity probability of a node should be the ratio of its traffic

load to the throughput expected to achieve when it is assumed to always take part in contention. According to (2.1), the n-th node is expected to achieve the throughput $\frac{L_n}{T_{exp,n}}$ when it always participates in contention, i.e., $p_n = 1$. Therefore, its activity probability should satisfy the condition $\min\left(\frac{\lambda_n}{L_n}T_{\text{exp},n}\left(=\frac{\lambda_n}{\frac{L_n}{T_{\text{exp},n}}}\right),\quad 1\right)$.

Note that $T_{\text{exp},n}$ is a function of p_j's $(j \neq n)$. Therefore, the n-th node's activity probability p_n is not solely determined by its own traffic load λ_n, but also by all the other nodes' activity probabilities. In other words, all the nodes' activity probabilities, i.e., p_n's $(n = 1, 2, \ldots, N)$ need to be jointly determined as the solution which jointly satisfies the following N conditions:

$$p_n = \min\left(\frac{\lambda_n}{L_n}T_{\text{exp},n},\quad 1\right), \qquad n = 1, 2, \ldots, N. \tag{2.14}$$

Here, the set of equations given by (2.14) for $n = 1, 2, \ldots, N$ are the general form of the simultaneous equations that all the activity probabilities should jointly satisfy, and we name the set of those equations as *activity probability equations*.

Taking a closer look at (2.2) and (2.14), $T_{\text{exp},n}$ is a non-decreasing function of p_j's $(j \neq n)$, and p_n becomes larger with increase of p_j's $(j \neq n)$. In other words, increasing the traffic load of a node ends up increasing its activity probability as well as the other nodes' activity probabilities, which means that activity probability equations (2.14) reflect the fundamental nature of DCF. As a node has more traffic load, thus taking more access opportunities, the other nodes win fewer access opportunities. Consequently, it increases the time duration that packets stay in the other nodes' transmission queues, meaning that the other nodes' activity probabilities increase.

Moreover, utilizing the fact that $T_{\text{exp},n}$ is a non-decreasing function of p_j's $(j \neq n)$, we develop an iterative update algorithm to find the joint solution of activity probability equations as described in Algorithm 1. At the beginning of the iteration process, p_n is initialized to be zero for $n = 1, 2, \ldots, N$, and U is initialized as the set including all nodes. In the first iteration, each p_n $(n \in U)$ is independently calculated by (2.14), assuming that $p_j = 0$, $\forall j \neq n$. After the first iteration, U becomes the set of nodes whose activity probability is smaller than 1. In the second iteration, each p_n $(n \in U)$ is

again independently calculated by (2.14), using the values of p_j's $(j \neq n)$ calculated in the first iteration.

As p_n's $(n = 1, 2, \ldots, N)$ are iteratively updated in this manner, they are non-decreasing. Moreover, they are bounded by 1 as shown in (2.14), and hence, they should become stable within a finite number of iterations. Once the difference of their updated values and current ones is smaller than a threshold δ, the iterative update algorithm stops.[1] The above reasoning guarantees that the algorithm always stops within a finite number of iterations, and, we can always find the joint solution of (2.14).

2.2.4 Considering Aggregate MAC Protocol Data Unit (A-MPDU)

Under the legacy DCF access mechanism described in Section 2.1, a packet contains a single MAC protocol data unit (MPDU). To reduce the protocol overhead needed per packet, e.g., back-off and ACK transmission overheads, IEEE 802.11n [36] allows a node to aggregate multiple MPDUs in a single packet, thus achieving MAC efficiency improvement. We here extend our performance analysis model to incorporate A-MPDU. When the n-th node uses MPDU aggregation, $T_{d,n}$ in (2.10) is modified to

$$\begin{aligned} &T_{d,n} = \\ &T_{\text{DIFS}} + T_{\text{PHY}} + \frac{a_n(L_{\text{MAC}}+L_n+L_{\text{CRC}})}{r_n} + T_{\text{SIFS}} + T_{\text{BACK},n}, \end{aligned} \tag{2.15}$$

where a_n is the average number of MPDUs aggregated in a packet, and $T_{\text{BACK},n}$ is the averge block ACK (BACK) transmission time. Generally, the number of aggregated MPDUs is bounded by a time bound T_{limit} [24, 36].

Then, a_n is given by

$$a_n = \left\lfloor \frac{T_{\text{limit}} - T_{\text{PHY}}}{\frac{(L_{\text{MAC}}+L_n+L_{\text{CRC}})}{r_n}} \right\rfloor, \tag{2.16}$$

[1] Too small values of δ would result in inaccurate activity probability estimation. Extensive simulations reveal that δ has negligible impact on activity probability estimation once the error bound per node becomes smaller than 0.01, i.e., $\delta/N \leq 0.01$.

Algorithm 1: Iterative algorithm for calculating activity probabilities.

Initialization:

$t = 0$

for $n = 1, 2, \ldots, N$ **do**
  $p_n[t] = 0$

end

$U = \{n | n = 1, 2, \ldots, N\}$

Activity probability calculation:

while $U \neq \emptyset$ **and** $\epsilon > \delta$ **do**
  $\epsilon = 0$

  for $n \in U$ **do**
    $T_{exp,n} = 0$

    for $A \in \Lambda_n$ **do**
      $P_n(A) = \prod_{j \in A} p_j[t] \prod_{j \in (A^C - \{n\})} (1 - p_j[t])$

      $T_{\text{exp},n} = T_{\text{exp},n} + T_{\text{avg},n}(A) P_n(A)$

    end

    $p_n[t+1] = \min\left(\frac{\lambda_n}{L_n} T_{\text{exp},n}, \quad 1\right)$

    if $p_n[t+1] = 1$ **then**
      $U = U - \{n\}$

    end

    $\epsilon = \epsilon + |p_n[t+1] - p_n[t]|$

  end

  for $n \notin U$ **do**
    $p_n[t+1] = 1$

  end

  $t = t + 1$

end

Return:

for $n = 1, 2, \ldots, N$ **do**
  $p_n = p_n[t]$

end

where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. Moreover, the n-th node's expected throughput in (2.1) is also modified to

$$\psi_n = \frac{p_n a_n L_n}{T_{\text{exp},n}}. \tag{2.17}$$

Finally, the n-th node's equation in the activity probability equations (2.14) is modified to

$$p_n = \min\left(\frac{\lambda_n}{a_n L_n} T_{\text{exp},n}, 1\right). \tag{2.18}$$

2.3 Contention Window Control

We first introduce Genie-aided theoretically ideal but practically unimplementable CW control. We apply the performance analysis developed in Section 2.2, and theoretically estimate the contention status under the ideal CW control. Finally, we consider a practically implementable CW control scheme which achieves the theoretically estimated contention status in an average sense.

2.3.1 Genie-Aided Ideal Contention Window Control

Let us assume that some of N nodes are active in contention at a given time. If W_{min} is set for those active nodes to achieve the maximum throughput at the moment, it could be considered optimal at least during a very short time period. However, since the active node set varies over time, the optimal W_{min} is also time varying. If nodes adopt the optimal W_{min} at every time instant, it is obviously the theoretically ideal CW control. However, this is apparently impossible to implement in reality because it is not possible to know which set of nodes are active at every time instant, and set all the active nodes' W_{min} in real-time.

Even though the ideal CW control is unrealizable, we apply the performance analysis model developed in Section 2.2 from a theoretical point of view to estimate the activity probabilities of nodes under the Genie-aided ideal CW control. While W_{min} is a constant in the analysis in Section 2.2, the ideal CW control continuously tunes

$W_{\min}$ for the active node set at every time instant. Accordingly, equations derived in Section 2.2 should be modified to be applicable to the ideal CW control.

When k nodes are always active in contention, it is shown that the optimal transmission probability of a node, p_t^*, should satisfy [1]

$$(1-p_t^*)^k = \frac{T_c}{T_{\text{slot}}}\left\{kp_t^* - \left(1-(1-p_t^*)^k\right)\right\}, \tag{2.19}$$

where T_c is the collision time under the assumption that nodes transmit the same sized packets using the same transmission rate.

Activity probability estimation under the ideal CW control requires determination of the $W_{\min}$ value tuned for an active node set at a given time instant, where only nodes in the active node set are active in contention. Therefore, for a given active node set, the optimal transmission probability should satisfy (2.19) while the number of active nodes in the active node set is substituted for k, and T_c is determined by the node whose transmission time is longest among nodes in the active node set.

From the above reasoning, when the n-th node as well as nodes in set A is active, the optimal transmission probability $p_t^*\left(A\cup\{n\}\right)$ should satisfy (2.19) with $p_t^*\left(A\cup\{n\}\right)$, $|A|+1$, and $T_c\left(A\cup\{n\}\right)$ substituted for p_t^*, k, and T_c respectively. Under the condition that $p_t^*\left(A\cup\{n\}\right) \ll 1$, the 2-nd order Taylor expansion provides an approximation for $\left(1-p_t^*\left(A\cup\{n\}\right)\right)^{|A|+1}$ as shown below.

$$\left(1-p_t^*\left(A\cup\{n\}\right)\right)^{|A|+1} \approx 1-\left(|A|+1\right)p_t^*\left(A\cup\{n\}\right)+\frac{|A|(|A|+1)}{2}\left(p_t^*\left(A\cup\{n\}\right)\right)^2. \tag{2.20}$$

The optimal transmission probability $p_t^*\left(A\cup\{n\}\right)$ then is approximately given by

$$p_t^*\left(A\cup\{n\}\right) \approx \frac{1}{|A\cup\{n\}|\sqrt{\frac{T_c(A\cup\{n\})}{2T_{\text{slot}}}}}. \tag{2.21}$$

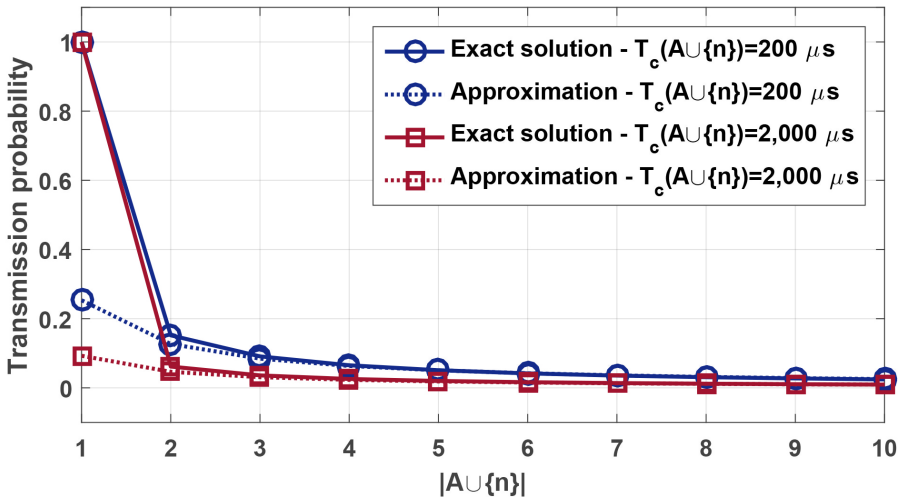


Figure 2.2: Comparison of the exact and approximate transmission probabilities.

Fig. 2.2 shows the exact solution and the approximate solution versus $|A \cup \{n\}|$ in two different $T_c(A \cup \{n\})$ cases, i.e., 200 and 2,000 μs.[2] The approximate solution deviates from the exact solution when $|A \cup \{n\}|$ is small. However, for $|A \cup \{n\}| \geq 2$, the approximate solution is very close to the exact solution. In addition, when $|A \cup \{n\}| = 1$, i.e., $A = \emptyset$, the exact solution obviously becomes $p_t^*(A \cup \{n\}) = 1$. Therefore, we simply present the optimal transmission probability as

$$p_t^*(A \cup \{n\}) = \begin{cases} 1, & |A \cup \{n\}| = 1, \\ \frac{1}{|A \cup \{n\}| \sqrt{\frac{T_c(A \cup \{n\})}{2T_{\text{slot}}}}}, & |A \cup \{n\}| = 2, 3, \dots . \end{cases} \tag{2.22}$$

Once the transmission probability is determined by (2.22), the collision probability $p_c^*(A \cup \{n\})$ is given by $p_c^*(A \cup \{n\}) = 1 - (1 - p_t^*(A \cup \{n\}))^{(|A|+1)-1}$. As we discussed in Section 2.2, the transmission probability, collision probability, and W_{min} satisfy (2.5). Therefore, if we solve (2.5) for W_{min} with $p_t^*(A \cup \{n\})$ and $p_c^*(A \cup \{n\})$ substituted for $p_t(k)$ and $p_c(k)$ respectively, it becomes the value of W_{min} tuned for the active node set at the time instant, which is denoted as $W_{\text{min}}^*(A \cup \{n\})$.

Once W_{min} is set to $W_{\text{min}}^*(A \cup \{n\})$, the average back-off time $T_{\text{BO}}^*(A \cup \{n\})$ at the time instant is given by (2.7) with $p_t^*(A \cup \{n\})$ substituted for $p_t(k)$. The probability $p_s^*(A \cup \{n\})$ that a transmission would be successful at the time instant is given by (2.8) with $p_t^*(A \cup \{n\})$ and $|A| + 1$ substituted for $p_t(k)$ and k respectively. Lastly, the average number of consecutive collisions per successful transmission $N_c^*(A \cup \{n\})$ is given by (2.9) with $p_s^*(A \cup \{n\})$ substituted for $p_s(k)$.

Accordingly, in the Genie-aided ideal CW control, the n-th node's average time interval between two consecutive successful medium accesses, $T_{\text{avg},n}^*(A)$, is given by (2.12) with $T_{\text{BO}}^*(A \cup \{n\})$ and $N_c^*(A \cup \{n\})$ substituted for $T_{\text{BO}}(|A| + 1)$ and $N_c(|A| + 1)$. Then, the n-th node's expected time interval between two consecutive successful medium accesses, $T_{\text{exp},n}^*$, is given by (2.2) with $T_{\text{avg},n}^*(A)$ substituted

[2] Assuming a packet size of 1,500 bytes, IEEE 802.11n MCS 0 and 7 with 20 MHz bandwidth and a single stream result in transmission time of about 200 and 2,000 μs respectively.

for $T_{\text{avg},n}(A)$. Lastly, the activity probability equations (2.14) are also modified with $T^*_{\text{exp},n}$ substituted for $T_{\text{exp},n}$. Hence, if we find the joint solution, i.e., p_n's $(n = 1, 2, \ldots, N)$ that simultaneously satisfy the activity probability equations (2.14) via Algorithm 1, we obtain the activity probabilities of nodes under the Genie-aided ideal CW control, and denote them as p^*_n's $(n = 1, 2, \ldots, N)$.

2.3.2 Proposed Contention Window Control

Now, we develop a realistically implementable CW control scheme which achieves the theoretically estimated ideal contention status in an average sense. Under the Genie-aided ideal CW control, the probability that nodes in set A are active at a given time is given by

$$P^*(A) = \prod_{j \in A} p^*_j \prod_{j \in (A^C)} \left(1 - p^*_j\right). \tag{2.23}$$

With the active node set A, the number of active nodes and the worst case collision time are respectively $|A|$ and $T_c(A)$.

