IEICE TRANSACTIONS

on Communications

VOL. E98-B NO. 8
AUGUST 2015

A PUBLICATION OF THE COMMUNICATIONS SOCIETY

The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Utility-Based Distributed Association Control Scheme with User Guidance for IEEE802.11 Wireless LANs**

Takahiro IWAMI[†*a)], **Irda ROSLAN**[†,††b)], ***Nonmembers***, **Yumi TAKAKI**[†c)], **Kyoko YAMORI**[†††d)], **Chikara OHTA**[†e)], ***Members***, ***and*** **Hisashi TAMAKI**[†f)], ***Nonmember***

SUMMARY At present, wireless local area networks (WLANs) based on IEEE802.11 are widely deployed in both private premises and public areas. In a public environment offering several access points (APs), a station (STA) needs to choose which AP to associate with. In this paper, we propose a distributed association control scheme with user guidance to increase users' utility based on uplink and downlink throughputs of individual stations. As part of the scheme, we also present a simple throughput estimation method that considers physical data rate, traffic demand, and frame length in both uplink and downlink. Basically, in the proposed scheme, an AP selects a user and suggests that the user moves to another AP if certain conditions are met. The user then decides whether to accept the suggestion or not in a self-interested manner or in a voluntary manner for the benefit of all users including the user's own self. Through simulations under this condition, we confirm that our distributed association control scheme can improve user utility and fairness even though the channel quality of the new AP is unknown in advance.

key words: *IEEE802.11, throughput estimation, association control, handover, user guidance, user utility*

1. Introduction

The emergence of the latest mobile technologies such as smart phones and tablet PCs indicates that users are demanding fast and easy access to information and other online services. This has led to the deployment of wireless local area networks (WLANs) not only in homes and private premises, but also in public places such as shopping malls, airport terminals and coffee shops. As the installation of the IEEE802.11 based WLANs in public area is effortless, multiple access points (APs) can be found anywhere and can be accessed by a station (STA). For an STA to associate with an AP, it usually chooses the highest rank AP on its AP-list or selects an AP that has the strongest received signal strength indicator (RSSI) which suggests better channel quality. Such association strategies, however, tend to overload the selected AP, which will deprive the joining STA and other existing STAs of sufficient throughput to satisfy their needs. A simple way to mitigate this situation is to associate the STA with a lightly loaded AP. Most of the association schemes proposed for IEEE802.11 WLANs e.g. [3], [7], [16], [20], [22], [23], [26] assume that greedy STAs always have frames to transmit, and/or simply try to increase aggregated throughput regardless of the user's satisfaction with the resulting throughput.

However, it is not the same for the case of STAs that do not always have frames to transmit, known as non-greedy STAs. According to the distributed coordination function (DCF) of IEEE802.11, only backlogged STAs which contend for channel access have almost equal opportunity of sending a frame. Therefore, an STA is not allowed to dominate the transmission opportunity and eventually occupy aggregated throughput more than the other STAs. To understand the situation, refer to Fig. 1. If a new STA (STA2) with small traffic demand associates with AP1 that is heavily loaded by STA1 due to its large traffic demand, STA2 has virtually at least the same transmission opportunity as STA1 thanks to the equal transmission opportunity of IEEE802.11 DCF. Hence, STA2 will have the same throughput as its traffic demand. However, this will affect STA1's user perception as its throughput is lowered by STA2's traffic. This phenomenon is discussed in Sect. 4. On the other hand, if STA2 associates with AP2 which is heavily loaded by multiple STAs with small traffic demand (STA3–5), all STAs, including STA2 may have the same chance of frame transmission and thus having similar throughput due to IEEE802.11 DCF.

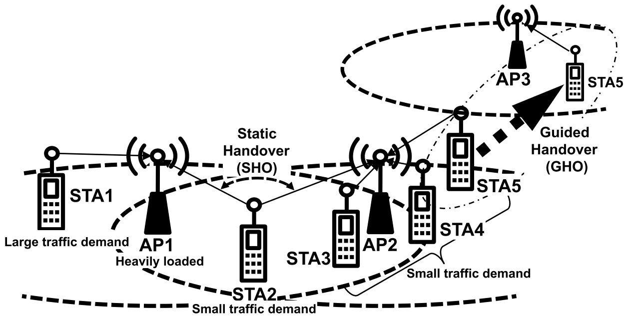


Fig. 1 Concept of the proposed distributed association control.

Manuscript received December 15, 2014.
Manuscript revised April 10, 2015.

[†]The authors are with the Graduate School of System Informatics, Kobe University, Kobe-shi, 657-8501 Japan.
[††]The author is also with the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia.
[†††]The author is with the School of Business Administration, Asahi University, Mizuho-shi, 501-0223 Japan.
[*]Presently, with the NTT DATA Corporation.

[**]This paper is an extended version of [9]–[11].

a) E-mail: iwami@al.cs.kobe-u.ac.jp
b) E-mail: irda@al.cs.kobe-u.ac.jp
c) E-mail: yumi@people.kobe-u.ac.jp
d) E-mail: yamori@alice.asahi-u.ac.jp
e) E-mail: ohta@port.kobe-u.ac.jp
f) E-mail: tamaki@al.cs.kobe-u.ac.jp

DOI: 10.1587/transcom.E98.B.1700

As a result, the STAs will have slightly smaller throughput than their traffic demand, which may affect their users' perception.

In [24], the authors agreed that the maximum aggregated throughput of an AP strongly depends on the frame length as well as the average traffic demand per STA. These are among of the factors that need to be considered in accomplishing a more sophisticated association control. Besides depending on the throughput attained, the degree of user satisfaction may also rely on the type of active application or the nature of the service contract (ex. premium or limited) consumed by the user. User satisfaction is indeed one of the metrics in choosing a better AP to be associated with. This can be achieved with user utility, which measures the levels of user satisfaction analogous to a set of characteristics offered by an access network [18]. Hence, user satisfaction can be modeled as a utility function.

There are a few methods that can be employed to determine a suitable AP for handover process. Most previous studies assume that the STA does not move during handover, which we call "static handover (SHO)." In this case, the STA simply compare the channel quality of adjacent APs based on the beacons transmitted by the APs. Another approach to select AP during handover is using spatial load balancing [13], [16]. To describe how spatial load balancing works, let us consider a situation where a large file needs to be downloaded via WLAN at an airport prior to departure. The nearest AP to the current location is very congested as there are many other passengers accessing WLAN as well. But if there is a nearby AP that can provide the required throughput, the user might be willing to walk there to receive the file in time. This handover approach where it proceeds at the will of the user is known as "guided handover (GHO)." With GHO, the user decides whether to accept the suggestion or ignore it in a self-interested manner. In Fig. 1, the user of STA5 is guided from the service area of AP2 to that of AP3, which is under-utilized. In general, the movement of a user will increase total user satisfaction at the old AP, with potential cost of degrading the total user satisfaction at the new AP. However, if the penalty is minor, it is likely to have an overall increase in user satisfaction and thus promotes social utility among all other users including the user's own self.

The difficulty with spatial load balancing is to predict the performance of the target STA with the newly associated AP in advance. The IEEE802.11a/b/g/n/ac standards allow multiple physical (PHY) data rates (hereafter, data rate, for simplicity), and the rates rise together with the signal to noise ratio (SNR). If the moving STA can utilize a higher data rate at the new AP, the negative impact on the existing STAs will become smaller since the frame transmit time is shorter, and vice versa. The work of [16] assumes that the moving STA knows the channel quality condition between itself and each AP within its moving distance. In practice, however, only the most conservative assumption is commonly adopted, which is using the lowest data rate. This assumption unfortunately reduces the actual benefits achieved by the STA's movement. Moreover, handover is less likely to be triggered due to the relatively small benefit.

In this paper, we propose a distributed association control scheme for IEEE802.11 WLANs, in which user diversity including traffic characteristics such as traffic demand and frame length as well as user characteristics such as user utility for throughput and movable distance are taken into account. A simple throughput prediction scheme is also provided as a part of the distributed association control scheme for uplink and downlink. Furthermore, we aim to clarify the effectiveness of GHO that works well without in advance knowledge such as the new channel quality and also the relationship between user cooperation and the social utility for WLAN users.

Note that this paper is an extended version of [9]–[11]. The simple simulations of [10], [11] examined the scenario of STAs on a square grid with the assumption that all throughput estimations were accurate. This paper verifies the effectiveness of the proposed scheme by applying Scenargie [31], a commercial network simulator, to a scenario where STAs are deployed in a field that mimics the departure lobby of Narita airport. Moreover, this paper describes our throughput estimation method in more detail than [9].

The rest of this paper is organized as follows: First, we describe related works in Sect. 2. In Sect. 3, we propose an association control scheme. The throughput estimation method used in the proposed association control scheme is then described in Sect. 4. In Sect. 5, simulation results are shown to investigate the effectiveness of SHO and GHO in terms of utility, fairness, and throughput. Finally, we conclude in Sect. 6.

2. Related Works

2.1 Association Control

In this section, we introduce related works on association control for IEEE802.11 WLANs.

In [7], a maximizing local throughput (MLT) scheme is proposed where each STA associates with the AP using the largest value of the metric defined by the frame error rate and the number of STAs. This scheme assumes that each STA is greedy to send frames and ignores frame collision. In practice, however, frame collision strongly affects the system throughput in IEEE802.11 WLAN [24]. A delay-based access point selection (DBS) scheme is proposed in [23]. In this scheme, the STA chooses the new AP according to the estimated packet delay (PD) to transmit a frame from the head-of-the-line of the transmission queue under the implicit assumption that all STAs are backlogged. Each STA tries to minimize its PD, so that it can increase its throughput, which results in load-balancing among APs. In [22], air time ratio (ATR), which is the ratio of channel busy time for transmission to overall time, is used to judge whether an AP is underutilized or not. In order to maximize aggregate throughput over APs while utilizing fewer APs, the ATR-based scheme hands over an STA currently connected to an overloaded AP to an underutilized adjacent AP, if any. Otherwise, handover is not executed. In this sense, the ATR-based scheme is not intended to achieve load-balancing.

A different approach is adopted by [26] who introduces a satisfaction ratio, the ratio of obtained throughput to the traffic demand. The paper considers only uplink throughput, with an assumption that frame-length is uniform and frame collision does not occur. Further, it is implicitly assumed that all STAs have the same data rate. In [20], mean opinion score (MOS), which represents user satisfaction based on perceived quality, is considered mainly for downlink video stream. In the MOS-based scheme, an AP periodically broadcasts beacon frames that carry the average value of MOSs. Each STA that associated with the AP estimates this MOS value by means of pseudo-subjective quality assessment (PSQA) based on metrics such as frame loss ratio and mean loss burst size. The target STA then chooses the AP with the highest average MOS among adjacent APs. This scheme, however, is only applicable to specific applications that permit service quality to be assessed, such as video with the MOS technique.

All the mentioned schemes above focused only on SHO approach. In this paper, we consider SHO as well as GHO. To the best of our knowledge, the concept of spatial load balancing was first introduced in [13], and its effectiveness was investigated from the viewpoint of call blocking probability in the context of cellular networks. In [16], an association control scheme that implements GHO in a centralized-manner for IEEE802.11 WLANs is proposed. However, it is assumed that STAs will always have backlogged frames and every STA has information on channel quality between itself and each AP within its moving distance. From these previous works, it can be seen that the effectiveness of GHO has not been studied in sufficient detail.

In addition, this paper takes account of variation in frame-length and traffic demand for both uplink and downlink among STAs.

2.2 IEEE802.11 DCF Analysis

In this section, we introduce related works on IEEE802.11 DCF analysis. So far, the case of multiple STAs has been considered, and some earlier works are summarized in [19].

In [2], Bianchi analyses a Markov chain that models IEEE802.11 DCF to compute saturated throughput for the case of a finite number of backlogged STAs. In [5], Bianchi's model is extended to an unsaturated environment where each STA with a buffer of one frame size is loaded with the same traffic demand. A similar case is considered in [4] except that each STA has an infinite buffer and Markovian analysis is applied. As for [15], Bianchi's model is extended to the case where each STA has a buffer of one frame but different traffic demand. As pointed in [19], it is necessary to solve $2n$-coupled non-linear equations numerically given n STAs. In [1], only two types of STAs with different traffic demands are analyzed using a Markov chain.

In [14], the authors provide an analysis based on a fixed-point equation that is simpler than the above works in order to estimate a collision probability and a transmission probability in a slot under the saturated condition. The analysis is then extended by [21], to the case of IEEE802.11e enhanced distributed channel access (EDCA). In [6] and [28], the case of homogeneous and unsaturated STAs is considered based on a fixed-point analysis. In the former, it is assumed that each STA has a buffer of one frame size, while in the latter, buffer size is equivalent to one or more frames. In [17], an IEEE802.11e EDCA WLAN with a mixture of saturated STAs and unsaturated STAs is considered. Each saturated STA has a finite buffer and different traffic demand. In this analysis, like [15], it is necessary to solve multiple simultaneous fixed-point equations numerically. If some unsaturated STAs are deemed to have become saturated, the simultaneous fixed-point equations need to be solved again.

In this paper, we extend Kumar's analysis [14] to the case of IEEE802.11 DCF WLAN with heterogeneous STAs that have buffers of finite capacity as shown in Sect. 4. Our analysis is simple and straightforward. A merit of the analysis is that it is only necessary to solve a fixed-point equation with respect to the number of nodes such as STAs and APs. Once a solution is obtained, it can be reused. In this analysis, we also consider downlink as well as uplink throughput for each STA-AP pair. Applying our methodology to IEEE802.11e EDCA is an issue for the future.

3. Distributed Association Control

In this section, we propose a distributed association scheme based on user utility. First of all, in Sect. 3.1, we describe the assumptions in this study. Proposed scheme is executed in a distributed manner analogous to that proposed in [3], which is directed at mobile cellular networks. In [3], distributed self-optimization is considered based on a Gibbs' sampler where the cost function, called "energy," is defined as the inverse of signal to interference plus noise ratio (SINR). In our case, the cost function is defined as the inverse of user utility as shown in Sect. 3.2. The proposed distributed association control aims to minimize the total energy according to the procedure of SHO and GHO as shown in Sect. 3.3. Notations used in this section are summarized in Table 1.

3.1 Assumptions

In this study, we assume the following.

1. Each AP chooses a non-overlapping channel so as not to interfere with each other.
2. Each AP can communicate with each other via wired networks.
3. Data rate is determined based on SNR. Each STA can estimate data rates for downlink between itself and each adjacent AP based on RSSI measurement of beacon frames periodically issued by APs. The data rate for uplink is the same as that for downlink because it is

Table 1 Notations and meanings for association control scheme.

Notation	Meaning
$\mathcal{A}$	A set of APs
$\mathcal{T}_i$	A set of STAs associated with AP i
U_u	User utility of STA u
ϵ_{total}	Total energy
ϵ	Partial energy
$d_{\max,u}$	Maximum movable distance of STA u
Δ	Control interval
t_{pro}	Protection time for SHO and GHO

supposed that there is not so much difference between both of them.

4. Each STA periodically notifies traffic demands for uplink and downlink, mean physical layer service data unit (PSDU) size of frames for uplink and downlink, utility function for uplink and downlink, maximum movable distance, vicinal APs' identifiers, and estimated uplink and downlink data rates for each adjacent AP.

3.2 Energy Function

In general, the user's utility will get larger as the user's satisfaction increase. We define the "energy" of a user as the reciprocal of the user's utility.

We introduce a total energy function ϵ_{total} as a measure of APs' state as follows:

$$\epsilon_{\text{total}} = \sum_{i \in \mathcal{A}} \sum_{u \in \mathcal{T}_i} \frac{1}{U_u}, \tag{1}$$

where U_u represents user u's utility when the STAs denoted as the set $\mathcal{T}_i$ are associated with AP i, and $0 \leq U_u \leq 1$. In this paper, user utility is expressed in terms of a non-decreasing function of estimated throughputs of uplink and downlink, and when its value reaches maximum, i.e. one, the throughputs are equal to traffic demands in both uplink and downlink. The estimation of uplink and downlink throughputs is shown in Sect. 4. Reducing the total energy leads to increasing users' utility as a whole. However, user's energy increases sharply (i.e. inversely) as the utility decreases. Thus, even if there are a few users who have lower utility than the others, the total energy is still large. To overcome this problem, the total energy function needs to be minimized in order to increase users' utility and fairness altogether as shown in Sect. 3.3.

According to Assumption 1, handover of the target STA from AP i to AP j does not affect user throughput in the other APs. Therefore, it is only necessary for AP i to communicate with just AP j in order to comprehend the influence of handover on the network. This enables distributed user association control. For convenience, we define the partial energy function $\epsilon(\cdot)$ as follows:

$$\epsilon(\mathcal{T}_i, \mathcal{T}_j) = \sum_{k \in \mathcal{T}_i} \frac{1}{U_k} + \sum_{l \in \mathcal{T}_j} \frac{1}{U_l}. \tag{2}$$

The first and the second terms are basically calculated by AP i and j, respectively.

3.3 SHO and GHO Operation

An AP, say AP i, selects a candidate STA, say STA u, for SHO or GHO among its associated STAs randomly every Δ seconds. AP i then chooses a candidate AP, say AP j, for handover within the maximum movable distance, $d_{\max,u}$, of the selected STA u. If STA u is within the service area of AP j, then SHO is triggered, otherwise GHO. Next, AP i and j cooperate to estimate the partial energies before and after handover as shown below:

$$\epsilon = \epsilon(\mathcal{T}_i, \mathcal{T}_j), \qquad \epsilon' = \epsilon(\mathcal{T}_i \setminus \{u\}, \mathcal{T}_j \cup \{u\}). \tag{3}$$

Calculation of ϵ' requires the uplink and downlink data rate between STA u and AP j. In the case of SHO, AP j can acquire them from AP i on the basis of Assumptions 3 and 4. On the other hand, in the case of GHO, AP j assumes that STA u uses the lowest data rate since STA u is out of AP j.