Therefore, the average number of active nodes at a given time under the ideal CW control is given by

$$N_{\text{avg}} = \sum_{A \in \Lambda} |A| P^*(A) = \sum_{n=1}^{N} p^*_n, \tag{2.24}$$

where Λ is the set of all possible 2^N subsets of a set composed of N nodes. In addition, the average worst case collision time is given by

$$\begin{aligned} T_{\text{avg},c} &= \textstyle\sum_{A \in \Lambda} T_c(A) P^*(A) \\ &= \textstyle\sum_{i=1}^{N} \left[p^*_{\Phi_i} T_{d|c,\Phi_i} \left\{ \prod_{j=1}^{i-1} \left(1 - p^*_{\Phi_j}\right) \right\} \right], \end{aligned} \tag{2.25}$$

where the second equality comes from the fact that $T_c(A)$ is determined by the largest value among $T_{d|c,j}$'s $(j \in A)$, and Φ is the ordered set of N nodes, in which nodes are ordered in the decreasing order of $T_{d|c,n}$, and the i-th element is denoted as Φ_i.

N_{avg} and $T_{\text{avg},c}$ reflect the theoretically ideal contention status under the ideal CW control in an average sense, and hence, the value of W_{min} tuned for them could be considered to achieve the ideal contention status in an average sense. Tuning the value

of $W_{\min}$ for N_{avg} and $T_{\text{avg},c}$ means that the transmission probability $p^*_{\text{avg},t}$ should satisfy (2.19) with N_{avg} and $T_{\text{avg},c}$ substituted for k and T_c respectively. Once the transmission probability is determined, the collision probability is given by $p^*_{\text{avg},c} = 1 - \left(1 - p^*_{\text{avg},t}\right)^{N_{\text{avg}}-1}$. Lastly, if we solve (2.5) for $W_{\min}$ with $p^*_{\text{avg},t}$ and $p^*_{\text{avg},c}$ substituted for $p_t(k)$ and $p_c(k)$ respectively, the recommended value of $W_{\min}$ is given by

$$W^*_{\min,\text{p}} = \frac{2 - 2p^*_{\text{avg},t} - p^*_{\text{avg},t}p^*_{\text{avg},c}\sum_{m=0}^{M-1}\left(2p^*_{\text{avg},c}\right)^m}{p^*_{\text{avg},t}\left(1 + p^*_{\text{avg},c}\sum_{m=0}^{M-1}\left(2p^*_{\text{avg},c}\right)^m\right)}. \tag{2.26}$$

Equation 2.26 enables realistically implementable CW control, where all the required parameters $p^*_{\text{avg},t}$ and $p^*_{\text{avg},c}$ are theoretically estimated considering node heterogeneity as discussed above.

Computational Complexity Reduction

As discussed in Section 2.3.2, the proposed CW control determines the value of $W_{\min}$ based on parameters N_{avg}, $T_{\text{avg},c}$, $p^*_{\text{avg},t}$, and $p^*_{\text{avg},c}$. All these parameters are determined based on nodes' activity probabilities under the Genie-aided ideal CW control, i.e., $p^*_n\ (n = 1, 2, \ldots, N)$. Activity probability estimation requires finding the expectation, $T^*_{\text{exp},n}$ over all possible active node combinations as explained in Section 2.3.1, which incurs computational complexity in the order of 2^{N-1}. It may be a substantial burden for an AP or an AP controller when N is sufficiently large, and hence, we develop a method which estimates activity probabilities in a polynomial time.

Under the condition that $p^*_t(A \cup \{n\}) \ll 1$ and $|A \cup \{n\}| \gg 1$, the 2-nd order Taylor expansion provides an approximation $\hat{p}^*_s(A \cup \{n\})$ for $p^*_s(A \cup \{n\})$ with (2.22) substituted for $p^*_t(A \cup \{n\})$.

$$\hat{p}^*_s(A \cup \{n\}) = \frac{1 - \sqrt{\frac{2T_{\text{slot}}}{T_c(A \cup \{n\})}} + \frac{1}{2}\frac{2T_{\text{slot}}}{T_c(A \cup \{n\})}}{1 - \frac{1}{2}\sqrt{\frac{2T_{\text{slot}}}{T_c(A \cup \{n\})}}}, \tag{2.27}$$

which is solely determined by $T_c(A \cup \{n\})$. Since $N^*_c(A \cup \{n\})$ is a function of $p^*_s(A \cup \{n\})$ as shown in Section 2.3.1, the above approximation makes $N^*_c(A \cup \{n\})$

also solely determined by $T_c\left(A \cup \{n\}\right)$.

From the above reasoning, the sum of both the back-off and collision time overhead in $T^*_{\text{avg},n}\left(A\right)$ is approximately given by

$$\begin{aligned} &T^*_{\text{BO}}\left(A \cup \{n\}\right) + \left(|A|+1\right) N^*_c\left(A \cup \{n\}\right) T_c\left(A \cup \{n\}\right) \approx \\ &\left(|A|+1\right)\left[\frac{\sqrt{T_c(A \cup \{n\}) T_{\text{slot}}}}{2} + T_c\left(A \cup \{n\}\right) \frac{1-\hat{p}^*_s(A \cup \{n\})}{\hat{p}^*_s(A \cup \{n\})}\right]. \end{aligned} \tag{2.28}$$

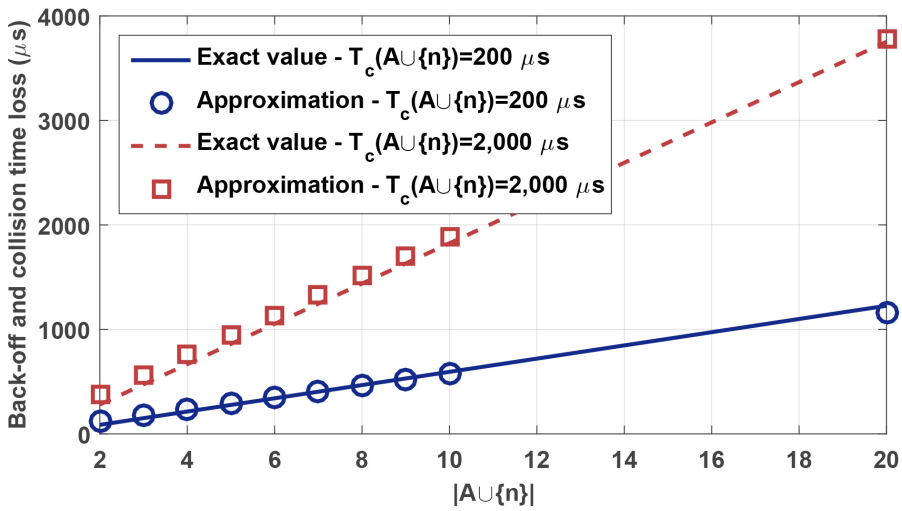


Figure 2.3: Comparison of the exact and approximate back-off and collision time loss.

Fig. 2.3 shows the exact back-off and collision time overhead, and its approximation given by (2.28) versus $|A \cup \{n\}|$ in two different $T_c(A \cup \{n\})$ cases. We observe that (2.28) provides a close approximation for the back-off and collision time overhead for $|A \cup \{n\}| \geq 2$. From (2.22), note that the back-off and collision time overhead becomes 0 when $|A \cup \{n\}| = 1$.

In (2.28), let us denote the sum of the two terms in the square brackets as $R(T_c(A \cup \{n\}))$. Then, in a similar manner to that shown in (2.25), $T^*_{\text{exp},n}$ could be efficiently calculated in a polynomial time as shown below.

$$\begin{aligned} T^*_{\text{exp},n} &= \textstyle\sum_{A \in \Lambda_n} T^*_{\text{avg},n}(A) P_n(A) \\ &= \textstyle\sum_{A \in \Lambda_n} \left[(|A|+1) R(T_c(A \cup \{n\})) + T_{d,n} + \sum_{j \in A} T_{d,j} \right] P_n(A) \\ &= \textstyle\sum_{A \in \Lambda_n} (|A|+1) R(T_c(A \cup \{n\})) P_n(A) + T_{d,n} + \sum_{j \neq n} p_j T_{d,j} \\ &= \textstyle\sum_{i=1,\Phi_i \neq n}^{N} \left[p_{\Phi_i} R\left(T_{d|c,\Phi_i}\right) \left(2 + \sum_{j=i+1,\Phi_j \neq n}^{N} p_{\Phi_j}\right) \left\{ \prod_{j=1,\Phi_j \neq n}^{i-1} \left(1 - p_{\Phi_j}\right) \right\} \right] \\ &\quad + T_{d,n} + \textstyle\sum_{j \neq n} p_j T_{d,j}. \end{aligned} \tag{2.29}$$

In (2.29), the forth equality holds from the fact that $R(T_c(A \cup \{n\}))$ becomes 0 with $A = \emptyset$, and $T_c(A \cup \{n\})$ is determined by the largest value of $T_{d|c,j}$'s ($j \in A \cup \{n\}$). Using (2.29), the computational complexity for calculating $T^*_{\text{exp},n}$ linearly increases with N, and hence, the computational complexity for calculating p^*_n's is given by $O(N)$.

Once p^*_n's are obtained, the proposed CW control calculates N_{avg}, $T_{\text{avg},c}$, $p^*_{\text{avg},t}$, and $p^*_{\text{avg},c}$ to determine $W^*_{\text{min},p}$ as explained in Section 2.3.2, which also incurs computational complexity of $O(N)$. Therefore, the total computational complexity for calculating $W^*_{\text{min},p}$ is also given by $O(N)$. Considering that an AP or AP controller calculates the CW size with a long interval, the computational complexity of $O(N)$ is assessed to be affordable for practical WLAN devices.

Contention Window Control in the Time-Varying Network

As explained in Section 2.3.2, the proposed CW control determines the CW size based on the knowledge of node properties including average transmission rates, traffic loads,

and packet sizes. In a practical WLAN, those node properties vary over time. The proposed CW control then adapts to the network time variation by updating the node property information, and recalculating the CW size with the updated information.

As an example, we obtain real channel data traces in an environment with high mobility, i.e., a conference poster presentation session. We deploy 9 mobile nodes across the space, and then, a transmitter is set to continuously broadcast short packets to those mobile nodes which accumulate the real time link conditions using the CSI measurement tool [37]. Then, with a specific realization of random traffic loads and packet sizes, we calculate the ideal $W^*_{\text{min},p}$ values at every time instant, where the transmission rates are assumed to be ideally selected as the largest rate among all possible rates satisfying the frame error rate smaller than 1% under the real time link condition. Then, the ideal $W^*_{\text{min},p}$ values are compared with those calculated with periodic node property information update.

In a WLAN, a beacon frame includes the CW information, which means that the minimum possible update interval is the beacon transmission interval that is typically configured to be 100 ms [35]. The 100 ms update interval might be a non-negligible burden for an AP equipped with a very poor processor, and hence, we also consider larger update intervals, i.e., 500 and 1,000 ms. Fig. 2.4 shows the CW size over time. Despite the high mobility of nodes, the proposed CW control keeps up with the ideal $W^*_{\text{min},p}$ by periodically updating node property information, and recalculating the CW size. When the error is normalized to the ideal $W^*_{\text{min},p}$ value, and averaged across the time, 100, 500 and 1,000 ms update intervals results in only about 3.29%, 7.35%, and 7.91% deviation from the ideal $W^*_{\text{min},p}$ value.

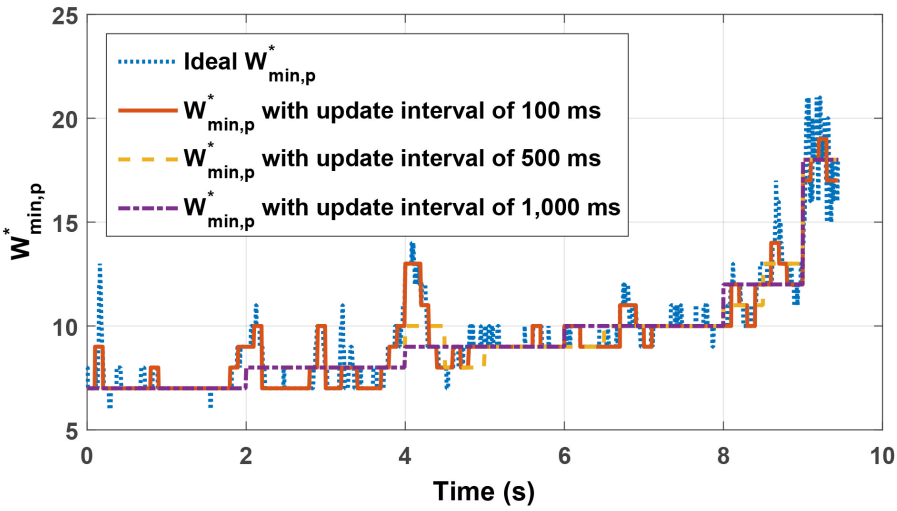


Figure 2.4: $W^*_{\text{min},p}$ variation with time in a real WLAN.

Table 2.1: Parameters used in simulations.

T_{SIFS}	T_{DIFS}	T_{slot}	T_{PHY}	L_{MAC}	L_{CRC}	ACK (BACK) frame size
16 µs	34 µs	9 µs	28 µs	26×8 bits	4×8 bits	$14\,(32) \times 8$ bits

Table 2.2: Node properties.

r_n	randomly selected among all possible IEEE 802.11n rates with a single stream and 20 MHz bandwidth, i.e., 6.5, 13, 19.5, 26, 39, 52, 58.5, and 65 Mb/s
L_n	randomly selected between $1,000$ and $1,500$ bytes.
λ_n	randomly selected between 0 and λ_{max}

2.4 Performance Evaluation

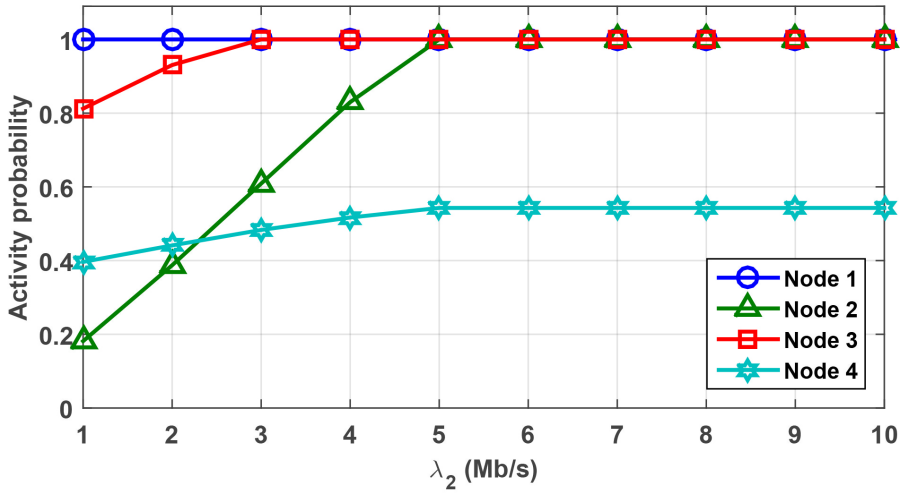
In Sections 2.4.1 and 2.4.2, we present NS-3 simulation results to validate both the proposed performance analysis model and CW control. All the simulation results are obtained in the 802.11n WLAN, and Table 2.1 shows the parameters used in the simulations. Unless specified otherwise, node properties such as transmission rates, packet sizes, and traffic loads are heterogeneously selected as shown in Table 2.2. In Section 2.4.3, we also present real testbed experiment results to validate the feasibility of the proposed CW control.