If $\epsilon \leq \epsilon'$, which means the state of AP i and j get worse or does not change, then the process is terminated. Otherwise, the procedure proceeds. In the case of SHO, AP i instructs STA u to perform SHO to AP j. On the other hand, in the case of GHO, further consideration needs to be performed. As mentioned in Sect. 1, GHO can be induced in a self-interested manner or in a voluntary manner for social benefit, a situation where appropriate advantages are received by all the users in the network including the user's own self. In the former case, it is natural to make the suggestion only if STA u will increase its utility. For this case, information for decision-making, e.g. the distance to the new AP candidate and the degree of utility improvement, need to be provided. Then, based on the information, the user of STA u decides whether to accept the suggestion or not, and inform the result to AP i. If the user accepts the suggestion, the user moves toward AP j. Otherwise, the user stays at the same place. In the latter case, depending on the user, the suggestion could be made even if STA u will not increase its utility. In Sect. 5, we will consider three types of GHO.

Note that any STA that is about to undergo SHO or to be suggested for GHO will not be selected during t_{pro} period for protection in order to prevent users from experiencing excessively frequent interruption of communications due to handover or movement.

4. Throughput Estimation

As mentioned in the previous section, we assume that user utility is determined based on the estimated throughput of IEEE802.11 DCF. For details of the behavior of IEEE802.11 DCF, see its specification [30].

In this section, we propose a throughput estimation scheme, which is applicable even in the case where uplink and downlink traffic demand and frame-length are heterogeneous among STAs. Additional notations used in this sec-

Table 2 Notations and meanings for throughput estimation.

Notation	Meaning
$\mathcal{T}$	A set of STAs (except AP).
$l_{\mathrm{u},u}$	Uplink mean PSDU size of STA u.
$l_{\mathrm{d},u}$	Downlink mean PSDU size of STA u.
$\mu_{\mathrm{u},u}$	Uplink data rate from STA u to AP.
$\mu_{\mathrm{d},u}$	Downlink data rate from AP to STA u.
$\lambda_{\mathrm{u},u}$	Uplink traffic demand from STA u to AP.
$\lambda_{\mathrm{d},u}$	Downlink traffic demand from AP to STA u.
$\hat{\lambda}_{\mathrm{u},u}$	Uplink throughput from STA u to AP.
$\hat{\lambda}_{\mathrm{d},u}$	Downlink throughput from AP to STA u.
$N_{\mathrm{u},u,\mathrm{DBPS}}$	The number of data bits per OFDM symbol of data frame from STA u to AP.
$N_{\mathrm{d},u,\mathrm{DBPS}}$	The number of data bits per OFDM symbol of data frame from AP to STA u.
$N_{\mathrm{u},u,\mathrm{DBPS,ACK}}$	The number of data bits per OFDM symbol of ACK frame from STA u to AP.
$N_{\mathrm{d},u,\mathrm{DBPS,ACK}}$	The number of data bits per OFDM symbol of ACK frame from AP to STA u.
$T_{\mathrm{u},u,\mathrm{ACK}}$	ACK transmission time from STA u to AP.
$T_{\mathrm{d},u,\mathrm{ACK}}$	ACK transmission time from AP to STA u.
$T_{\mathrm{u},u,\mathrm{DATA}}$	Data frame transmission time from STA u to AP.
$T_{\mathrm{d},u,\mathrm{DATA}}$	Data frame transmission time from AP to STA u.
T_{slot}	A slot time.
T_{SIFS}	SIFS time.
T_{DIFS}	DIFS time.
T_{sym}	Transmission time of a symbol.
T_{P}	Transmission time of the physical preamble.
T_{PHY}	Transmission time of the PHY header.
T_{ext}	Signal extension time.
$L_{\mathrm{H\_DATA}}$	Total header size [bytes] (28 bytes).
L_{ACK}	ACK size [bytes] (14 bytes).
$\mathrm{CW}_{\mathrm{min}}$	Minimum contention window size.
$\mathrm{CW}_{\mathrm{max}}$	Maximum contention window size.
τ	Propagation delay.
ρ_{max}	Maximum channel utilization.

tion are summarized in Table 2.

4.1 Basic Idea

Suppose that an AP accommodates multiple STAs. We collectively call the STAs and the AP "nodes" for simplicity.

First of all, we suppose that all nodes have frames to send, i.e. they are backlogged. The backlogged nodes have probabilistically equal opportunity of sending a data frame according to IEEE802.11 DCF. Each node starts to decrease its backoff counter just after the channel is sensed idle during a DCF interframe space (DIFS). When the counter reaches zero, the node attempts to send the data frame, occupying the head-of-line (HOL) in its transmission queue. If the destined STA receives the data frame successfully, it replies with an acknowledgment (ACK) frame after a short interframe space (SIFS). The data frame, however, can collide others because the nodes operate in a distributed manner. In such cases, the node can retransmit the data frame if a predetermined retry limit is not reached. Otherwise, it discards the data frame.

We model the above behavior to a round-robin service for multiple queues with switch-over time. In this model, during a polling period, each node either succeeds in transmitting a data frame at the HOL of its transmission queue or discards it if the retry limit is reached. Consequently, after a polling period finishes, only a data frame is removed from the transmission queue of each node, which models the equal transmission opportunity in IEEE802.11 DCF. Aside from these, backoff events are also occur in a polling period.

A polling period consists of a successful transmission period, a collision period, and a backoff period. The successful transmission period represents a total of the average time for each node to succeed in transmitting at most one data in a polling period. The collision period on the other hand, includes several frame collisions. Note that according to IEEE802.11 DCF, every node runs the backoff period down in parallel. Otherwise, one or more nodes would transmit data frames, which contradicts the concept of the backoff period.

We assume that transmission buffers are managed on a per-frame basis (not octet-basis) and arriving frames from downlink-sessions are accommodated in a single transmission queue at the AP. If a queue has some unused capacity, frames from the sessions will be successfully stored in proportion to the frame arrival rate. Otherwise, they are dropped. As a result, AP transmission opportunities are also allocated to each session in proportion to the frame arrival rate.

Recall that Table 2 summarizes the notations.

4.2 Collision and Transmission Probabilities[14]

Collision probability and transmission probability are given by the fixed-point analysis provided in [14]. Here, an overview of the analysis is conducted.

Let $\mathcal{T}_{\mathrm{bl}}$ denote a set of nodes that have transmission frames, namely backlogged nodes, and let n denote the number of nodes in the set, i.e. $n = |\mathcal{T}_{\mathrm{bl}}|$. Let $\gamma(n)$ denote the probability that a transmitted data frame collides with others in the case of n backlogged nodes. Let $g(n)$ denote the probability that a node (re)transmits a data frame in the case of n backlogged nodes. Let K denote the maximum retry limit of frame transmission.

Let $S(n)$, $R(n)$, and $X(n)$ denote the average number of frame successfully transmitted, the average number of frame (re)transmissions, and the average number of backoff slots, respectively. The values of $S(n)$, $R(n)$, and $X(n)$ are given by

$$S(n) = \sum_{k=1}^{K+1} \gamma(n)^{k-1}(1-\gamma(n)) = 1 - \gamma(n)^{K+1}, \tag{4}$$

$$\begin{aligned} R(n) &= (K+1)\gamma(n)^{K+1} + \sum_{k=1}^{K+1} k\gamma(n)^{k-1}(1-\gamma(n)) \\ &= 1 + \gamma(n) + \cdots + \gamma(n)^{K}, \end{aligned} \tag{5}$$

$$\begin{aligned} X(n) &= b_K\gamma(n)^{K+1} + \sum_{k=1}^{K+1} b_{k-1} \sum_{i=0}^{k-1} b_i\gamma(n)^{k-1}(1-\gamma(n)) \\ &= b_0 + b_1\gamma(n) + \cdots + b_k\gamma(n)^{K}, \end{aligned} \tag{6}$$

where b_k, the average number of backoff slots in the $(k+1)$st (re)transmission trial, is given by

$$b_k = \frac{2^{\min\{k,m\}}(\mathrm{CW}_{\min}+1)-1}{2}, \tag{7}$$

where m is the number of consecutive collisions to reach the maximum contention window size, which satisfies $\mathrm{CW}_{\max} = 2^m(\mathrm{CW}_{\min}+1)-1$. Note that $S(n) \le 1$, and $1 - S(n) = \gamma(n)^{K+1}$ denotes the expected number of frames discarded due to successive $(K+1)$ times (re)transmissions.

Since a data frame is (re)transmitted $R(n)$ times during $X(n)$ slots, the transmission probability, $g(n)$, in a slot, per node, is given by

$$g(n) = R(n)/X(n). \tag{8}$$

Since every node has the same collision probability, $\gamma(n)$, it is given by

$$\gamma(n) = 1-(1-g(n))^{n-1}. \tag{9}$$

As proved in [14], Eqs. (8) and (9) have a unique fixed point in terms of $\gamma(n)$, and the value of $\gamma(n)$ is obtained by solving Eqs. (8) and (9), numerically. Given $\gamma(n)$, we can calculate the values of $S(n)$, $R(n)$, $X(n)$, and $g(n)$.

Note that the value of $\gamma(n)$ mainly depends on n, the number of backlogged nodes. Thus, for each n, it is easy to calculate them in advance and store in the APs.

4.3 Polling Time

In this section, we explain how to calculate the polling time, $\rho_{\mathrm{poll}}(\mathcal{T}_{\mathrm{bl}})$, which is the time required to complete one round of polling when the set of backlogged nodes is $\mathcal{T}_{\mathrm{bl}}$. Note that $l_{\mathrm{u},u}$, $l_{\mathrm{d},u}$, $\lambda_{\mathrm{u},u}$ and $\lambda_{\mathrm{d},u}$, which are defined in Table 2, include overhead due to upper layer headers. Table 3 summarizes data size and delay time in IEEE802.11a/g [25]. Physical data rate and the corresponding N_{DBPS}, number of data bits per OFDM (Orthogonal Frequency Division Multiplexing) symbol are shown in Table 4 [25].

Let $T_{v,\mathrm{DATA}}$ and $T_{v,\mathrm{ACK}}$ denote the time taken for node v to transmit a data frame and that for node v to receive an ACK frame, respectively. Here, without loss of generality, we can number each node as $T_{i,\mathrm{DATA}} \le T_{j,\mathrm{DATA}}$ for $1 \le i < j \le n = |\mathcal{T}_{\mathrm{bl}}|$. According to [25], if node v is identical to STA u, we have

$$\begin{aligned} T_{v,\mathrm{DATA}} &= T_{\mathrm{sym}} \cdot \left\lceil \frac{16+6+8L_{\mathrm{H\_DATA}}+8l_{\mathrm{u},u}}{N_{\mathrm{u},u,\mathrm{DBPS}}} \right\rceil \\ &\quad + T_{\mathrm{P}} + T_{\mathrm{PHY}} + T_{\mathrm{ext}}, \end{aligned} \tag{10}$$

$$\begin{aligned} T_{v,\mathrm{ACK}} &= T_{\mathrm{sym}} \cdot \left\lceil \frac{16+6+8L_{\mathrm{ACK}}}{N_{\mathrm{d},u,\mathrm{DBPS,ACK}}} \right\rceil \\ &\quad + T_{\mathrm{P}} + T_{\mathrm{PHY}} + T_{\mathrm{ext}}, \end{aligned} \tag{11}$$

where $\lceil x \rceil$ denotes the smallest integer more than or equal to x. Similarly, if node v is the AP, we have

$$T_{v,\mathrm{DATA}} = \sum_{u\in\mathcal{T}} \left(T_{\mathrm{d},u,\mathrm{DATA}} \frac{\frac{\lambda_{\mathrm{d},u}}{l_{\mathrm{d},u}}}{\sum_{w\in\mathcal{T}} \frac{\lambda_{\mathrm{d},w}}{l_{\mathrm{d},w}}} \right), \tag{12}$$

$$T_{v,\mathrm{ACK}} = \sum_{u\in\mathcal{T}} \left(T_{\mathrm{d},u,\mathrm{ACK}} \frac{\frac{\lambda_{\mathrm{d},u}}{l_{\mathrm{d},u}}}{\sum_{w\in\mathcal{T}} \frac{\lambda_{\mathrm{d},w}}{l_{\mathrm{d},w}}} \right), \tag{13}$$

where

$$\begin{aligned} T_{\mathrm{d},u,\mathrm{DATA}} &= T_{\mathrm{sym}} \cdot \left\lceil \frac{16+6+8L_{\mathrm{H\_DATA}}+8l_{\mathrm{d},u}}{N_{\mathrm{d},u,\mathrm{DBPS}}} \right\rceil \\ &\quad + T_{\mathrm{P}} + T_{\mathrm{PHY}} + T_{\mathrm{ext}}, \end{aligned} \tag{14}$$

$$\begin{aligned} T_{\mathrm{d},u,\mathrm{ACK}} &= T_{\mathrm{sym}} \cdot \left\lceil \frac{16+6+8L_{\mathrm{ACK}}}{N_{\mathrm{u},u,\mathrm{DBPS,ACK}}} \right\rceil \\ &\quad + T_{\mathrm{P}} + T_{\mathrm{PHY}} + T_{\mathrm{ext}}. \end{aligned} \tag{15}$$

In this case, the polling time $\rho_{\mathrm{poll}}(\mathcal{T}_{\mathrm{bl}})$ is given by

$$\rho_{\mathrm{poll}}(\mathcal{T}_{\mathrm{bl}}) = T_{\mathrm{suc}}(\mathcal{T}_{\mathrm{bl}}) + T_{\mathrm{col}}(\mathcal{T}_{\mathrm{bl}}) + T_{\mathrm{idle}}(\mathcal{T}_{\mathrm{bl}}), \tag{16}$$

where $T_{\mathrm{suc}}(\mathcal{T}_{\mathrm{bl}})$, $T_{\mathrm{col}}(\mathcal{T}_{\mathrm{bl}})$, and $T_{\mathrm{idle}}(\mathcal{T}_{\mathrm{bl}})$ denote a successful transmission period, a collision period, and a backoff period, respectively; they are given by

$$\begin{aligned} T_{\mathrm{suc}}(\mathcal{T}_{\mathrm{bl}}) &= S(n) \sum_{v\in\mathcal{T}_{\mathrm{bl}}} (T_{\mathrm{DIFS}} + T_{v,\mathrm{DATA}} + T_{\mathrm{SIFS}} \\ &\quad + T_{v,\mathrm{ACK}} + 2\tau), \end{aligned} \tag{17}$$

$$\begin{aligned} T_{\mathrm{col}}(\mathcal{T}_{\mathrm{bl}}) &= R(n) \sum_{r=2}^{n} g(n)^{r-1}(1-g(n))^{n-r} \\ &\quad \times \sum_{k=r}^{n} \binom{k-1}{r-1} (T_{\mathrm{DIFS}} + T_{k,\mathrm{DATA}} + \tau), \end{aligned} \tag{18}$$

$$T_{\mathrm{idle}}(\mathcal{T}_{\mathrm{bl}}) = X(n) T_{\mathrm{slot}}. \tag{19}$$

The derivation of Eq. (18) is explained in the Appendix. Equation (19) comes from the fact that all backlogged nodes count down the backoff period in parallel.

4.4 Estimation Algorithm

In this section, we introduce an algorithm to estimate the uplink and downlink throughputs of individual nodes. To simplify the procedure, we consider the number of frames arrive

Table 3 Time parameters in IEEE802.11a/g.

Notation	802.11a	802.11g	Notation	802.11a	802.11g
T_{slot}	9 μs	20 μs**	T_{PHY}	4 μs	48 μs
T_{SIFS}	16 μs	10 μs	τ	1 μs	1 μs
T_{DIFS}	34 μs	50 μs***	L_{ACK}	14 bytes	14 bytes
T_{sym}	4 μs	4 μs	$L_{\mathrm{H\_DATA}}$	64 bytes	64 bytes
T_{P}	16 μs	16 μs	T_{ext}	0 μs	6 μs

** 9μs, *** 28μs in case of short slot.

Table 4 Number of data bits per OFDM symbol.

Rate	N_{DBPS}	Rate	N_{DBPS}
6 Mbps	24 bits	24 Mbps	96 bits
9 Mbps	36 bits	36 Mbps	144 bits
12 Mbps	48 bits	48 Mbps	192 bits
18 Mbps	72 bits	54 Mbps	216 bits

at the queue of a node and the number of frames transmitted or discarded from the queue per unit time, e.g. one second. Suppose that frames generated in a previous unit time arrive at the queue at the beginning of the current unit time on a batch-wise basis, and then, all or some of them are transmitted during the unit time. If more frames have been stored than transmitted, the remaining frames will stay in the queue at the end of the unit time. Given that this case represents overloading, it is convenient to remove the remaining frames since this has no effect on the result. On the other hand, if a queue becomes empty during the current unit time, its node is no longer contends for channel access, so it can be removed from set $\mathcal{T}_{\mathrm{bl}}$. The number of frames transmitted in the unit time can be calculated by Algorithm 1; as outlined below:

1. Decide the number, Δf, of additional polling within the range for which node set $\mathcal{T}_{\mathrm{bl}}$ does not change (in the 4th line).
2. If the total channel utilization ρ exceeds the maximum channel utilization $\rho_{\max}$ during additional polling Δf, go to 6 (in the 10th line).
3. Otherwise, update the number, f_v, of successful transmissions for every node $v \in \mathcal{T}_{\mathrm{bl}}$ (in the 6th line).
4. Update the total polling frequency f (in the 7th line).
5. Update the total channel utilization ρ (in the 8th line).
6. Update the node set $\mathcal{T}_{\mathrm{bl}}$, and return to 1 (in the 9th line).
7. Allocate the number of polling events to each node in $\mathcal{T}_{\mathrm{bl}}$ up to the unused portion of the unit time, i.e., $\rho_{\max}$ minus the current ρ (in the 11th line), and update the number, f_v, of successful transmissions for every node $v \in \mathcal{T}_{\mathrm{bl}}$ (in the 12th line).