2.4.1 Evaluation of Proposed Performance Analysis Model

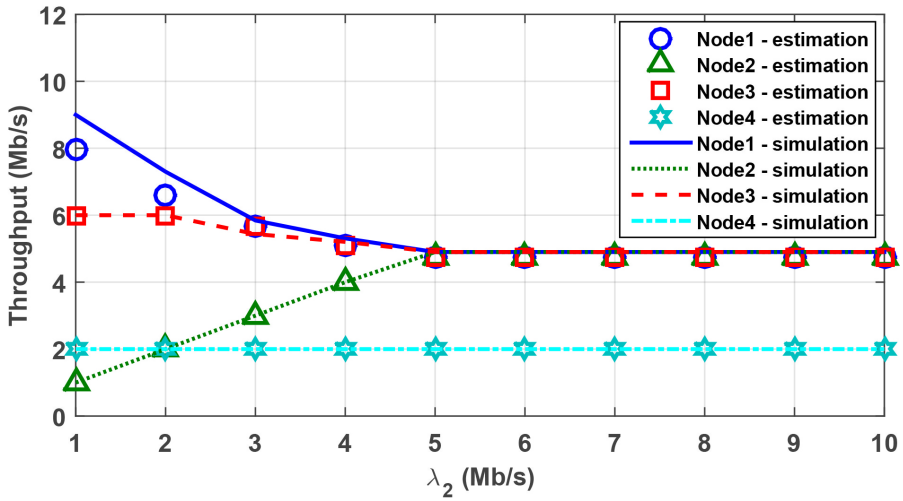
We firstly validate the throughput prediction accuracy of the proposed performance analysis model in a specific topology composed of four nodes $(N = 4)$. Nodes have heterogeneous transmission rates, $r_1 = 52$ Mb/s, $r_2 = 26$ Mb/s, $r_3 = 39$ Mb/s, and $r_4 = 13$ Mb/s, respectively. Nodes 1, 3, and 4 have heterogeneous traffic loads, $\lambda_1 = 9$ Mb/s, $\lambda_3 = 6$ Mb/s, and $\lambda_4 = 2$ Mb/s, respectively, and we vary the traffic load of node 2, i.e., λ_2 from 1 Mb/s to 10 Mb/s. The MAC payload size is set to 1,500 bytes for all nodes $(L_n = 1,500 \times 8\,\text{bits}\,, \forall n = 1, 2, 3, 4)$, which leads to throughput fairness

among nodes which are always active in contention.[3]

[3]The 802.11 DCF provides access fairness characteristic, i.e., when nodes are always active in contention, they achieve the same number of access opportunities in the long term [15]. In this simulation topology, the MAC payload size is equal for all nodes, and hence, access fairness means throughput fairness.



(a) Activity probability estimation.



(b) Per-node throughput estimation.

Figure 2.5: Throughput estimation in a heterogeneous transmission rate and traffic load topology.

Fig. 2.5(a) shows the activity probabilities, i.e., p_1, p_2, p_3, and p_4 calculated as the joint solution of (2.14) for given λ_2. As λ_2 increases from 1 Mb/s to 10 Mb/s, all nodes' activity probabilities jointly increase because they have relations with one another through (2.14). As λ_2 increases, node 2 obtains more access opportunities, which ends up with fewer access opportunities for the other nodes, i.e., they become more active in contention. p_1 and p_3 become 1 once λ_2 reaches 3 Mb/s, and p_2 also becomes 1 once λ_2 reaches 5 Mb/s. p_4 remains less than 1 regardless of λ_2 due to its very light traffic load.

Fig 2.5(b) shows the per-node throughput estimation results, where each node's throughput is estimated by (2.1) using the activity probabilities. When $\lambda_2 \geq 3$ Mb/s, the proposed performance analysis model predicts that the two nodes achieve the same throughput due to the access fairness characteristic of DCF. Moreover, when $\lambda_2 \geq 5$ Mb/s, nodes 1, 2, and 3 are predicted to achieve the same throughput. Lastly, node 4 is predicted to achieve the same throughput as its traffic load because its activity probability is less than 1 regardless of λ_2. The simulation result shows that the proposed performance analysis model reflects the access fairness characteristic of DCF well, and accurately estimates each node's throughput.

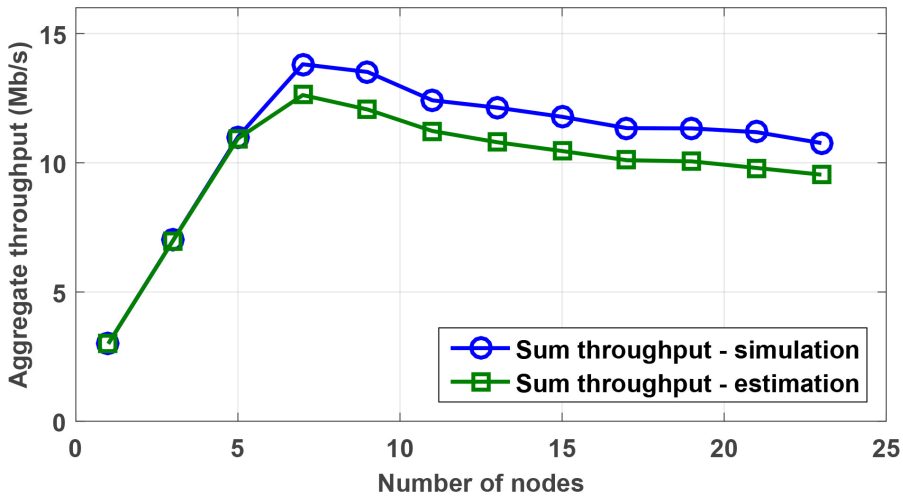


Figure 2.6: Aggregate throughput estimation.

Next, we validate the back-off and collision time overhead analysis incorporated in the proposed model. Node properties are heterogeneously selected according to Table 2.2 with $\lambda_{\max} = 4$ Mb/s. Fig. 2.6 shows both the aggregate throughput estimated by our proposed model and that obtained through NS-3 simulations. When N is smaller than 7, the medium is underutilized, and hence, the aggregate throughput increases as N increases. However, as we further increase N, a larger amount of time resource is wasted by collisions, which leads to aggregate throughput decrease. The simulation result shows that the incorporated collision probability and collision time analysis reasonably reflect the back-off and collision time overhead of DCF in heterogeneous node environment.

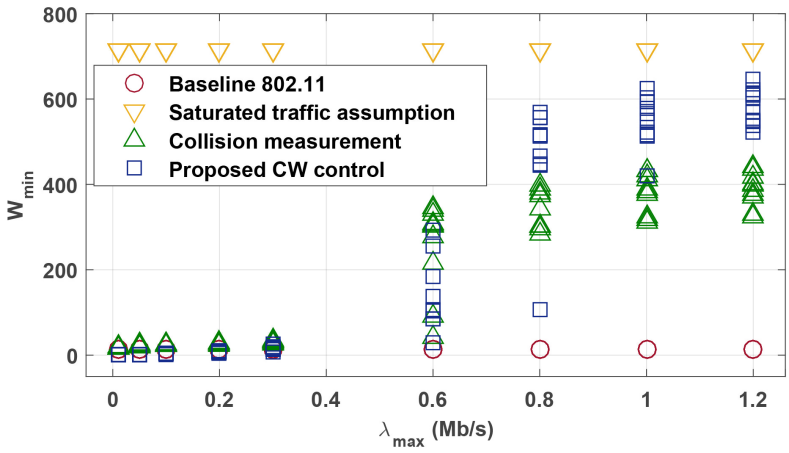
2.4.2 Evaluation of Proposed Contention Window Control

In this subsection, we present NS-3 simulation results for evaluation of the proposed CW control. The following CW control approaches are compared: 1) the baseline 802.11 CW control defined in IEEE 802.11 [35], 2) the saturated traffic assumption-based CW control derived in [1],[4] and 3) the collision measurement-based CW control proposed in [18] which adjusts the CW size according to the run-time collision probability measurement.

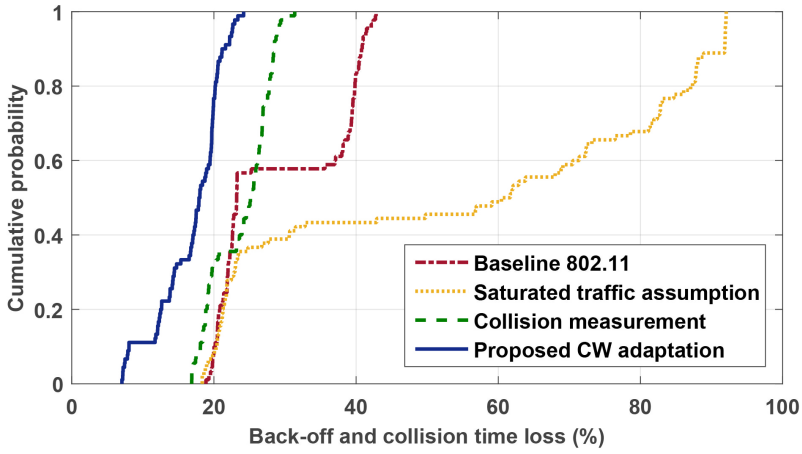
We consider a topology of two APs each of which has 20 associated nodes, i.e., $N = 40$, where all the APs and nodes can carrier-sense each other, and one of those 40 nodes is set to always have saturated uplink traffic with 1,500 bytes MAC payload size and 65 Mb/s transmission rate for observation of maximum achievable throughput. Each AP determines the CW size considering node properties of its associated nodes and the other AP's ones using (2.26), which leads to the same CW size calculated at the two APs. We compare the network performance under different CW control

[4]Because the analysis derived in [1] assumes homogeneous node properties, it is not directly applicable to the following simulation topologies with node heterogeneity. Therefore, we apply the analysis using the lowest transmission rate and largest MAC payload size in each simulation topology.

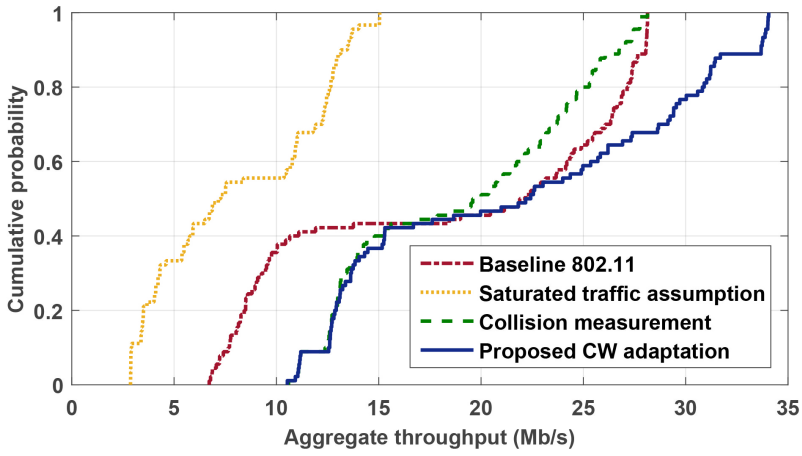
approaches with λ_{max} varying from 0.01 Mb/s to 1.2 Mb/s. For given λ_{max}, we execute 10 different runs with different realizations of node properties, where each run lasts for 60 seconds.



(a) $W_{\min}$.



(b) Back-off and collision time loss.



(c) Aggregate throughput.

Figure 2.7: CW control with varying traffic load.

Fig. 2.7(a) shows selected W_{min} according to given λ_{max}. Both the baseline 802.11 CW control and saturated traffic assumption-based CW control use fixed W_{min} without considering node heterogeneity. The collision measurement-based CW control adjusts the CW size according to the measured collision probability only, and its contention level estimation analysis does not consider node heterogeneity as mentioned in Section 1.1. Accordingly, the selected W_{min} values are noticeably different from those selected under the proposed CW control which determines W_{min} with comprehensive consideration of node heterogeneity.

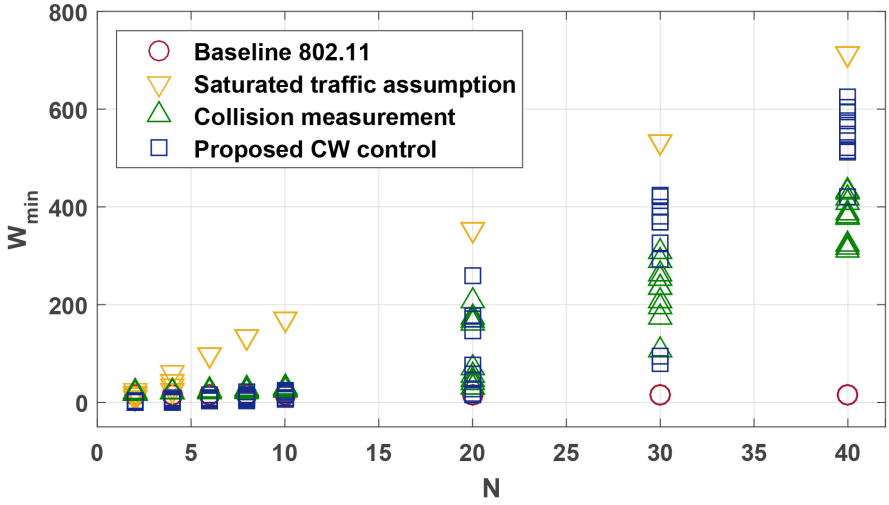
Fig. 2.7(b) shows the CDF of back-off and collision time loss, i.e., the total time portion that is not utilized by successful transmissions. Note that the CDF is obtained by collecting all the results obtained in all runs with different realizations of node properties, and the leftmost line means the largest time portion utilized for successful transmissions. The baseline CW control incurs large back-off time loss when λ_{max} is small, e.g., about 22% of back-off time loss when $\lambda_{\text{max}} \leq 0.2$ Mb/s. On the other hand, it incurs severe collision time loss when λ_{max} is large, e.g., about 40% of collision time loss when $\lambda_{\text{max}} \geq 0.8$ Mb/s. The saturated traffic assumption-based CW control incurs severe back-off time loss across the entire range of λ_{max}. When $\lambda_{\text{max}} \leq 0.2$ Mb/s, the back-off time loss is about 83% while it reduces to about 21% when $\lambda_{\text{max}} \geq 0.8$ Mb/s.

The collision measurement-based CW control reduces the collision time loss compared with the baseline CW control, e.g., about 20% of collision time loss when $\lambda_{\text{max}} \geq 0.8$ Mb/s. However, its contention level estimation is not valid with node heterogeneity, thus resulting in particularly large performance degradation compared with the proposed CW control when λ_{max} is small, i.e., node heterogeneity in the traffic load is significant with a single heavily loaded node and 39 lightly loaded nodes. The proposed CW control leads to considerably less back-off and collision time loss than the compared approaches. The back-off time loss reduces to about 12% when $\lambda_{\text{max}} \leq 0.2$ Mb/s while the collision time loss reduces to about 19% when $\lambda_{\text{max}} \geq 0.8$ Mb/s. On average, 29.34%, 53.87%, and 23.83% of time resources

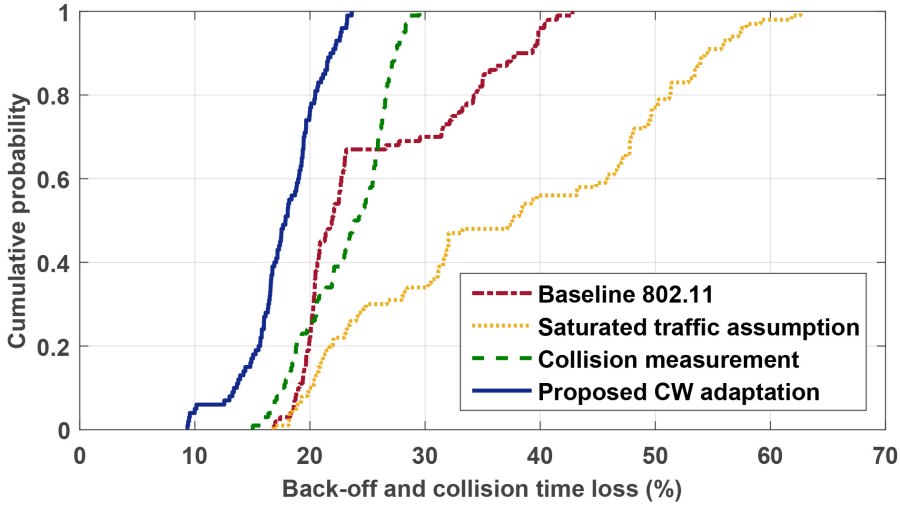
are wasted by back-off and collisions in the baseline CW control, saturated traffic assumption-based, and collision measurement CW control, respectively. On the other hand, under the proposed CW control, the back-off and collision time loss reduces to 16.82% on average.