Here let $f_{\mathrm{u},u}$ and $f_{\mathrm{d},u}$ denote the number of successful transmissions of STA u per unit time in uplink and that in downlink, respectively. If node v is equivalent to STA u, then we have

$$f_{\mathrm{u},u} = f_v. \tag{20}$$

On the other hand, if node u is equivalent to the AP, then we have

$$f_{\mathrm{d},u} = \frac{\frac{\lambda_{\mathrm{d},u}}{l_{\mathrm{d},u}}}{\sum_{v \in \mathcal{T}} \frac{\lambda_{\mathrm{d},v}}{l_{\mathrm{d},v}}} f_{\mathrm{AP}}. \tag{21}$$

Hence, uplink and downlink throughputs in medium access control (MAC) layer are given by

$$\hat{\lambda}_{\mathrm{u},u} = f_{\mathrm{u},u} \cdot l_{\mathrm{u},u}, \tag{22}$$

$$\hat{\lambda}_{\mathrm{d},u} = f_{\mathrm{d},u} \cdot l_{\mathrm{d},u}. \tag{23}$$

Note that the above throughputs include headers in the upper layers.

Algorithm 1: Throughput estimation

1: $\mathcal{T}_{\mathrm{bl}} \leftarrow \mathcal{T} \cup \{\mathrm{AP}\}, \forall v \in \mathcal{T}_{\mathrm{bl}}; f_v \leftarrow 0$
2: $\rho \leftarrow 0, f \leftarrow 0, \Delta f \leftarrow 0$
3: **while** $n = |\mathcal{T}_{\mathrm{bl}}| > 0$ **do**
4: $\quad \Delta f \leftarrow \min_{v \in \mathcal{T}_{\mathrm{bl}}} \left(\frac{\lambda_v}{l_v}\right) - f$
5: $\quad$ **if** $\rho + \Delta f \cdot \rho_{\mathrm{poll}}(\mathcal{T}_{\mathrm{bl}}) < \rho_{\max}$ **then**
6: $\quad\quad f_v \leftarrow f_v + S(n)\Delta f, \forall v \in \mathcal{T}_{\mathrm{bl}}$
7: $\quad\quad f \leftarrow f + \Delta f$
8: $\quad\quad \rho \leftarrow \rho + \Delta f \cdot \rho_{\mathrm{poll}}(\mathcal{T}_{\mathrm{bl}})$
9: $\quad\quad \mathcal{T}_{\mathrm{bl}} \leftarrow \left\{v \in \mathcal{T}_{\mathrm{bl}} | \frac{\lambda_v}{l_v} > f\right\}$
10: $\quad$ **else**
11: $\quad\quad \Delta f \leftarrow (\rho_{\max} - \rho)/\rho_{\mathrm{poll}}(\mathcal{T}_{\mathrm{bl}})$
12: $\quad\quad f_v \leftarrow f_v + S(n)\Delta f, \forall v \in \mathcal{T}_{\mathrm{bl}}$
13: $\quad\quad$ break
14: $\quad$ **end if**
15: **end while**

4.5 Validation

We verify our throughput estimation method by comparing its outputs with simulation results. Scenargie is used as the network simulator.

Here, let us define the offered load in terms of messages in the application layer by ρ_{msg} as

$$\rho_{\mathrm{msg}} = \sum_{u \in \mathcal{T}} \left(\frac{\lambda^*_{\mathrm{u},u}}{\mu_{\mathrm{u},u}} + \frac{\lambda^*_{\mathrm{d},u}}{\mu_{\mathrm{d},u}} \right), \tag{24}$$

where $\lambda^*_{\mathrm{u},u}$ and $\lambda^*_{\mathrm{d},u}$ denote uplink and downlink traffic demands of STA u, respectively, in the application layer. Note that they do not include overhead due to Internet protocol (IP) headers, and user datagram protocol (UDP) headers unlike $\lambda_{\mathrm{u},u}$ and $\lambda_{\mathrm{d},u}$.

We explore the effect of frame length, traffic demand, and data rate on uplink and downlink throughputs of individual nodes. Table 5 summarizes the simulation scenario. In this scenario, 10 STAs generate constant bit rate (CBR) traffic, whose parameters are, for the case of $\rho_{\mathrm{msg}} = 0.1$, shown in Table 6. Note that PSDU size is the sum of message size, 20 bytes for IP header, and 8 bytes for UDP header. The dividing of the traffic rate by message size yields frame generation rate, which increases in the order of STA1, 2, 3, 4, 7, 8, 5, 9, 6, and 10. We varied the offered load, defined as Eq. (24), by changing the frame generation rate while keeping the proportions of individual traffic demand the same.

Figure 2 shows estimated and simulated uplink throughputs while the downlink equivalents are shown in Fig. 3. Only the results of seven STAs are drawn; plots for STA2, 5, and 7, are omitted as they overlap the others. In both estimated and simulated uplink throughputs, each STA peaks in descending order of frame generation rate regardless of data rate. Before reaching the peak, each STA has the same throughput as its traffic demand. After peaking, STAs with the same frame length have the same throughput asymptotically. Moreover, regardless of the data rate,

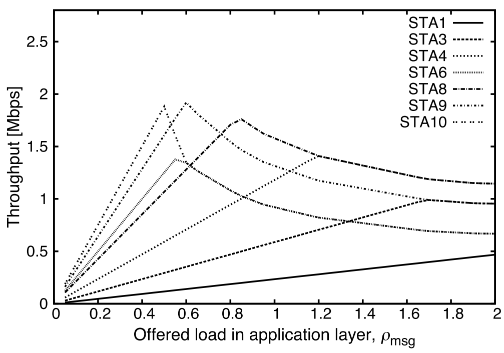
Table 5 Simulation settings for validation of throughput estimation.

Item	Value/Type
Simulator	Scenargie 1.6
Simulation Time [sec]	50
Simulation Number	10
Protocol	IEEE802.11g
Number of APs	1
Number of STAs	10
Traffic Model	CBR
Transport Layer	UDP

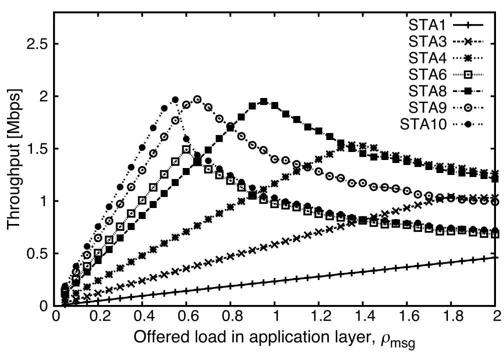
Table 6 Traffic parameters and data rate for validation of throughput estimation in the case of $\rho_{\mathrm{msg}} = 0.1$.

	Uplink			Downlink		
STA	Message size [bytes]	Traffic demand* [Mbps]	Data rate [Mbps]	Message size [bytes]	Traffic demand* [Mbps]	Data rate [Mbps]
1	1,200	0.024	12	700	0.036	12
2	700	0.036	12	1,000	0.084	12
3	1,000	0.060	12	1,200	0.096	12
4	1,200	0.120	24	700	0.120	24
5	700	0.216	24	1,000	0.096	24
6	700	0.252	36	1,200	0.252	36
7	1,000	0.144	36	700	0.108	36
8	1,200	0.216	54	1,000	0.270	54
9	1,000	0.324	54	1,200	0.216	54
10	700	0.378	54	700	0.216	54

* Traffic demand in application layer.

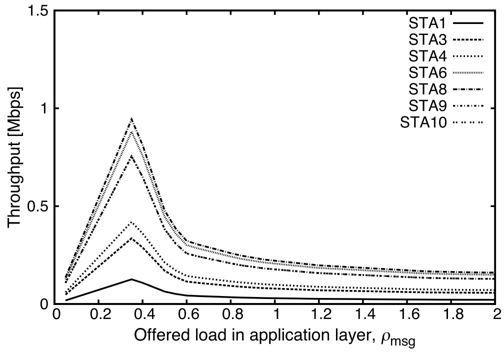


(a) Results by estimation

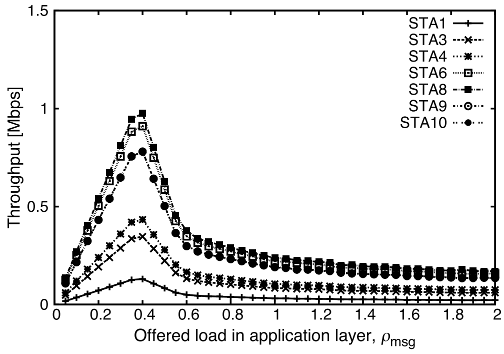


(b) Results by simulation

Fig. 2 Characteristics of individual uplink throughput.



(a) Results by estimation



(b) Results by simulation

Fig. 3 Characteristics of individual downlink throughput.

the longer the frame length is, the higher the asymptotic throughput is. STAs with smaller frame generation rates maintain the same throughput as the traffic demand even under heavier traffic loads. The above effects are due to the equal transmission opportunity of IEEE802.11 DCF. As for both estimated and simulated downlink throughputs, each STA peaks at the same traffic demand. Further, individual downlink throughput is proportional to individual downlink traffic demand, which is expected from Eqs. (21) and (23).

For both uplink and downlink, estimated and simulated throughputs basically agree; the worst error was about 0.36 Mbps. We confirmed that this error is due to the fact that our estimation scheme does not take account of extended inter-frame space (EIFS) and ACK timer for a shorter frame expires before that of a longer frame when different length frames collide. Eliminating this estimation error is a future challenge.