Fig. 2.7(c) shows the CDF of the aggregate throughput, where the rightmost line means the largest aggregate throughput. As shown in Fig. 2.7(b), the proposed CW control considerably reduces the back-off and collision time loss, thus leading to larger aggregate throughput than the compared approaches. The maximum throughput gains over the baseline CW control, saturated traffic assumption-based, and collision measurement-based CW control are 66.59%, 1,084.80%, and 28.53%, respectively.

Here, we vary the number of nodes, N, from 2 to 40 for a fixed $\lambda_{\max} = 1$ Mb/s, where two APs have the same number of associated nodes, and again, one of those N nodes is set to always have saturated uplink traffic with 1,500-byte MAC payload and 65 Mb/s transmission rate for observation of maximum achievable throughput.



(a) $W_{\min}$



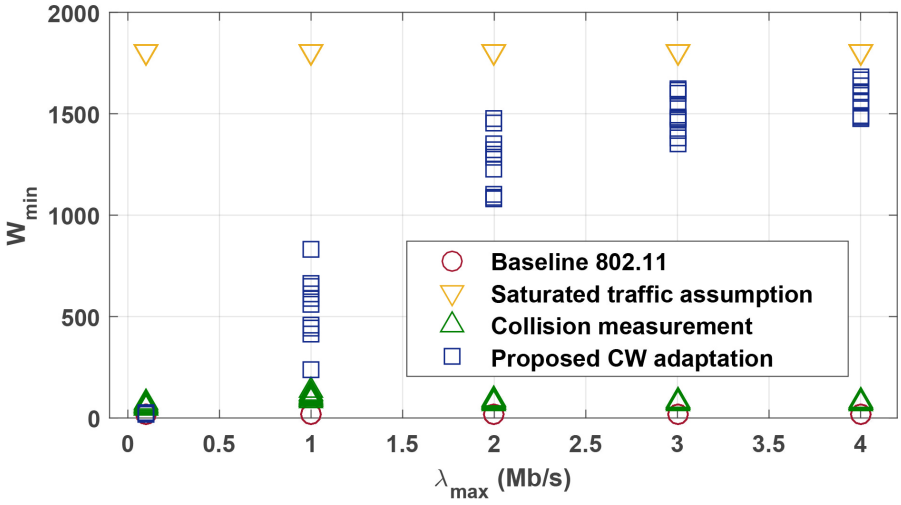
(b) Back-off and collision time loss.

Figure 2.8: CW control with varying number of nodes.

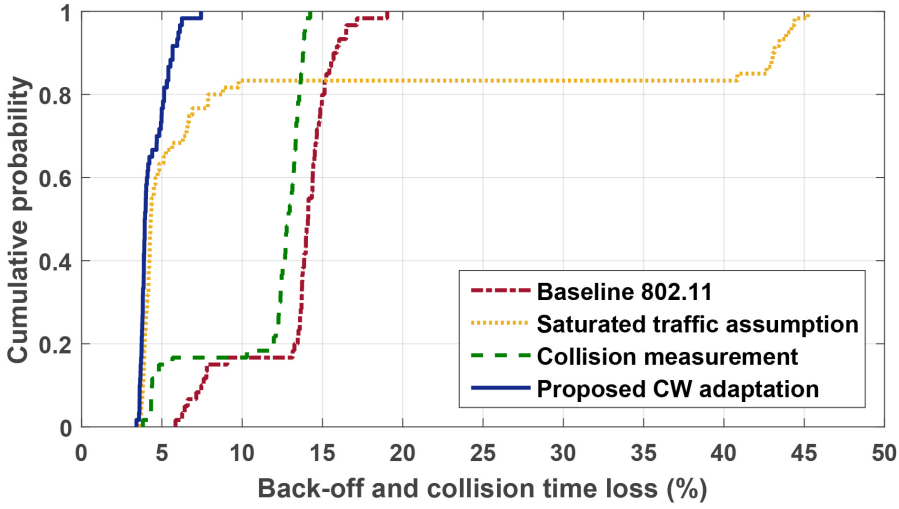
Fig. 2.8(a) shows the selected W_{min} according to given N. In contrast to Fig. 2.7(a), both the saturated traffic assumption-based and collision measurement-based CW control increase W_{min} with N. However, the W_{min} values selected under the compared approaches are still noticeably different from those selected under the proposed CW control.

Fig. 2.8(b) shows the CDF of back-off and collision time loss, which shows a trend similar to that in Fig. 2.7(b). The baseline CW control incurs about 21% of back-off time loss when $N \leq 8$ while it incurs about 33% of collision time loss when $N \geq 20$. The saturated traffic assumption-based CW control incurs about 45% of back-off time loss when $N \leq 8$. The collision measurement-based CW control reduces collision time loss to about 20% when $N \geq 20$. However, all those CW control approaches do not consider node heterogeneity, and still show performance degradation compared with the proposed CW control.

On average, 25.46%, 38.38%, and 23.58% of time resources are wasted by back-off and collisions under the baseline CW control, saturated traffic assumption-based, and collision measurement-based CW control, respectively. On the other hand, the proposed CW control maintains the back-off and collision time loss at 17.13% on average. The proposed CW control achieves considerable throughput improvement over the compared CW control approaches, e.g., maximum throughput gains of 65.20%, 157.78%, and 25.21% over the baseline CW control, saturated traffic assumption-based, and collision measurement-based CW control, respectively.



(a) $W_{\min}$.



(b) Back-off and collision time loss.

Figure 2.9: CW control with AMPDU and varying traffic load.

Next, we consider the topology with a fixed number of nodes, $N = 40$, again with MPDU aggregation enabled, where the maximum aggregate time bound is fixed to 10 ms [36]. Fig. 2.9(a) shows the values of W_{min} as λ_{max} increases. Given that the frame size increases by A-MPDU, a collision results in severe time loss due to the increased transmission time. Therefore, the proposed CW control rapidly increases W_{min} according to the increase of λ_{max}, and employs much larger W_{min} values compared with the result shown in Fig. 2.7(a). On the other hand, the collision measurement-based CW control does not reflect the increased time loss per collision. The collision probability rather decreases due to increased throughput with A-MPDU, and hence, the collision measurement-based CW control rather employs W_{min} values smaller than those in Fig. 2.7(a). The saturated traffic assumption-based CW control still employs too conservative values for W_{min}.

Fig. 2.9(b) shows the CDF of back-off and collision time loss. Compared with Fig. 2.7(b), the overall back-off and collision time loss becomes smaller due to reduced overhead per medium access with A-MPDU. The baseline CW control incurs about 14% of collision time loss with A-MPDU when $\lambda_{\text{max}} \geq 1$ Mb/s. On the other hand, the saturated traffic assumption-based CW control incurs about 43% of back-off time loss when $\lambda_{max} = 0.1$ Mb/s. The collision measurement-based CW control slightly reduces collision time to about 13% when $\lambda_{\text{max}} \geq 1$ Mb/s. The proposed CW control sill leads to considerably less back-off and collision time loss than the compared approaches.

On average, 13.35%, 11.23%, and 11.60% of time resources are wasted by back-off and collisions under the baseline CW control, saturated traffic assumption-based, and collision measurement-based CW control, respectively. On the other hand, the proposed CW control maintains the back-off and collision time loss at 4.38% on average. The reduced back-off and collision time loss leads to considerable throughput improvement, e.g., maximum throughput gains of 59.07%, 84.67%, and 26.56% over the baseline CW control, saturated traffic assumption-based, and collision measurement

based CW control, respectively.

All the above simulation results reveal that considering node heterogeneity is critical for CW control in practical IEEE 802.11 WLANs. The proposed CW control considers node heterogeneity, and determines the CW size achieving theoretically ideal contention status in an average sense. It leads to considerable performance improvement over the compared CW control approaches which do not consider node heterogeneity.

2.4.3 Testbed Experiments

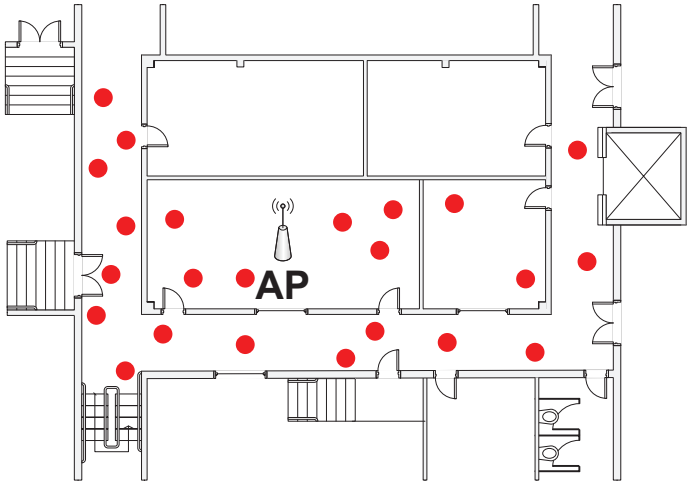


Figure 2.10: Testbed floor plan.

In this subsection, we verify the feasibility of the proposed scheme by presenting results of testbed experiments conducted in a controlled office environment. The operating channel number is 149, i.e., 5,745 MHz center frequency without any external interfere detected. We use a programmable 802.11n device, Qualcomm Atheros AR9380, along with *hostAP* to build an AP on linux environment [38]. All stations are equipped with Intel Ultimate-N 6300 Network Interface Card (NIC), and data traffic is generated using Iperf 2.0.5 [39]. Fig. 2.10 illustrates the floor plan, where red points represent different locations on which stations are randomly located. 12 stations have heterogeneous traffic loads, MAC payload sizes, and transmission rates, where the transmission rate at a given location is selected to achieve the maximum throughput based on measurement. One of 12 stations has saturated uplink traffic with 65 Mb/s transmission rate and 1,500 bytes MAC payload size. The results are averaged over 10 different runs with different node properties, where each run lasts for 30 seconds.

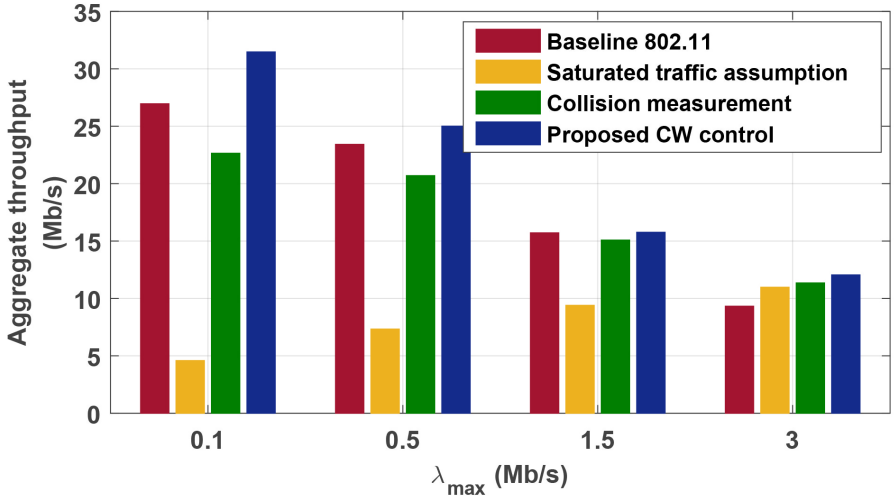


Figure 2.11: Aggregate throughput from testbed measurement.

Fig. 2.11 shows the average aggregate throughput results when $\lambda_{\max}$ is 0.1, 0.5, 1.5, and 3 Mb/s. We observe a similar trend as that in NS-3 simulations. When $\lambda_{\max}$ is small, e.g., $\lambda_{\max} \leq 1$ Mb/s, the baseline 802.11 CW control makes the medium underutilized while it incurs large collision time loss when we further increase $\lambda_{\max}$. The saturated traffic assumption-based CW control renders the medium underutilized at all values of $\lambda_{\max}$. The collision measurement-based CW control mitigates collisions compared to the baseline CW control when $\lambda_{\max}$ is sufficiently large. However, its contention level estimation is not valid with node heterogeneity, which ends up with performance degradation when $\lambda_{\max}$ is small, i.e., heterogeneity in the traffic load is significant with one heavily loaded station and 11 lightly loaded stations. The proposed CW control comprehensively considers node heterogeneity, thus achieving considerable performance improvement over the existing CW control approaches.

Chapter 3

Sounding Node Set and Sounding Interval Determination for IEEE 802.11ac MU-MIMO

3.1 MU-MIMO in IEEE 802.11ac

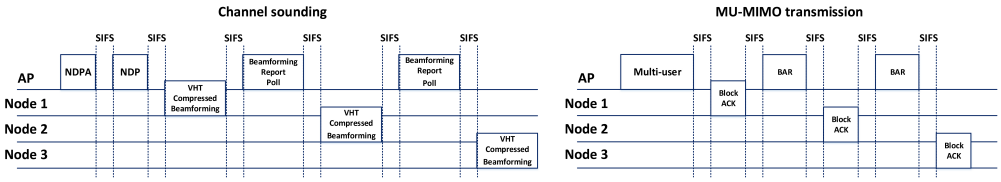


Figure 3.1: Channel sounding and MU-MIMO transmission protocol.

Fig. 3.1 shows the channel sounding and MU-MIMO transmission protocols defined in IEEE 802.11ac. When the AP performs channel sounding at a given time, it announces the beginning of a sounding process by transmitting a null-data packet announcement (NDPA). After a short inter-frame space (SIFS), the AP transmits a null-data packet (NDP) in which each AP antenna sequentially transmits a known signal for channel estimation. After a SIFS, a pre-designated node feeds back the CSI. After a SIFS again, the AP polls a next node and the polled node feeds back the CSI after a SIFS until there is no remaining node to be polled for CSI feedback.

When the AP serves multiple users through MU-MIMO transmission, it firstly contends for medium access with the packet at the head of the AP queue whose destination node becomes a primary user. Then, the AP selects some other packets in the AP queue whose destination node(s) become secondary user(s), and simultaneously transmits packets to the primary and secondary users. After the AP transmits packets to the primary and secondary users, the primary user transmits a block acknowledgement (BA) after a SIFS. Then, the AP transmits a block acknowledgement request (BAR) for one of the secondary users after a SIFS, and the polled node transmits a BA after a SIFS until there is no remaining node to be polled.

3.2 System Description

We consider a WLAN composed of an AP equipped with M antennas and N mobile nodes equipped with one antenna. The DL channel from the AP to the n-th node varies with time according to propagation environment variation. At time t, the channel is denoted as $\mathbf{h}_n(t) \in \mathbb{C}^{M\times 1}$ $(n = 1, 2, \ldots, N)$. The channel coherence time for $\mathbf{h}_n(t)$ is denoted as $T_{c,n}$ which means the expected time duration during which $\mathbf{h}_n(t)$ is expected to remain virtually constant. The long-term average DL SNR at the n-th node is denoted as γ_n.