5. Performance Evaluation

As shown in the previous section, non-greedy STAs can have different throughput according to the traffic demand and frame length. Our association control scheme is applicable to such cases. In this section, we conduct simulations using Scenargie to investigate the effectiveness of SHO and GHO even when the moving STA is uncertain about the data rate at the new AP. We also investigate how much social benefit GHO provides for WLAN users. To do so, we consider the following four cases.

Case 1: SHO only
Case 2: SHO/GHO (WTM)
Case 3: SHO/GHO (lossless)
Case 4: SHO/GHO (sacrificial)

In Case 1, only SHO is executed. Whereas in Cases 2 through 4, GHO is executed along with SHO. It is expected that the total energy of the old and the new APs will decrease everytime GHO is conducted, which could enhance social utility. In Case 2, GHO is suggested if the user utility as well as social utility, is improved. The selected user expediently determines whether to accept the suggestion based on his or her "willingness to move (WTM)," which depends on the magnitude of improvement in user utility as described in Sect. 5.1. In Case 3, the selected user cooperatively accepts the suggestion even if the user utility increases slightly provided that it does not cause any loss, that is $U' > U$ where U and U' denote the current estimated utility and the expected utility after conducting GHO, respectively. In Case 4, unlike Cases 2 and 3, GHO can be triggered even if the selected user utility degrades. The selected user is assumed to accept the instruction for the benefit of society. In this sense, sacrificial cooperation is conducted.

5.1 User Model

Based on reference [18], we assume that user utility U_u of STA u is determined in terms of throughput as

$$U_u = \frac{u(\hat{\lambda}_{\mathrm{u},u}, \lambda_{\mathrm{u},u}) + u(\hat{\lambda}_{\mathrm{d},u}, \lambda_{\mathrm{d},u})}{2}, \tag{25}$$

where

$$u(\hat{\lambda}, \lambda) = \begin{cases} \frac{(2\hat{\lambda}/\lambda)^4}{1+(2\hat{\lambda}/\lambda)^4}, & 0 \leq \hat{\lambda} \leq \lambda/2, \\ 1 - \frac{\{2(\lambda-\hat{\lambda})/\lambda\}^4}{1+\{2(\lambda-\hat{\lambda})/\lambda\}^4}, & \lambda/2 < \hat{\lambda} \leq \lambda. \end{cases} \tag{26}$$

For convenience, we treat the impact of uplink and downlink throughput on user utility equivalently, i.e. we weight each 0.5. This selection, however, is not limited in practice.

For Case 2, based on [27], we assume that the selected user follows the suggestion of GHO if the distance to the indicated spot is shorter than the acceptable distance, d_{WTM}, which is determined in meters as

$$d_{\mathrm{WTM}} = 21.995 \ln\left(\frac{U'}{U}\right) + 91.11, \tag{27}$$

where U and U' denote the current estimated utility and the expected utility after conducting GHO, respectively. Otherwise, the suggestion is rejected.

5.2 Scenario

Table 7 summarizes the values of the parameters used in the simulations. Each STA has different message size and traffic demand as shown in Table 8. The reason of using variety of message sizes and traffic demands is to prove that our association control scheme works well in heterogeneous case. In Table 8, AP that is associated with each STA is shown.

Table 7 Major simulation settings for evaluation of SHO/GHO.

Item	Value/Type
Simulator	Scenargie 1.5
Number of simulation trials	10
Simulation time	3,000 s
MAC protocol	IEEE802.11g
Propagation model	Wall count
Penetration loss	10 dB
Number of APs	9
Number of STAs	40
Travel speed	1 m/s
Control interval Δ	30 s
Maximum movable distance†	300 m
Maximum channel utilization $\rho_{\max}$	1
Protection time t_{pro}	60 s

Figure 4 shows the simulation field which mimics the departure lobby (third floor) of Narita International Airport, the locations of APs, and the initial position of STAs. Since the actual locations of APs in the airport are not disclosed, we assume that APs (AP1–6) are located on Free Wi-Fi Desks [29]. Since these APs could not cover the entire area, we place three additional APs (AP7–9) to eliminate the dead spots. The service area of each AP is line-of-sight. At the beginning of the simulation, each STA associates with the AP with the strongest RSSI.

To evaluate the effects of SHO and GHO, we introduce two scenarios. In the first scenario which we will call as Scenario 1 hereafter, STAs are initially located so that AP2, AP6, and AP7 are overloaded while others are not. As for the second scenario which we will call as Scenario 2 hereafter, the location of the STAs has been changed so that the new overloaded AP will be AP4, AP6, and AP8.

For adjacent APs, an STA estimates the physical data rate between itself and each of the APs based on the SNR of beacon frames which are periodically issued by APs. Table 9 shows the rate adaptation table used in the simulations. For farther APs that are outside the STA's communication range, the STA assumes that only the minimum physical data rate, i.e. 6 Mbps, is available.

It is also assumed that users walk at one meter per second when they migrate from one AP to another. While moving, their STAs cannot communicate with any AP.

5.3 Performance Measures

Measures for evaluation are average user utility, average total throughput, and fairness of utility. As the metric of fairness, we use Jain's fairness index, F, which is given by

†How the maximum movable distance is used is explained in Sect. 3.3. The maximum movable distance is applied to Cases 2, 3 and 4. In addition, equation (27) is applied to Case 2 as well where a STA can deny to move if the suggested AP is far away.

Table 8 STA traffic parameters for evaluation of SHO/GHO and AP which each STA is initially associated with in Scenario 1 and 2.

STA	Uplink		Downlink		Initial AP	
	Message size [bytes]	Traffic demand* [Mbps]	Message size [bytes]	Traffic demand* [Mbps]	Scenario 1	Scenario 2
1	1,400	0.5	100	2	2	3
2	100	2.5	500	4	2	3
3	400	0.8	1,000	0.4	2	3
4	900	2.4	600	3.2	2	3
5	1,300	0.2	700	0.7	2	3
6	800	0.8	500	0.8	2	3
7	600	1	200	2	2	3
8	700	3.5	800	2.5	2	3
9	800	0.8	500	0.5	3	2
10	500	0.5	900	0.2	3	2
11	900	0.3	500	0.1	3	2
12	1,100	0.08	600	4	5	5
13	1,200	1.2	900	1.5	5	5
14	300	0.5	1,400	0.7	5	5
15	1,400	0.02	1,200	0.4	4	4
16	500	0.5	1,000	0.2	4	4
17	200	1	500	0.8	7	5
18	800	2	600	2.5	7	5
19	400	1.6	300	0.3	7	5
20	100	0.01	400	0.8	7	5
21	600	0.1	200	1.6	7	5
22	100	1	500	4	7	5
23	800	0.2	200	0.4	7	5
24	800	0.02	1,000	0.05	5	7
25	1,000	0.05	400	1	5	7
26	200	1	500	1.6	5	7
27	900	1.5	500	3.2	6	7
28	1,300	0.5	600	1	6	7
29	700	5	200	0.5	6	4
30	600	1	400	3.2	6	7
31	800	0.05	1,300	0.5	8	9
32	600	0.15	500	0.8	8	9
33	500	0.8	1,000	0.2	8	9
34	700	2	900	2	6	6
35	500	2.5	600	1.2	6	6
36	400	1.6	300	1	6	8
37	600	0.5	200	0.4	9	8
38	800	0.05	1,100	1	9	8
39	1,200	1.5	400	0.01	9	8
40	100	0.089	900	1.6	9	8

∗ Traffic demand in application layer.

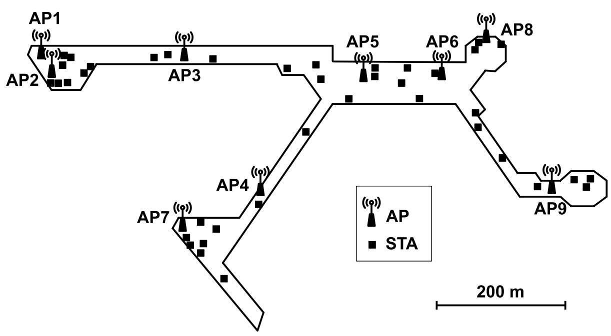
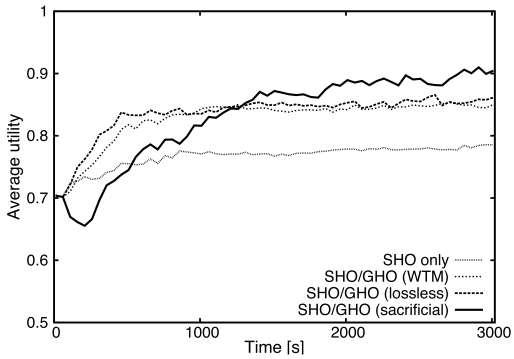


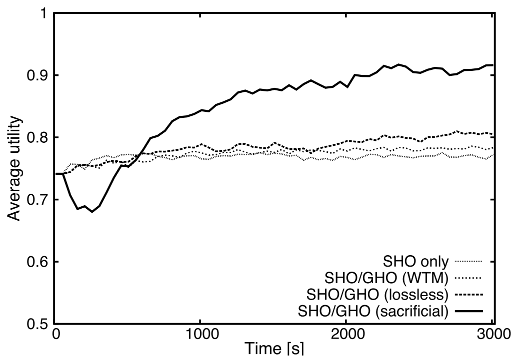
Fig. 4 Locations of nine APs and initial positions of 40 STAs in simulation field which mimics the departure lobby of Narita International Airport.