All the N nodes have different DL traffic loads, and they are denoted as p_n $(0 \leq p_n \leq 1)$

which means the probability that the AP has packets destined for the n-th node at a given time. The DL time channel utilization for the n-th node is denoted as ρ_n $(0 \leq \rho_n \leq 1)$ which means the time portion that the AP utilizes for DL data transmission to the n-th node. As p_n increases, the wireless medium is utilized more for the DL transmission to the n-th node, thus resulting in increase of ρ_n.

The AP periodically collects CSI, i.e., performs channel sounding with a period T_s, and uses zero-forcing (ZF) MU-MIMO precoding based on the collected CSI. The time durations for the NDPA, NDP, SIFS, CSI report poll, and CSI report are denoted as T_{NDPA}, T_{NDP}, T_{SIFS}, T_{poll}, and T_{F}, respectively. The block acknowledgement and block acknowledgement request transmission times are denoted as T_{BA} and T_{BAR}, respectively.

3.3 MIMO Channel Characteristics in real IEEE 802.11ac WLANs

Before we propose a sounding control scheme, we firstly investigate channel coherence time characteristics in real WLAN environments including typical office and conference environments. We put an AP and distribute 9 mobile nodes across the space. The AP is set to continuously broadcast short packets, and we concurrently measure DL channels between the AP and each mobile node using the CSI measurement tool [37]. From the collected channel data, we calculate the channel coherence times for the 9 nodes.

ZF precoding eliminates inter-user interference based on the directional information of channels, and hence, we calculate the channel coherence time to be the largest expected time duration that the directional information does not significantly change, i.e., $T_{c,n}$ $(n = 1, 2, \ldots, 9)$ is calculated to be the largest value of $\tau > 0$, which satisfies the following constraint:

$$\mathbb{E}_t \left| \frac{\mathbf{h}_n(t+\tau)^H \mathbf{h}_n(t)}{\|\mathbf{h}_n(t+\tau)\| \|\mathbf{h}_n(t)\|} \right|^2 \geq \delta, \tag{3.1}$$

where $\mathbf{x}^H$ and $\|\mathbf{x}\|$ are respectively the conjugate transpose and norm of a vector $\mathbf{x}$, and δ is a threshold. In [40], it is shown that inter-user interference cancellation requires very accurate CSI, and we take a strict threshold $\delta = 0.95$ for the coherence time.

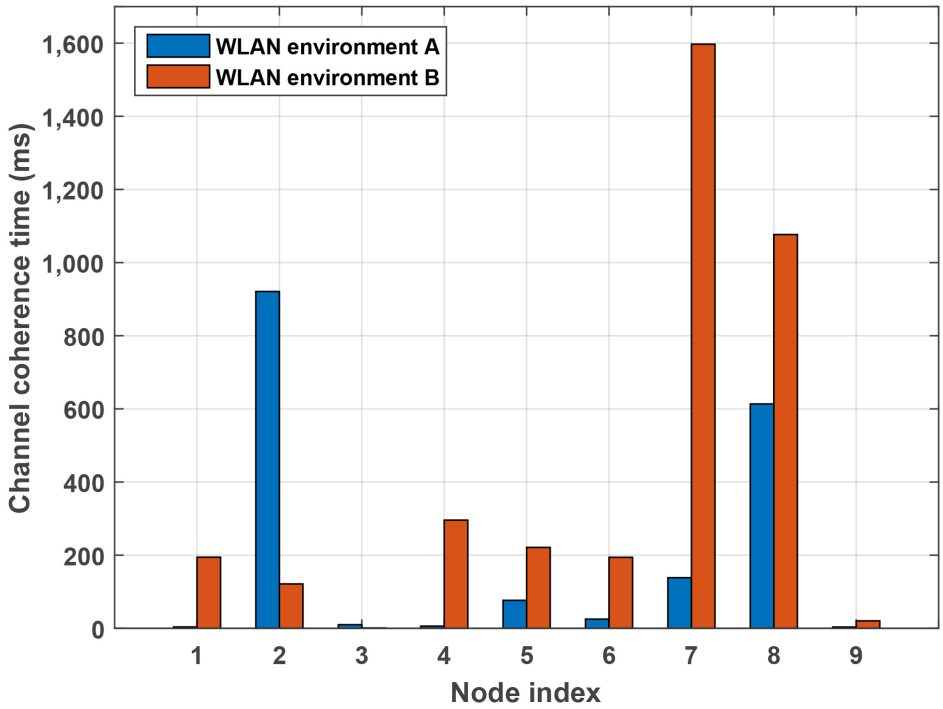


Figure 3.2: MIMO channel coherence time characteristics in real environments.

Fig. 3.2 shows the measurement results in two different real WLAN environments. Even though all the mobile nodes are in the same space, different wireless links experience different propagation environment variation due to mobility, behavior of nearby people and so on. For example, some wireless links experience stationary propagation environments such that their channel coherence times are several hundreds of milliseconds. On the other hand, some of the other wireless links are moderately time-varying such that their channel coherence times are several tens of milliseconds. Lastly, the rest wireless links vary quite fast such that their channel coherence times are several milliseconds.

When wireless links have different coherence times, accurate CSI is guaranteed only for the nodes whose channel coherence times are larger than the sounding interval. Accordingly, determining the sounding interval is interpreted as determining the sounding node set, i.e., the set of nodes to guarantee accurate CSI at the AP. As we will discuss in the following, determination of the sounding node set requires comprehensive consideration of the network environment including channel coherence times, traffic loads, and so on.

3.4 Proposed Sounding Control

Let us assume that sounding is periodically performed with a specific sounding interval T_s. As T_s becomes smaller, accurate CSI is guaranteed for more devices, i.e., the MU-MIMO throughput gain increases at the cost of the sounding overhead increase. On the other hand, as T_s becomes larger, the sounding overhead decreases, but the MU-MIMO throughput gain decreases. Considering the trade-off relationship, we formulate the sounding interval determination problem as

$$T_s^* = \underset{T_s>0}{\arg\max} \left[T_G\left(T_s\right) - T_O\left(T_s\right) \right], \tag{3.2}$$

where $T_G\left(T_s\right)$ and $T_O\left(T_s\right)$ are respectively the long-term expected MU-MIMO time gain per second and the expected sounding time overhead per second for a given sound-

ing interval T_s. In the following subsection, we will justify (3.2) by deriving both $T_G(T_s)$ and $T_O(T_s)$.

3.4.1 Derivation of $T_G(T_s)$ and $T_O(T_s)$

When the sounding process is periodically performed with an interval T_s, the AP has accurate CSI only for nodes whose channel coherence times are larger than T_s. The set of such nodes is referred to as the sounding node set, and it is formally defined as $\Lambda(T_s) = \{n | T_{c,n} \geq T_s,\ n = 1, 2, \ldots, N\}$.

If MU-MIMO transmission involves nodes not in $\Lambda(T_s)$, inter-user interference severely degrades the throughput due to imperfect CSI at the AP. Therefore, involving them in sounding only ends up increasing sounding overhead without throughput improvement. Only the members of $\Lambda(T_s)$ are involved in both sounding and MU-MIMO transmission. Then, the time overhead per sounding process is given by

$$\begin{aligned} &T_{\text{sounding}}(T_s) = \\ &T_{\text{NDPA}} + T_{\text{NDP}} + |\Lambda(T_s)|\,(T_{\text{SIFS}} + T_{\text{F}}) + (|\Lambda(T_s)| - 1)\,T_{\text{poll}}, \end{aligned} \tag{3.3}$$

where $|\Lambda(T_s)|$ is the cardinality of set $\Lambda(T_s)$. Because sounding is periodically performed with an interval T_s, the sounding time overhead per second is given by

$$T_O(T_s) = \frac{T_{\text{sounding}}(T_s)}{T_s}. \tag{3.4}$$

Now, we derive $T_G(T_s)$ to complete the problem formulation of (3.2). Let us assume that the AP achieves a medium access opportunity with a packet destined for a node. Then, if the destination node is an element of $\Lambda(T_s)$, it becomes a primary user, and MU-MIMO transmission is available with some other node(s) in $\Lambda(T_s)$ as secondary user(s). Therefore, we formulate the long-term expected MU-MIMO time gain $T_G(T_s)$ to be the expectation over all possible primary user cases as shown below.

$$T_G(T_s) = \sum_{i \in \Lambda(T_s)} \rho_i T_{G|i}(T_s), \tag{3.5}$$

where ρ_i is interpreted here as the probability that node i becomes a primary user, and $T_{G|i}(T_s)$ is the expected MU-MIMO time gain per second when node i is a primary user.

When a packet destined for node i is at the head of the AP queue, there are $2^{|\Lambda(T_s)|-1}$ cases according to whether the AP queue also has packets destined for each of the other nodes in $\Lambda_i(T_s) = \Lambda(T_s) - \{i\}$. Hence, $T_{G|i}(T_s)$ is also given by the expectation over the possible cases as shown in the following equation.

$$T_{G|i}(T_s) = \sum_{\Upsilon \in \Omega_i(T_s)} P(\Upsilon, \Lambda_i(T_s)) T_{G|i,\Upsilon}(T_s), \tag{3.6}$$

where $\Omega_i(T_s)$ is the set of all $2^{|\Lambda_i(T_s)|}$ subsets of $\Lambda_i(T_s)$, $T_{G|i,\Upsilon}(T_s)$ is the expected MU-MIMO time gain per second given that a packet destined for node i is at the head of the AP queue and it also has packets destined for nodes in $\Upsilon \subset \Lambda_i(T_s)$, and $P(\Upsilon, \Lambda_i(T_s))$ means the probability that the AP queue has packets destined for nodes in Υ, which is given by

$$P(\Upsilon, \Lambda_i(T_s)) = \prod_{j \in \Upsilon} p_j \prod_{j \in \Lambda_i(T_s) - \Upsilon} (1 - p_j). \tag{3.7}$$

Then, we need to derive $T_{G|i,\Upsilon}(T_s)$ to calculate $T_G(T_s)$ and complete the problem formulation of (3.2). Let us assume that a packet destined for node i is at the head of the AP queue at a given time, and the AP queue also has packets destined for nodes in set $\Upsilon \subset \Lambda_i(T_s)$. In IEEE 802.11ac, the maximum number of nodes simultaneously served with MU-MIMO is defined to be $k_{\max} = \min(M, 4)$. Therefore, we define $\Gamma(\Upsilon)$ to be the set of nodes which have the $(k_{\max} - 1)$ largest SNRs among nodes in set Υ, and the long-term achievable ergodic DL sum rate (bps/Hz) with ZF is approximated by the upper bound [41] as shown in the following equation.

$$\begin{aligned} &C(\{i\} \cup \Gamma(\Upsilon)) = \\ &\textstyle\sum_{j \in \{i\} \cup \Gamma(\Upsilon)} \log_2 \left(1 + \frac{M - |\{i\} \cup \Gamma(\Upsilon)| + 1}{|\{i\} \cup \Gamma(\Upsilon)|} \gamma_j\right). \end{aligned} \tag{3.8}$$

When the packet destined for node i is solely transmitted, the ergodic sum rate is

$C(\{i\})$, and hence, the average sum rate gain through MU-MIMO is given by $C(\{i\} \cup \Upsilon) - C(\{i\})$.

If we additionally consider the protocol overhead as shown in Fig. 3.1, $T_{G|i,\Upsilon}(T_s)$ is given by

$$T_{G|i,\Upsilon}(T_s) = \frac{C(\{i\} \cup \Gamma(\Upsilon)) - C(\{i\})}{C(\{i\})} - |\Gamma(\Upsilon)| \frac{T_{\text{SIFS}} + T_{\text{BAR}} + T_{\text{SIFS}} + T_{\text{BA}}}{T_i}, \tag{3.9}$$

where T_i is the transmission time when the packet destined for node i is solely transmitted. We obtain $T_G(T_s)$ by substituting (3.9) and (3.7) for $T_{G|i}(T_s)$ and $P(\Upsilon, \Lambda_i(T_s))$ in (3.6), and again substituting (3.6) for $T_{G|i}(T_s)$ in (3.5).

Now, as per our above discussion, we have formulated $T_G(T_s)$ and $T_O(T_s)$ to have the same unit, which means that they are directly comparable. Therefore, maximizing $T_G(T_s) - T_O(T_s)$ could be interpreted as maximization of the net time gain per second considering both MU-MIMO time gain and sounding time overhead, and we obtain T_s^* as the solution of (3.2). If $T_G(T_s) - T_O(T_s) \leq 0$ at T_s^*, it means that sounding overhead is expected to be larger than the long-term expected MU-MIMO gain in the network environment, and hence, T_s^* is set to ∞, i.e., MU-MIMO is not used at all.

3.4.2 Efficient Determination of Sounding Node Set and Sounding Interval

Exhaustively finding the solution of (3.2) among all possible values of T_s (> 0) incurs extremely heavy computational complexity. Therefore, we develop a method to efficiently find T_s^* and $\Lambda(T_s^*)$ from the following lemma.

Lemma 1.

$$T_s^* = \arg\max_{T_s > 0} [T_G(T_s) - T_O(T_s)] = \arg\max_{T_s \in \Phi} [T_G(T_s) - T_O(T_s)],$$

where $\Phi = \{T_{c,1}, T_{c,2}, \ldots, T_{c,N}\}$.

Proof. $T_G(T_s)$ and $T_{\text{sounding}}(T_s)$ have relations with T_s through $\Lambda(T_s)$. Therefore, as long as $\Lambda(T_s)$ does not change, $T_G(T_s)$ and $T_{\text{sounding}}(T_s)$ remain constant for varying T_s. Therefore, if T_s is increased between any two consecutive values among all channel coherence times $T_{c,n}$'s $(n = 1, 2, \ldots, N)$, $T_G(T_s)$ and $T_{\text{sounding}}(T_s)$ do not change. However, $T_O(T_s)$ decreases because its numerator $T_{\text{sounding}}(T_s)$ remains constant while its denominator increases. From the above reasoning, we can conclude that T_s^* should be one of $T_{c,n}$'s $(n = 1, 2, \ldots, N)$, and the original feasible set $T_s > 0$ in (3.2) can be replaced with $T_s \in \Phi$. □

From the above lemma, T_s^* could be obtained by comparing $T_G(T_s) - T_O(T_s)$ at only N candidate values. When we compare $T_G(T_s) - T_O(T_s)$ at candidate values in Φ, calculating (3.6) becomes a bottleneck in terms of computation complexity due to summation over all $2^{|\Lambda(T_s)|-1}$ cases. However, it can be calculated in a polynomial time if we make use of the fact that $\Gamma(\Upsilon)$ in (3.9) is determined by nodes with $(k_{\max} - 1)$ largest SNRs in Υ. Let Ψ_i be an ordered set of nodes in $\Lambda_i(T_s)$, where they are ordered in a decreasing order of γ_n and its subset composed of the $l_1, l_2, \ldots$-th elements is denoted as $\Psi_i^{(l_1,l_2,\ldots)}$. When $k_{\max}$ has the largest value of 4, $\Gamma(\Upsilon) = \Psi_i^{(1,2,3)}$ whenever $\Psi_i^{(1,2,3)} \subset \Upsilon$. If $\Psi_i^{(1)} \not\subset \Upsilon$ and $\Psi_i^{(2,3,4)} \subset \Upsilon$, $\Gamma(\Upsilon) = \Psi_i^{(2,3,4)}$. Therefore, (3.6) can be represented as shown below.

$$T_{G|i}(T_s) = \sum_{l_1=1}^{|\Lambda_i(T_s)|} \sum_{l_2=l_1+1}^{|\Lambda_i(T_s)|} \sum_{l_3=l_2+1}^{|\Lambda_i(T_s)|} \left[T_{G|i,\Psi_i^{(l_1,l_2,l_3)}} \prod_{j \in \Psi_i^{(l_1,l_2,l_3)}} p_j \prod_{j \in \Psi_i^{(1,2,\ldots,l_3)} - \Psi_i^{(l_1,l_2,l_3)}} (1 - p_j) \right]. \tag{3.10}$$

If we additionally consider the summation in (3.5), the computation complexity for calculating $T_G(T_s) - T_O(T_s)$ increases in the order of N^4 when $k_{\max} = 4$. Considering that the determination of the sounding interval and sounding node set is performed at an AP or AP controller with long-term intervals, we assess that the polynomial time calculation is not a severe burden.

Table 3.1: Two different traffic load cases in the measured WLAN environment A.

	Node index	1	2	3	4	5	6	7	8	9
	$T_{c,n}$ (ms)	3.74	921.04	9.91	6.96	76.65	25.17	138.50	613.46	3.63
Case 1	p_n	0.3548	0.1094	0.4891	0.4330	0.5805	0.7942	0.5987	0.0625	0.0883
	$T_G(T_{c,n}) - T_O(T_{c,n})$	−0.2594	−0.0005	0.2831	0.1277	0.1518	**0.3387**	0.0227	0.0001	−0.4023
Case 2	p_n	0.8657	0.0752	0.1333	0.4043	0.9327	0.0898	0.8127	0.8414	0.8323
	$T_G(T_{c,n}) - T_O(T_{c,n})$	−0.3633	−0.0005	0.1963	−0.0718	**0.4334**	0.3670	0.2209	0.0166	−0.3744

3.5 Performance Evaluation

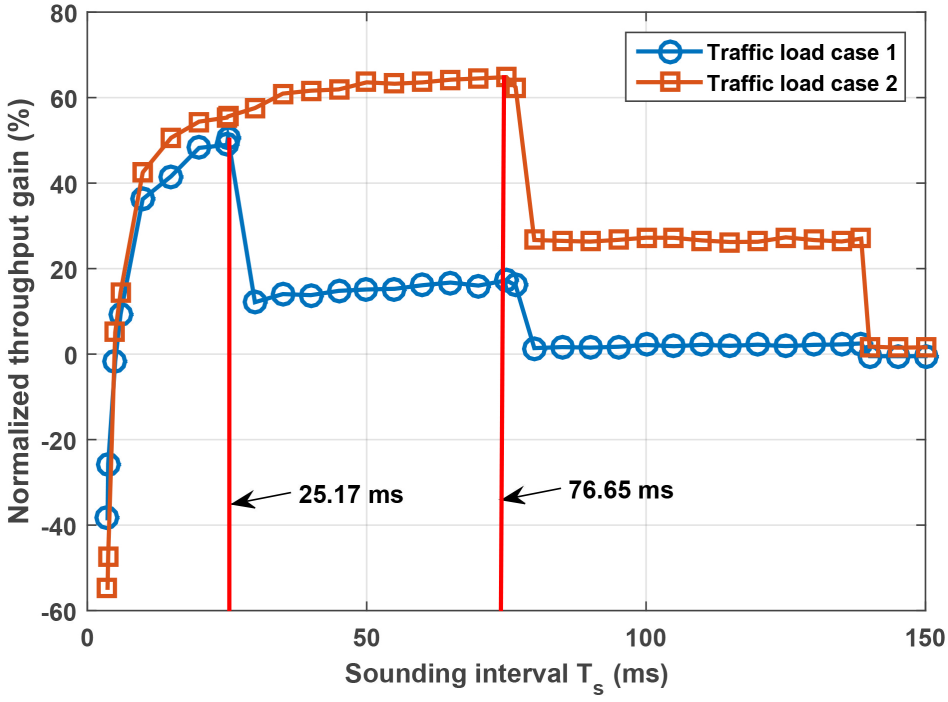
In this section, we evaluate the performance of the proposed scheme by employing the real channel data traces measured in the real WLAN environments introduced in Section 3.3. We consider a WLAN with $M = 3$ and $N = 9$. Both $\mathbf{h}_n(t)$ and γ_n $(n = 1, 2, \ldots, 9)$ are given by the real channel data traces. Nodes do not have uplink traffic, and the n-th node's DL traffic load, p_n, is randomly selected between $[0, 1]$. We model the AP queue to have the DL traffic loads, and set the transmission time T_n to 1 ms. The DL time channel utilization ρ_n $(n = 1, 2, \ldots, 9)$ is obtained by measuring the long-term time portion used for the n-th node when MU-MIMO is not used. IEEE 802.11ac parameters are employed for T_{NDPA}, T_{NDP}, T_{SIFS}, T_{poll}, T_{F}, T_{BA}, and T_{BAR}, where T_{F} is determined with quantization of $\psi = 7$, $\phi = 9$ bits without subcarrier grouping [24]. For MU-MIMO transmission, the secondary users are selected to maximize the instantaneous sum rate with ZF based on the channel knowledge. In the following simulations, the MU-MIMO throughput gain is normalized over the throughput obtained without MU-MIMO through maximum ratio transmission (MRT) beamforming [41].

To show that the proposed scheme adaptively determines the sounding interval and sounding node set according to the network environment, we consider two different traffic load cases in each of the measured WLAN environments as shown in Tables 3.1 and 3.2. As per our discussion in Section 4.2.3, the proposed scheme makes a candidate

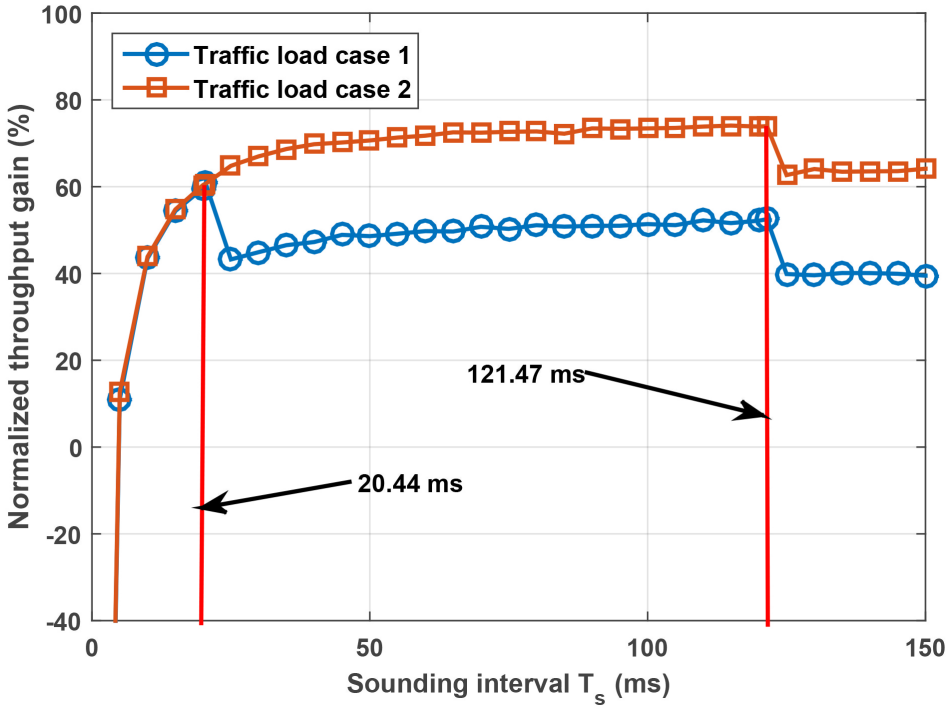
Table 3.2: Two different traffic load cases in the measured WLAN environment B.

	Node index	1	2	3	4	5	6	7	8	9
	$T_{c,n}$ (ms)	194.49	121.47	1.04	295.85	221.26	194.23	1597.2	1076.8	20.44
Case 1	p_n	0.2700	0.7721	0.5268	0.8385	0.1103	0.0807	0.4722	0.9849	0.9739
	$T_G(T_{c,n}) - T_O(T_{c,n})$	0.2683	0.4300	-2.6479	0.2219	0.2285	0.2773	-0.0003	0.1039	$\mathbf{0.5237}$
Case 2	p_n	0.8678	0.9078	0.9437	0.9698	0.9340	0.6310	0.8259	0.7996	0.0170
	$T_G(T_{c,n}) - T_O(T_{c,n})$	0.3770	$\mathbf{0.6141}$	-2.6333	0.1820	0.2694	0.4847	-0.0003	0.1063	0.4823

sounding interval set which consists of nodes' channel coherence times, and compares $T_G(T_s) - T_O(T_s)$ at each candidate sounding interval. Tables 3.1 and 3.2 also show the candidate sounding interval set, and $T_G(T_s) - T_O(T_s)$ at each candidate sounding interval for the two different traffic load cases in the two measured WLAN environments. The proposed scheme suggests T_s^* to be 25.17 ms and 76.67 ms respectively for the two traffic load cases in the measured WLAN environment A, and to be 20.44 ms and 121.47 ms for those in the measured WLAN environment B. For given sounding interval, the sounding node set is composed of nodes whose channel coherence times are larger than T_s^*.



(a) Measured WLAN environment A.



(b) Measured WLAN environment B.

Figure 3.3: Normalized MU-MIMO throughput gain versus T_s.

Fig. 3.3(a) shows the average normalized throughput gain versus the sounding interval for the two traffic load cases in the measured WLAN environment A. We observe that the proposed scheme adaptively determines the sounding interval and sounding node set according to the network environment variation. Compared to the the first traffic load case, node 6 has a smaller DL traffic load in the second traffic load case. Therefore, the proposed scheme excludes node 6 from the sounding node set by increasing the sounding interval from 25.17 ms to 76.67 ms. We can see that the maximum throughput is actually achieved when the sounding interval is set to the recommended values in the two traffic load cases. Sounding interval smaller or larger than the recommended values results in throughput degradation due to increased sounding overhead or inaccurate CSI for nodes which lead to a significant MU-MIMO throughput gain with accurate CSI.

Fig. 3.3(b) shows the average normalized throughput gain versus the sounding interval for the two traffic load cases in the measured WLAN environment B. Although the channel coherence time characteristics are quite different from those in the measured WLAN environment A, we also observe that the proposed scheme adaptively finds the proper sounding intervals which lead to the maximum MU-MIMO throughput gain by comprehensively considering the network environment.

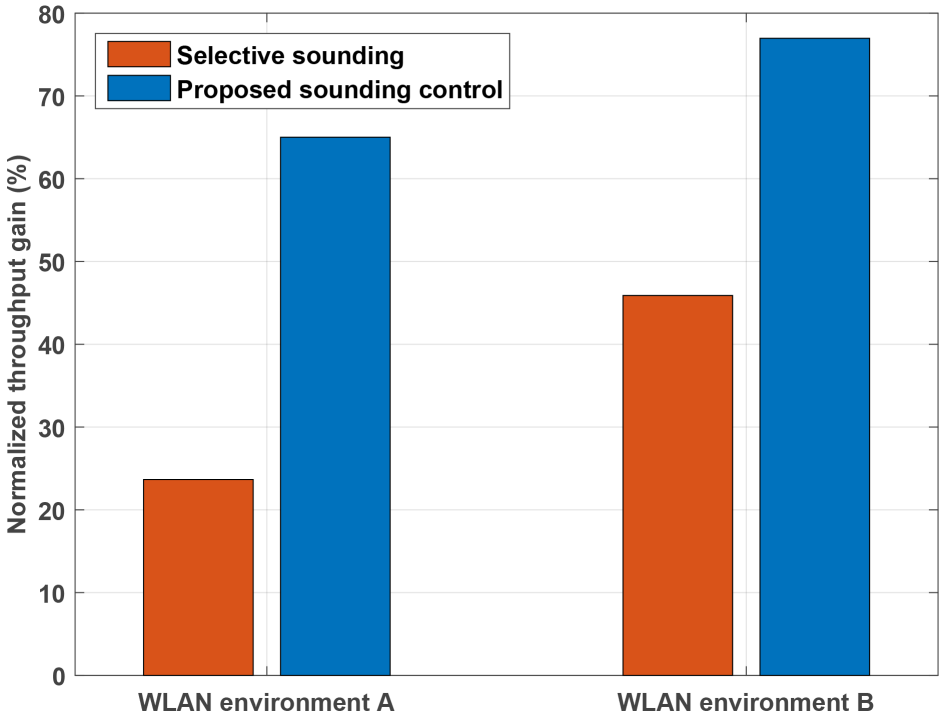


Figure 3.4: Comparison with selective sounding.

Fig. 3.4 shows the average normalized throughput gain over different realizations of traffic loads in the measured WLAN environments. The selective sounding scheme [32] considers channel coherence times only, and hence, it wastes precious time resources to guarantee accurate CSI for nodes which do not lead to a large throughput gain due to small DL traffic loads and/or small channel coherence times. On the other hand, the proposed scheme analytically finds the sounding interval which maximizes the MU-MIMO throughput gain in comprehensive consideration of the network environment. Therefore, the proposed scheme leads to about 65% and 76% of throughput improvements respectively in the two measured WLAN environments while the selective sounding scheme only leads to about 24% and 45% of throughput improvements.

Chapter 4

Adaptive Group ID Control for Idle Power Consumption Mitigation in IEEE 802.11ac WLANs

4.1 Group ID and Power Saving Mechanism in IEEE 802.11ac MU-MIMO

A node group consists of up to four nodes, where a GID is assigned to each node group while specific positions are also assigned to member nodes. Node groups are then distinguished by GIDs, and members in a node group are distinguished by their positions. When an AP transmits multiple data streams to a group of nodes, the target node group and destination nodes are notified by the GID and positions, both found at the VHT-SIG-A field in the PHY header. Utilizing the fact that the PHY header contains the target GID and node positions, IEEE 802.11ac defines a built-in power saving mechanism in which non-destination nodes of a transmission stop decoding the incoming packet after decoding the PHY header and identifying the GID and position information.

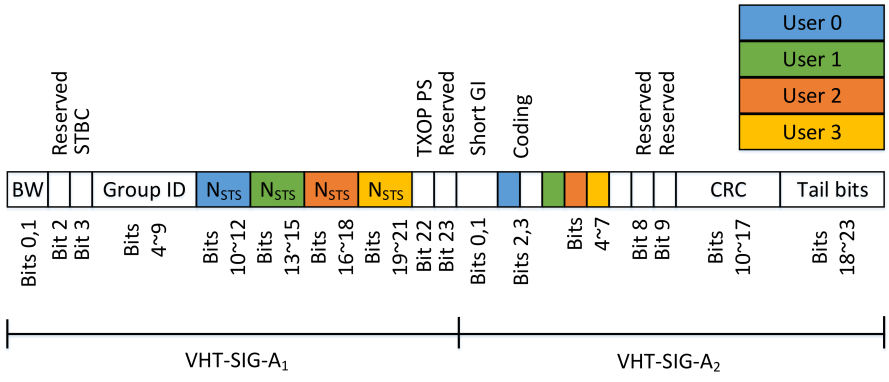


Figure 4.1: VHT-SIG-A field in IEEE 802.11ac.

Fig. 4.1 shows the detailed format of the VHT-SIG-A field, where only 6 bits are allocated for GIDs, which enables $2^6 = 64$ different combinations. Two of the 64 combinations are reserved for other purposes, and hence, the number of available GIDs becomes only 62. Once the number of associated nodes reaches 8, the number of possible node groups, i.e., $\binom{8}{4} = 70$ becomes larger than the number of available GIDs. Due to the limited number of available GIDs, IEEE 802.11ac allows multiple node groups to share the same GID, i.e., GIDs could be overloaded.