Table 9 Rate adaptation table in terms of SINR.

SNR [dB]	Data rate [Mbps]
40–∞	54
35–40	48
30–35	36
25–30	24
20–25	18
15–20	12
10–15	9
−∞–10	6



(a) Scenario 1



(b) Scenario 2

Fig. 5 Time variation of average user utility.

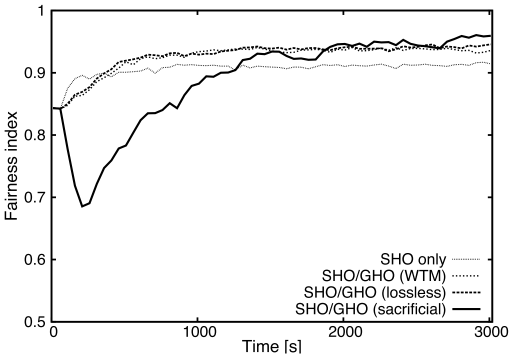
$$F = \frac{\left(\sum_{u=1}^{N_{\mathrm{STA}}} U_u\right)^2}{N_{\mathrm{STA}} \sum_{u=1}^{N_{\mathrm{STA}}} U_u^2}, \tag{28}$$

where N_{STA} is the number of STAs [12]. The closer this value to one, the fairer the user utility is.

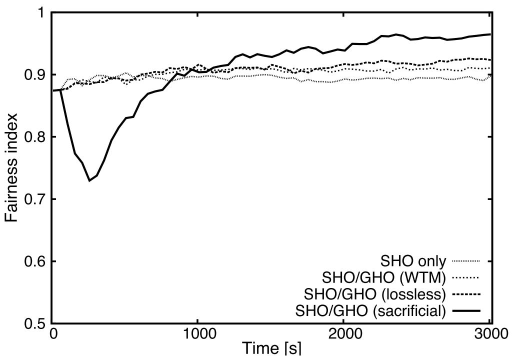
5.4 Simulation Results

Figures 5 through 7 show the time-varying characteristics of the average user utility, the fairness index, and the total throughput, over simulation trials for both scenarios, respectively.

For Scenario 1 as shown in Figs. 5(a), 6(a), and 7(a), we can confirm that the proposed distributed association schemes can eventually improve overall user utility, fairness, and throughput better than the initial values. In particular, GHO provides a significant improvement in these performance measurements even though the least data rate

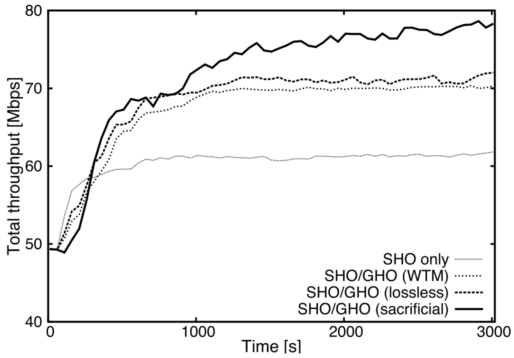


(a) Scenario 1

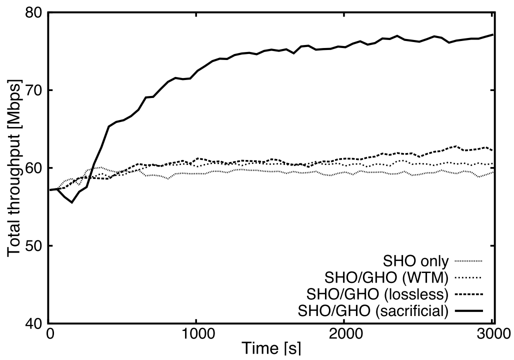


(b) Scenario 2

Fig. 6 Time variation in fairness index.



(a) Scenario 1



(b) Scenario 2

Fig. 7 Time variation in total throughput.

is assumed to be available in the new APs. For instance, let say the initial average utility is 0.7. The improvement for SHO only is just about 0.78. For the case of both SHO/GHO (WTM) and SHO/GHO (lossless), the average utility increased to about 0.85 and around 0.9 in the case of SHO/GHO (sacrificial). As for fairness index, there is no significant difference among the three GHO variations where the index results achieved are in the range of 0.9 to 0.95. The total throughput gained in Scenario 1 exhibits similar pattern of enhancement as in average utility result. Note that, in the case of SHO/GHO (sacrificial), the performance measures drop temporarily at the beginning of each trial, except for total throughput. This is because most GHOs are suggested at the beginning phase to balance the user utility. Besides that, STAs are in inactive state when users are moving. Such behavior does not exhibit in the total throughput measurement mainly because the static STAs still utilize bandwidth even if the moving STAs do not.

To verify the effectiveness of this scheme, we investigate Scenario 2 where the STA traffic parameters are maintained for each STA while being located at a different position from that in Scenario 1. The results are shown in Figs. 5(b), 6(b), and 7(b). As shown in Fig. 7(b), there is not much improvement achieved by SHO only, SHO/GHO (WTM), and SHO/GHO (lossless) cases compared with Scenario 1. This is because the initial throughput is 57 Mbps, which is higher than that in Scenario 1, and traffic load is more balanced over APs compared with Scenario 1. As discussed later, the standard deviation of traffic demand over AP for Scenario 2 is smaller than that for Scenario 1, which means that there is only small difference in the AP traffic load. This indicates that moving to the other AP will not improve user utility enough or will not at all. As a result, a selected user has less chance to improve user utility.

For the case of SHO/GHO (WTM), even if the selected user is guaranteed to have an improved user utility, the user may not move if the acceptable distance d_{WTM} becomes short due to small improvement. Let us consider a situation where an STA with 1 Mbps of traffic demand associated to an overloaded AP. If there is a nearby AP that is lightly loaded and can offer 2 Mbps throughput, the STA will be willingly to move to the lightly loaded AP when being suggested with GHO, while being guaranteed with improved user utility. Now, consider another situation where an STA with 10 Mbps of traffic demand associated to a heavily loaded AP. If it is suggested to handover to a far away AP which can offer 11 Mbps of throughput, the selected user will refuse to move far away just to gain only small improvement, that is the acceptable distance is short. In such a case, the selected user prefer to stay at the current AP. In this case, this situation will keep continue. One of the ways to overcome this problem is to make the selected user to handover to the other AP like SHO/GHO (sacrificial) case, so that the user utility can be improved. To accomplish this, we can give some incentive such as points or any kinds of token of appreciation to the users so that they will agree to move and satisfy with the gained user utility.

Table 10 The numbers of SHOs executed and GHOs accepted.

	Scenario 1		Scenario 2	
Association method	SHO	GHO	SHO	GHO
SHO only	22.9	N/A	22	N/A
SHO/GHO (WTM)	15.1	1.8	12.8	0.3
SHO/GHO (lossless)	17.6	2.8	13.4	1.3
SHO/GHO (sacrificial)	28.0	26.8	19.4	20.3

Table 11 Total traffic demands for each AP in Mbps.

AP	Scenario 1	Scenario 2
1	0	0
2	27.3	2.4
3	2.4	27.3
4	1.12	6.62
5	11.7	24.29
6	26.2	7.7
7	16.31	14.12
8	2.5	7.75
9	5.15	2.5

The SHO/GHO (lossless) case managed to achieve much better improvement compared to SHO/GHO (WTM). In this case, selected user will cooperatively handover to the other AP even though only slightly improvement can be achieved. Consider a situation where there is an STA with 5 Mbps traffic demand associated to an overloaded AP. If it is suggested to handover to a lightly overloaded AP that can offer 5.5 Mbps of throughput, the selected user will accept the suggestion and handover to that AP, provided that the action will not cause the user any loss.

Unlike SHO/GHO (WTM) and SHO/GHO (lossless), SHO/GHO (sacrificial) case takes only total energy into account. This means that if the total energy is improved and the GHO has been triggered, the selected user will agree to handover for the benefit of all users, even though the action will cause decrement to its user utility.

In Table 10, the number of SHOs executed and that of GHOs accepted are shown. For Scenario 1, we find that GHO suggestions are accepted only a few times in the case of SHO/GHO (WTM). The GHO is executed slightly more often in the case of SHO/GHO (lossless) but still not frequently. Nevertheless, they managed to improve the performance measures compared to SHO only. In contrast, SHO/GHO (sacrificial) yields about 10 times as many GHOs as SHO/GHO (lossless), which eventually achieves larger average user utility and larger total throughput. As for Scenario 2, even though the numbers of SHOs executed and GHOs accepted is fewer than the former scenario, it can be seen that the behavior shown by all cases are quite similar.

As mentioned earlier, the aim of this evaluation is to prove that our scheme works well in heterogeneous environment. We want to investigate which of these scenarios conform to this criteria. Table 11 shows the total traffic demands for each AP for both Scenario 1 and Scenario 2. Referring to

Table 12 Average and standard deviation of total traffic demand over AP.