Determining which GIDs to be assigned to which node groups is an implementation-dependent algorithmic problem, and besides, overloaded GIDs degrade the performance of the aforementioned built-in power saving mechanism. Let us assume that two different node groups share the same GID, and an AP transmits multiple data streams to nodes in the first group using the shared GID. Then, nodes in the second group are not able to realize that the transmission is not destined to them even after identifying the GID information in the PHY header, which ends up with idle power consumption at nodes in the second group. Accordingly, as more GIDs overloaded, idle power consumption increases. Accordingly, we need to develop a GID control scheme considering power saving performance.

4.2 Proposed GID Control

4.2.1 System Description

We consider a WLAN composed of an AP and N associated nodes. We assume that N is larger than or equal to 8, and hence, the number of possible combinations of node groups is larger than the number of available GIDs. At the AP, nodes have heterogeneous downlink (DL) traffic loads, and hence, the DL time resources are also heterogeneously used for them according to their traffic loads. The DL time channel utilization for the n-th node is denoted as ρ_n $(0 \leq \rho_n \leq 1)$ representing the portion of DL time resources utilized for the n-th node $(n = 1, 2, \ldots, N)$. Apparently, as a node

has a larger DL traffic load, its DL time channel utilization increases.

4.2.2 Definition of GID Overloading Node Set and Idle Power Consumption Minimization Problem

Let us assume that a set of nodes share a common GID and position. Then, whenever AP serves one of them with the shared GID, all the other nodes suffer from idle-listening. As more nodes share the same GID and position, and as the AP more frequently serve them using the shared GID, idle power consumption obviously increases. Therefore, it could considerably mitigate idle power consumption that common GIDs and positions are assigned to nodes with small DL traffic loads. To this end, idle power consumption should be regulated to have effect only on intended nodes. We newly develop a GID assignment mechanism by defining the concept of GID overloading node set, where its member nodes share the same GID and position. Accordingly, when the AP serves one of those member nodes, only the rest nodes suffer from idle-listening.

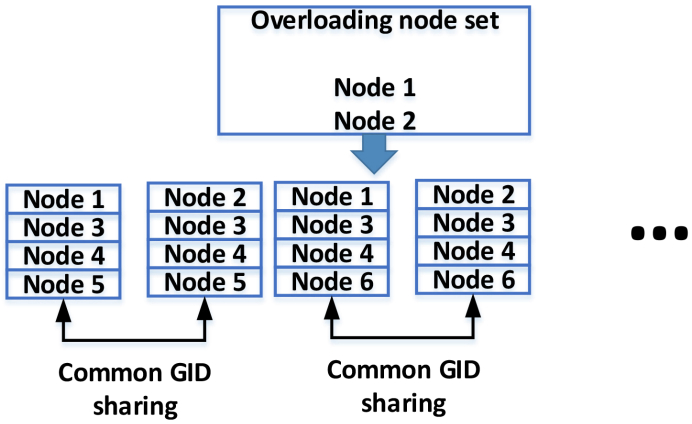


Figure 4.2: The concept of GID overloading node set.

For example, let us assume that a GID overloading node set is composed of nodes 1 and 2. Then, as shown in Fig. 4.2, common GIDs are assigned to all node groups whose members are respectively nodes 1, n_3, n_4, n_5 and nodes 2, n_3, n_4, n_5 for all combinations of nodes n_3, n_4, n_5 while nodes 1 and 2 take the same position in node groups sharing the same GID. Accordingly, whenever the AP serves node 1 via MU-MIMO, only node 2 suffers from idle-listening, and vice versa.

Let us now consider a general case in which we generate a set $\Theta = \{\Psi^{(1)}, \Psi^{(2)}, \ldots\}$, where $\Psi^{(n)}$ is a GID overloading node set. Then, nodes in each GID overloading set always share common GIDs and positions as discussed above, which leads to all nodes in each GID overloading node set being regarded as a single node. Accordingly, a node does not belong to multiple GID overloading node sets, and the number of GIDs required for handling node groups is reduced to

$$\phi\left(\Theta, N\right) = \binom{N - \sum_{\mu\in\Theta} |\mu| + |\Theta|}{4}, \tag{4.1}$$

where $|\mu|$ is the cardinality of set μ.

Remind that whenever the AP serves a node in a GID overloading node set, all the other nodes in the GID overloading node set suffer from idle-listening. Therefore, GID overloading node sets in Θ end up with total idle power consumption given by

$$\varphi\left(\Theta\right) = \sum_{\mu\in\Theta} \left\{ \left(|\mu| - 1\right) P_{\text{loss}} \sum_{j\in\mu} \rho_j \right\}, \tag{4.2}$$

where P_{loss} is the idle power loss at a node when it suffers from idle-listening. Eq. (4.2) reflects the fundamental nature of GID overloading. The number of nodes suffering from idle-listening increases as more nodes belong to a GID overloading node set. Accordingly, total idle power consumption also increases as GID overloading node sets consist of a larger number of nodes. In addition, member nodes of a GID overloading node set more frequently suffer from idle-listening as the AP more frequently serves them. Therefore, total idle power consumption also increases as member nodes of GID overloading node sets have larger DL time channel utilization.

Now, the above analysis leads to an idle power consumption minimization problem as follows.

$$\begin{aligned} \underset{\Theta \in \Omega}{\text{Minimize}} \quad & \varphi(\Theta) \\ \text{subject to} \quad & \phi(\Theta, N) \leq 62, \end{aligned} \tag{4.3}$$

where all possible realizations of GID overloading set generation is denoted as Ω, and the total expected idle power consumption is minimized while the number of GIDs required for handling node groups is smaller than or equal to the number of available GIDs, i.e., 62.

4.2.3 Efficient GID Overloading Algorithm

Once we generate GID overloading node sets as the solution of (4.3), node groups are handled with the available GIDs while minimizing expected idle power consumption induced from the overloaded GIDs. However, the number of possible realizations in Ω exponentially increases with growth of N, and hence, the computational complexity for exhaustively finding the solution of (4.3) could be unacceptable for a practical AP processing unit when the network scale grows. In this subsection, we derive some lemmas which give insight into the solution of (4.3), and develop an algorithm to efficiently find the solution.

Let us denote the solution of (4.3) as Θ^*. Then, Θ^* should satisfy the following lemmas.

Lemma 2. *If K ($\leq N$) nodes belong to GID overloading node sets in Θ^*, i.e., $K = \sum_{\mu \in \Theta^*} |\mu|$, they should have the K smallest DL time channel utilization values among all the N nodes.*

Proof. Let Υ be the set of nodes involved in generating GID overloading node sets in Θ^*, i.e., $\Upsilon = \{j | j \in \mu, \mu \in \Theta^*\}$. If nodes in set Υ do not have the smallest time channel utilization, the following equation holds.

$$\max_{j \in \Upsilon} \rho_j < \min_{j \in \Upsilon^C} \rho_j. \tag{4.4}$$

Then, idle power consumption could be further reduced by utilizing the fact that (4.1) has relations only with the cardinality of GID overloading node sets, and (4.2) decreases with smaller time channel utilization. Let us replace the node whose time channel utilization is largest in Υ with that whose one is smallest in Υ^C, which leads to further reduction of (4.2) without any change in (4.1). From the above reasoning, Lemma 1 holds. □

Lemma 3. *Let us assume that any two overloading node sets $\Psi^{(i)}$ and $\Psi^{(j)}$ in Θ^* have different cardinality such that $\left|\Psi^{(i)}\right| > \left|\Psi^{(j)}\right|$. Then, the following equation holds.*

$$\max_{w \in \Psi^{(i)}} \rho_w \leq \min_{w \in \Psi^{(j)}} \rho_w.$$

Proof. Let us assume that $\max_{w \in \Psi^{(i)}} \rho_w > \min_{w \in \Psi^{(j)}} \rho_w$. If the two nodes with the largest and smallest time channel utilization in $\Psi^{(i)}$ and $\Psi^{(j)}$ are exchanged with each other, the cardinality of $\Psi^{(i)}$ and $\Psi^{(j)}$ does not change, and the expected idle power consumption (4.2) further reduces by $\left(\left|\Psi^{(i)}\right| - \left|\Psi^{(j)}\right|\right)\left(\max_{w \in \Psi^{(i)}} \rho_w - \min_{w \in \Psi^{(j)}} \rho_w\right)$. From the above reasoning, Lemma 2 holds. □

Lemma 4. *If we consider all possible cases in which GID overloading node sets are differently generated with a set of nodes,* (4.2) *is minimized when the number of GID overloading node sets with least cardinality is maximized.*

Proof. Looking into the form of (4.2), each node in a GID overloading node set contributes to expected idle power consumption proportional to both its DL time channel utilization and the cardinality of the GID overloading node set that it belongs to. Accordingly, considering all possible cases in which GID overloading node sets are differently generated with a set of nodes, (4.2) is minimized when the number of GID overloading node sets with least cardinality is maximized. □

Let us assume that $\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(L)} \in \Theta^*$, where L is the total number of GID overloading node sets in Θ^*. Then, Lemma 1 says that nodes belonging to $\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(L)}$ should have $\sum_{l=1}^{L} \left|\Psi^{(l)}\right|$ smallest DL time channel utilization

Algorithm 2: Greedy search-based efficient GID control algorithm

Initialization

$\Xi =$ ordered set of nodes in the increasing order of ρ_n

$P = \infty$

Searching for the solution of (4.3)

for $m = 1, 2, \ldots$ **do**

  $\Pi = \left\{\Xi_1, \Xi_2, \ldots, \Xi_{N-6+(m-1)}\right\}$

  $\Omega^{(m)} =$ GID overloading node set generation cases with Π

  for $l = 1, 2, \ldots, \left|\Omega^{(m)}\right|$ **do**

    if $\phi\left(\Omega_l^{(m)}, N\right) \leq 62$ **then**

      if $P > \varphi\left(\Omega_l^{(m)}\right)$ **then**

        $\Theta = \Omega_l^{(m)}$

        $P = \varphi\left(\Omega_l^{(m)}\right)$

      end

      if $l = 1$ **then**

        Stop searching

      end

    end

  end

end

Return: Θ

values among N associated nodes. Without loss of generality, Lemma 2 says that $\left|\Psi^{(i)}\right| \geq \left|\Psi^{(j)}\right|$ and $\max\limits_{w \in \Psi^{(i)}} \rho_w \leq \min\limits_{w \in \Psi^{(j)}} \rho_w$ for $i < j$. Accordingly, we need not exhaustively search for Θ^* among all possible realizations of GID overloading node set generation, i.e., the computational complexity could be considerably reduced by a greedy search-based algorithm, where nodes with the smallest DL time channel utilization are sequentially selected as shown in Algorithm 2.

We firstly generate an ordered set Ξ of all the associated nodes, where nodes are

ordered in the increasing order of ρ_n, and Ξ_i denotes the i-th element. If we consider all possible cases in which GID overloading node sets are differently generated with nodes in Υ, (4.1) is minimized when a single GID overloading node set is generated. When $|\Upsilon| < N - 6$, the minimized value of (4.1) is still larger than 62, which means that GIDs should be overloaded using at least $N - 6$ nodes. Accordingly, at the m-th stage, we select $N - 6 + (m - 1)$ nodes with smallest DL time channel utilization, i.e., $\Xi_1, \Xi_2, \ldots, \Xi_{N-6+(m-1)}$, and then consider all possible cases in which GID overloading node sets are differently generated with those nodes.

We define $\Omega^{(m)}$ to be the set of all possible cases in which GID overloading node sets are differently generated at the m-th stage. For example, when Ξ_1 and Ξ_2 are selected, $\Omega^{(m)}$ consists of one case in which those two nodes generate a single GID overloading node set presented by $\Psi^{(1)} = \{\Xi_1, \Xi_2\}$. When $\Xi_1, \Xi_2, \ldots, \Xi_6$ are selected, $\Omega^{(m)}$ consists of 4 different cases according to Lemmas 1 and 2. The first case is composed of three GID overloading node sets $\Psi^{(1)} = \{\Xi_1, \Xi_2\}$, $\Psi^{(2)} = \{\Xi_3, \Xi_4\}$, and $\Psi^{(3)} = \{\Xi_5, \Xi_6\}$. The second case is composed of two GID overloading node sets $\Psi^{(1)} = \{\Xi_1, \Xi_2, \Xi_3\}$, and $\Psi^{(2)} = \{\Xi_4, \Xi_5, \Xi_6\}$. The third case is composed of two GID overloading node sets $\Psi^{(1)} = \{\Xi_1, \Xi_2, \Xi_3, \Xi_4\}$, and $\Psi^{(2)} = \{\Xi_5, \Xi_6\}$. Lastly, the fourth case is composed of one GID overloading node set $\Psi^{(1)} = \{\Xi_1, \Xi_2, \Xi_3, \Xi_4, \Xi_5, \Xi_6\}$.

At the m-th stage, we denote the l-th case as $\Omega_l^{(m)}$, and $\Omega_1^{(m)}$ as the case in which the number of GID overloading node sets with least cardinality is maximized among all cases in $\Omega^{(m)}$. According to Lemma 3, $\Omega_1^{(m)}$ has the minimum expected idle power consumption at the m-th stage. The algorithm stops searching at the M-th stage, where GID overloading node sets in $\Omega_1^{(M)}$ leads to (4.1) smaller than the number of available GIDs. According to Lemmas 1, 2, and 3, further searching only ends up with cases that have larger idle power consumption, and hence, the algorithm returns the best case among all the cases considered up to that point.

4.3 Performance Evaluation

In this section, we evaluate the performance of the proposed GID control. An AP has heterogeneous packet enqueue rates for the N associated nodes which do not have uplink traffic. Each node's packet enqueue rate at the AP queue is randomly selected according to the uniform distribution with range $[0, 1]$, where the packet enqueue rate for a node represents the probability that a packet destined to it is enqueued at a given time. Lastly, MU-MIMO transmission is available for up to 4 nodes at a time [24], and P_{loss} is set 1,220 mW [33].

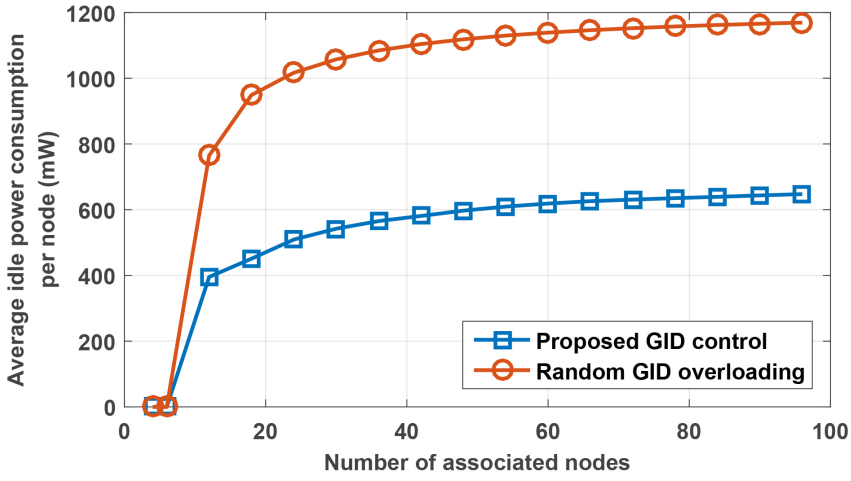


Figure 4.3: Power saving performance of the proposed scheme.

Fig. 4.3 shows the average idle power consumption of a node per transmission according to the number of associated nodes, N. When $N < 8$, the number of possible node groups is smaller than the number of available GIDs, thus resulting in no idle power consumption. With further increase of N, idle power consumption increases due to the limited number of available GIDs. However, as discussed in Section 4.2.2, the proposed GID control assigns GIDs in consideration of nodes' DL traffic loads, thus mitigating idle power consumption by about 46% compared to random GID overloading.

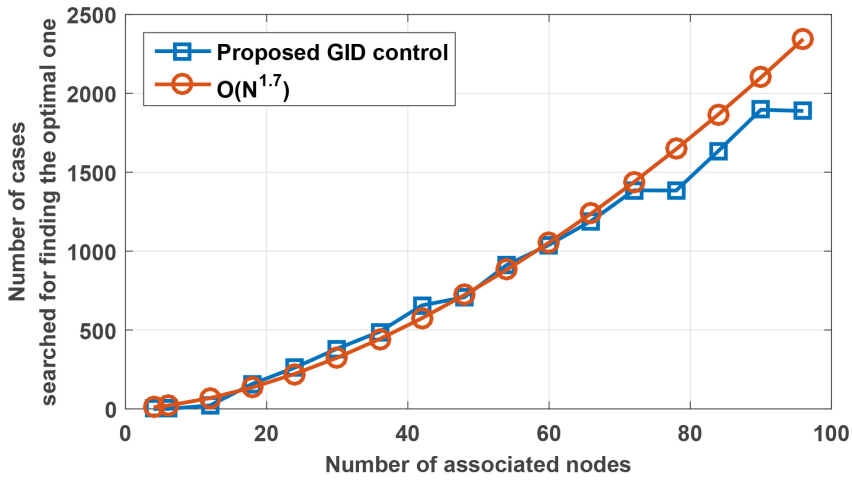


Figure 4.4: Computational complexity versus number of nodes.

Fig. 4.4 shows the computational complexity of Algorithm 2, i.e., the number of cases searched before obtaining Θ^* with growth of N. As N increases, the computational burden becomes heavier. However, Algorithm 2 efficiently finds Θ^*, which leads to the computational burden increasing in an order similar to $O\left(N^{1.7}\right)$. Considering that GID control is performed with long-term intervals, the polynomial calculation time is assessed to be affordable for a practical AP or AP controller.

Chapter 5

Conclusion

In part I of this dissertation, we have developed a new theoretical performance analysis model and CW control scheme for practical IEEE 802.11 WLANs with node heterogeneity in terms of the transmission rates, packet sizes, and traffic loads. We have derived the theoretically ideal contention status considering node heterogeneity, and developed a CW control scheme that achieves the ideal contention status in an average sense. Simulation results have verified that the developed performance analysis model accurately estimates the network status in terms of throughput, collision time, and collision probability considering node heterogeneity, and that the proposed CW control scheme achieves considerable performance improvement over the existing schemes which do not consider node heterogeneity. Lastly, the feasibility of the proposed CW control has been verified with real testbed experiments. For future research directions, the proposed performance analysis model and CW adaption could be extended to incorporate multi-user multiple-input multiple-output (MU-MIMO) defined in the IEEE 802.11ac.

In part II of this dissertation, we have developed a sounding control scheme for IEEE 802.11ac MU-MIMO. We have analytically formulated a sounding node set and interval determination problem in which the expected MU-MIMO gain is maximized considering sounding overhead. We have collected real channel data in practi-

cal WLAN environments, and employed the collected data for performance evaluation of the proposed scheme. Simulation results have revealed that determination of the sounding interval and sounding node set requires comprehensive consideration of the network environment including DL traffic loads, SNRs, and channel coherence times. The proposed scheme achieves remarkable performance improvement over the existing scheme which considers channel coherence times only.

In part III of this dissertation, we have developed an adaptive GID control scheme for IEEE 802.11ac WLANs. A novel GID assignment mechanism has been developed, and an idle power consumption minimization problem has been analytically formulated. We have also developed an efficient algorithm that finds the solution in a polynomial time. Simulation results demonstrate that the proposed GID control considerably reduces idle power consumption compared with random GID overloading.

Bibliography

[1] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[2] J. W. Robinson and T. S. Randhawa, "Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 917–928, June 2004.

[3] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions," *IEEE/ACM Trans. Networking*, vol. 15, no. 1, pp. 159–172, Feb. 2007.

[4] G. Prakash and P. Thangaraj, "Throughput Analysis of IEEE 802.11b WLAN under a Non-Saturated Condition," in *Proc. Int. Conf. and Workshop on Emerging Trends in Technology*, New York, NY, USA, Mar. 2010, pp. 298–303.

[5] D.-Y. Yang, T.-J. Lee, K. Jang, J.-B. Chang, and S. Choi, "Performance Enhancement of Multi-Rate IEEE 802.11 WLANs with Geographically-Scattered Stations," *IEEE Trans. Mobile Computing*, vol. 5, no. 7, pp. 906–919, July 2006.

[6] B. Kwak, N. Song, and L. E. Miller, "Performance Analysis of Exponential Backoff," *IEEE/ACM Trans. Networking*, vol. 13, no. 2, pp. 343–355, Apr. 2005.

[7] O. Tickoo and B. Sikdar, "Queuing Analysis and Delay Mitigation in IEEE 802.11 Random Access MAC Based Wireless Networks," in *Proc. IEEE INFOCOM*, Hong Kong, China, Mar. 2004, pp. 1404–1413.

[8] ——, “A Queueing Model for Finite Load IEEE 802.11 Random Access MAC,” in *Proc. IEEE ICC*, Sydney, Australia, June 2004, pp. 175–179.

[9] F. Daneshgaran, M. Laddomada, F. Mastiff, and M. Mondin, “Unsaturated Throughput Analysis of IEEE 802.11 in Presence of Non Ideal Transmission Channel and Capture Effects,” *IEEE Trans. Wireless Communications*, vol. 7, no. 4, pp. 1276–1286, Apr. 2008.

[10] K. Medepalli and F. A. Tobagi, “Towards Performance Modeling of IEEE 802.11 based Wireless Networks: A Unified Framework and its Applications,” in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–12.

[11] J. Hui and M. Devetsikiotis, “A Unified Model for the Performance Analysis of IEEE 802.11e EDCA,” *IEEE Trans. Communications*, vol. 53, no. 9, pp. 1498–1510, Sep. 2005.

[12] F. Eshghi and A. K. Elhakeems, “Performance Analysis of Ad hoc Wireless LANs for Real-Time Traffic,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 204–215, Feb. 2003.

[13] X. J. Dong and P. Varaiya, “Saturation Throughput Analysis of IEEE 802.11 Wireless LANs for Lossy Channel,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 100–102, Feb. 2005.

[14] Z. Hadzi-Velkov and B. Spasenovski, “Saturation Throughput - Delay Analysis of IEEE 802.11 DCF in Fading Channel,” in *Proc. IEEE ICC*, Anchorage, AK, May 2003, pp. 121–126.

[15] E. Garcia, D. Viamonte, R. Vidal, and J. Paradells, “Achievable Bandwidth Estimation for Stations in Multi-Rate IEEE 802.11 WLAN Cells,” in *Proc. IEEE WoWMoM*, Espoo, Finland, June 2007, pp. 1–8.

[16] G. Bianchi and I. Tinnirello, "Kalman Filter Estimation of the Number of Competing Terminals in an IEEE 802.11 Networks," in *Proc. INFOCOM*, San Francisco, CA, USA, Apr. 2003, pp. 844–852.

[17] J. S. Kim, E. Serpedin, and D. R. Shin, "Improved Particle Filtering Based Estimation of the Number of Competing Stations in IEEE 802.11 Networks," *IEEE Signal Process. Lett.*, vol. 15, pp. 87–90, 2008.

[18] D. Deng, C. Ke, H. Chen, and Y. Huang, "Contention Window Optimization for IEEE 802.11 DCF Access Control," *IEEE Trans. Wireless Communications*, vol. 7, no. 12, pp. 5129–5135, Dec. 2008.

[19] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, pp. 1774–1786, Sept. 2000.

[20] ——, "Dynamic tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Trans. Networking*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

[21] H. Lee, H. Lee, Y. Yi, S. Chong, B. Nardelli, and M. Chiang, "Making 802.11 DCF Near-Optimal: Design, Implementation, and Evaluation," in *Proc. IEEE SECON*, New Orleans, LA, June 2013, pp. 86–94.

[22] J. Ni, B. Tan, and R. Srikant, "Q-CSMA: Queue-Length Based CSMA/CA Algorithms for Achieving Maximum Throughput and Low Delay in Wireless Networks," *IEEE/ACM Trans. Networking*, vol. 20, no. 3, pp. 825–836, June 2012.

[23] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification: Medium Access Control (MAC) Quality of Service Enhancements*, IEEE Std. 802.11e, 2005.

[24] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification: Enhancements for Very High Throughput for Operation in Bands Below 6 GHz*, IEEE Std. 802.11ac, 2013.

[25] L. Choi and R. Murch, "A Transmit Preprocessing Technique for Multiuser MIMO Systems Using a Decomposition Approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, Jan. 2004.

[26] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates with Downlink Training and Channel State," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[27] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[28] N. A. E. Aryafar, T. Salonidis, and E. W. Knightly, "Design and Experimental Evaluation of Multi-User BeamForming in Wireless LANs," in *Proc. ACM MobiCom*, 2010, pp. 197–208.

[29] B. Bellalta, J. Barcelo, D. Staehle, and A. Vindel, "On the Performance of Packet Aggregation in IEEE 802.11ac MU-MIMO WLANs," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1588–1591, Oct. 2012.

[30] D. Nojima, L. Lanante, Y. Nagao, and M. Kurosaki, "Performance Evaluation for Multi-User MIMO IEEE 802.11ac Wireless LAN Systems," in *Proc. International Conference on Advanced Communication Technology (ICACT)*, 2012, pp. 804–808.

[31] G. Redieteab, L. Cariou, P. Christin, and J. F. Helard, "SU/MU-MIMO in IEEE 802.11ac: PHY+MAC Performance Comparison for Single Antenna Stations," in *Proc. IEEE Wireless Telecommunications Symposium (WTS)*, 2012, pp. 1–5.

[32] O. Bejarano, E. Magistretti, O. Gurewitz, and E. W. Knightly, "MUTE: Sounding Inhibition for MU-MIMO WLANs," in *Proc. IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2014.

[33] X. Zhang and K. G. Shin, "E-MiLi: Energy-Minimizing Idle Listening in Wireless Networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 9, pp. 1441–1454, 2012.

[34] S. P. Abraham, S. Vermani, and H. Sampath, "Efficient Group Definition and Overloading for Multiuser MIMO Transmissions," Patent US 13 028 016, Feb., 2011.

[35] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 2012.

[36] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification: Enhancements for Higher Throughput*, IEEE Std. 802.11n, 2009.

[37] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool Release: Gathering 802.11n Traces with Channel State Information," *ACM SIGCOMM CCR*, vol. 41, no. 1, p. 53, Jan. 2011.

[38] HostAP: IEEE 802.11 AP, IEEE 802.1X/WPA/WPA2/EAP/RADIUS Authenticator. [Online]. Available: http://hostap.epitest.fi/hostapd/

[39] Iperf: TCP/UDP Bandwidth Measurement Tool. [Online]. Available: http://dast.nlanr.net/Projects/iperf/

[40] J. Zhang, M. Kountouris, J. G. Andrews, and R. W. Heath, "Multi-Mode Transmission for the MIMO Broadcast Channel with Imperfect Channel State Information," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 803–814, Mar. 2011.

[41] Y. Lim, C. Chae, and G. Caire, “Performance Analysis of Massive MIMO for Cell-Boundary Users,” *IEEE Trans. Wireless Commun.*, in Press, Available at: http://arxiv:1309.7817v2[cs.IT].

초 록

본 학위 논문에서는 IEEE 802.11 무선 랜 매체접근제어 (MAC) 효율화 기법을 제안한다. 전반부에서는 단말들이 서로 다른 트래픽 양, 전송 속도, 패킷 크기를 가지는, 즉 단말 상이성이 있는 실제적인 무선 랜 환경을 위한 contention window 제어 기법을 제안한다. 각 단말이 특정 시점에 무선 매체에 접근하기 위하여 contention에 참여할 확률을 의미하는 activity probability 개념을 도입하여, 단말 상이성이 있는 무선 랜 네트워크 환경에서 패킷 충돌 확률, 패킷 충돌 시간, 백오프 시간, 단위 시간당 처리량 등을 예상할 수 있는 성능 분석 모델을 제안한다. 새롭게 제안된 성능 분석 모델을 기반으로 하여, 이론적 관점에서 이상적인 contention 상태를 유도하고, 무선 랜 네트워크를 평균적인 관점에서 이상적인 contention 상태에서 동작하도록 하는 contention window 제어 기법을 제안한다. NS-3 시뮬레이션과 실측 실험을 진행하여, 새롭게 제안한 성능 분석 모델이 단말 상이성을 갖는 무선 랜 네트워크 환경의 contention 상태를 상당한 정확도로 예측하는 것을 확인할 뿐만 아니라, 제안한 contention window 제어 기법이 단말 상이성을 고려하지 않는 기존의 제어 기법과 비교하여, 상당한 성능 향상을 얻는 것을 확인한다.

본 학위 논문의 중반부에서는 IEEE 802.11ac 다중 사용자 다중 안테나 전송을 위한 채널 사운딩 제어 기법을 제안한다. 제안하는 사운딩 제어 기법은 다중 사용자 다중 안테나 전송으로부터 얻어지는 장기적 성능 향상의 기대치를 최대화하기 위하여, 단말들의 하향링크 트래픽 양과 AP로부터 단말들로의 무선 채널 coherence time을 고려하여, 사운딩에 참여할 단말과 사운딩을 수행할 주기를 결정한다. 이를 위하여, 무선 랜 네트워크 환경과 사운딩에 따르는 손실을 고려하여, 다중 사용자

다중 안테나 전송으로부터 얻어지는 성능 향상 최대화 문제를 이론적으로 유도한다. 실제 무선 랜에서 채널 환경을 측정하고, 측정된 채널 데이터를 기반으로 시뮬레이션을 진행하여, 제안한 기법이 무선 채널 coherence time만을 고려하는 기존의 제어 기법보다 좋은 성능을 나타내는 것을 확인한다.

본 학위 논문의 종반부에서는 IEEE 802.11ac 무선 랜 환경에서 대기 상태 파워 소모를 완화하는 그룹 아이디 제어 기법을 제안한다. 공통의 그룹 아이디와 위치를 공유하는 단말들의 대기 상태 파워 소모 기대치를 이론적으로 유도하여, 대기 상태 파워 소모가 단말들의 하향링크 트래픽 양과 관련이 있음을 보인다. 이를 바탕으로 하여, 대기 상태 파워 소모 최소화 문제를 유도하고, 계산 복잡도를 줄이기 위한 효율적인 그룹 아이디 제어 알고리즘을 함께 제안한다. 시뮬레이션을 통하여, AP에 많은 단말들이 접속되어 있는 환경에서 상당한 대기 상태 파워 소모가 발생하며, 제안한 그룹 아이디 제어 기법이 무작위로 그룹 아이디를 제어하는 것과 비교하여 대기 상태 파워 소모를 상당히 완화하는 것을 확인한다.

주요어: 무선 랜, 충돌 회피에 의한 반송파 감지 다중 액세스, 다중 사용자 다중 안테나 전송, 채널 사운딩

학번: 2010-20908