Statistic	Scenario 1	Scenario 2
Mean	10.3	10.3
Standard Deviation	10.72	9.71

Table 12, even though the mean of total traffic demands for both scenario is same, the standard deviation of total traffic demand over AP for Scenario 1 is much larger than that of Scenario 2. Having high standard deviation means that all of the APs have different traffic demands. In this kind of scenario, the SHO/GHO (WTM) and SHO/GHO (lossless) schemes work efficiently.

6. Concluding Remarks

In this paper, we proposed a distributed association control scheme with user guidance to improve not only user utility but also its utility fairness among users based on uplink and downlink throughputs. As part of the scheme, we also provided a simple method to estimate throughput for non-greedy STAs, which captures the essential feature of the equal transmission opportunity provided by IEEE802.11 DCF. Our simulations confirmed that GHO improves user utility and the fairness index compared to the case of using only SHO. The result is quite satisfying even if the moving AP is unaware of the channel quality and solely depends on the users' decision to accept the suggestion of GHO in a self-interested manner. Considering two different scenarios gives us insight on the behavior of STAs for each cases where the initial position of each STA and the variability of traffic demands may deliver different results in measurements. Furthermore, we found that the improvement can be achieved by the movement of just a few users. When GHO is conducted in a sacrificial voluntary manner, more improvement is realized at the cost of moving many users.

In this study, we assumed that each AP chooses a non-overlapping channel so as not to interfere with each other. This situation can be carried out via the IEEE802.11k function [8]. In practice, however, it may not be possible if the APs are located in a highly dense network. In such a case, neighboring APs which utilize same channel probably need to set the maximum channel utilization $\rho_{\max}$ of Algorithm 1 adequately. Investigation of this is left as a future challenge.

Acknowledgment

This work was supported by the NICT (National Institute of Information and Communications Technology), Japan, as a part of "New generation network R&D program for innovative network vitalization platform and its application," and JSPS KAKENHI Grant Number 24500080.

References

[1] H.M.K. Alazemi, A. Margolis, J. Choi, R. Vijaykumar, and S. Roy,

"Stochastic modelling and analysis of 802.11 DCF with heterogeneous non-saturated nodes," Comput. Commun., vol.30, no.18, pp.3652–3661, Dec. 2007.
[2] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE J. Sel. Areas. Commun., vol.18, no.3, pp.535–547, March 2000.
[3] C.S. Chen and F. Baccelli, "Self-optimization in mobile cellular networks: Power control and user association," Proc. 2010 IEEE International Conference on Communications, pp.1–6, May 2010.
[4] L. Dai and X. Sun, "A unified analysis of IEEE 802.11 DCF networks: Stability, throughput, and delay," IEEE Trans. Mobile Comput., vol.12, no.8, pp.1558–1572, Aug. 2013.
[5] K. Duffy, D. Malone, and D.J. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," IEEE Commun. Lett., vol.9, no.8, pp.715–717, Aug. 2005.
[6] L. Fu, X. Wang, and Q. Zhang, "Unified fixed point analysis of IEEE 802.11(e) WLAN under saturated and unsaturated conditions," Proc. 2009 International Conference on Wireless Communications and Mobile Computing Connecting the World Wirelessly (IWCMC'09), pp.800–804, 2009.
[7] Y. Fukuda, A. Fujiwara, M. Tsuru, and Y. Oie, "Analysis of access point selection strategy in wireless lan LAN," Proc. VTC-2005-Fall, 2005 IEEE 62nd Vehicular Technology Conference, 2005, pp.2532–2536, Sept. 2005.
[8] V. Gazis, K. Sasloglou, N. Frangiadakis, P. Kikiras, A. Merentitis, K. Mathioudakis, and G. Mazarakis, "Cooperative communication in channel assignment strategies for IEEE 802.11k WLAN systems," Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp.1924–1929, Sept. 2013.
[9] T. Iwami, Y. Takaki, C. Ohta, and H. Tamaki, "A study on estimation of session-based throughput in IEEE802.11 wireless access network IEEE802.11," IEICE Technical Report, NS, vol.111, no.468, pp.493–497, March 2012. (in Japanese)
[10] T. Iwami, T. Inamoto, Y. Takaki, K. Yamori, C. Ohta, and H. Tamaki, "Performance evaluation of distributed user association based on user utility in wireless networks," IEICE Technical Report, vol.112, no.350, pp.103–108, Dec. 2012. (in Japanese)
[11] T. Iwami, Y. Takaki, K. Yamori, C. Ohta, and H. Tamaki, "Distributed association control considering user utility and user guidance in IEEE802.11 networks," Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp.2125–2130, Sept. 2013.
[12] R. Jain, D.M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC Research Report TR-301, Sept. 1984.
[13] S. Kaneda, Y. Akinaga, N. Shinagawa, and A. Miura, "Traffic control by influencing users' behavior in mobile networks," Proc. 19th International Teletraffic Congress, pp.583–592, Aug.–Sept. 2005.
[14] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs," IEEE/ACM Trans. Netw., vol.15, no.3, pp.588–601, June 2007.
[15] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions," IEEE/ACM Trans. Netw., vol.15, no.1, pp.159–172, Feb. 2007.
[16] S. Miyata, T. Murase, and K. Yamaoka, "Novel access-point selection for user QoS and system optimization based on user cooperative moving," IEICE Trans. Commun., vol.E95-B, no.6, pp.1953–1964, June 2012.
[17] S.H. Nguyen, H.L. Vu, and L.L.H. Andrew, "Performance analysis of IEEE 802.11 WLANs with saturated and unsaturated sources," IEEE Trans. Veh. Technol., vol.61, no.1, pp.333–345, Jan. 2012.
[18] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," Proc. NOMS 2008, 2008 IEEE Network Operations and Management Symposium, pp.144–151, April 2008.
[19] Q. Ni, D. Malone, P. Clifford, D. Duffy, D. Leith, and T. Li, "Modelling and simulation analysis of the 802.11/802.11e MAC layer," in Advances in Wireless Networks: Performance Modelling, Analysis and Enhancement, G. Ming, Y. Pan, and P. Fan, eds., pp.103–145, Nova Science Publishers, New York, 2008.
[20] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin, "QoE-based network selection for multimedia users in IEEE 802.11 wireless networks," Proc. 2008 33rd IEEE Conference on Local Computer Networks (LCN), pp.388–394, Oct. 2008.
[21] V. Ramaiyan, A. Kumar, and E. Altman, "Fixed point analysis of single cell IEEE 802.11e WLANs: Uniqueness and multistability," IEEE/ACM Trans. Netw., vol.16, no.5, pp.1080–1093, Oct. 2008.
[22] O. Shagdar, S. Tang, A. Hasegawa, T. Shibata, and S. Obana, "Association control for throughput maximization and energy efficiency for wireless LANs," Proc. IARIA EMERGING 2011, pp.112–117, Nov. 2011.
[23] V.A. Siris and D. Evaggelatou, "Access point selection for improving throughput fairness in wireless LANs," Proc. 2007 10th IFIP/IEEE International Symposium on Integrated Network Management, pp.469–477, May 2007.
[24] Y. Takaki, S. Kaneda, Y. Tanaka, C. Ohta, M. Takai, and Y. Okada, "Study on throughput performance of IEEE802.11 and LTE networks in multiple-terminal environment," IEICE Technical Report, MoMuC, vol.112, no.44, pp.129–134, May 2012. (in Japanese)
[25] Y. Xiao and J. Rosdahl, "Throughput and delay limits of IEEE 802.11," IEEE Commun. Lett., vol.6, no.8, pp.355–357, Aug. 2002.
[26] T. Yoshida and S. Ishihara, "An evaluation of load balancing of wireless LAN with network triggered handover by negotiations between neighboring access points," IPSJ SIG Note UBI, vol.2007, no.14, pp.101–108, Feb. 2007. (in Japanese)
[27] B. Yu, B. Gu, K. Yamori, and Y. Tanaka, "Relationship between user mobility and influential factors in wireless data communication services," IEICE General Conference 2013, pp.S80–S81, March 2013.
[28] Q. Zhao, D.H.K. Tsang, and T. Sakurai, "A simple and approximate model for nonsaturated IEEE 802.11 DCF," IEEE Trans. Mobile Comput., vol.8, no.11, pp.1539–1553, Nov. 2009.
[29] Free Wireless LAN Access Area, Terminal 1, 3rd Floor and Passport Control (Departures), http://www.narita-airport.jp/en/guide/service/list/map/mlan_t1_3f.html, accessed March 24, 2014.
[30] IEEE Std 802.11™-2012, IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, March 2012.
[31] Scenargie, http://www.spacetime-eng.com/jp/index.html, accessed March 24, 2014.

Appendix: Derivation of Collision Period

Let p_{col} denote the probability that data frames collide in a slot in the case of $n = |\mathcal{T}_{\mathrm{bl}}|$. Since this event occurs when at least two nodes transmit simultaneously, it is given by

$$\begin{aligned} p_{\mathrm{col}} &= \sum_{r=2}^{n} \binom{n}{r} g(n)^r (1-g(n))^{n-r} \\ &= 1 - \{1 + (n-1)g(n)\}(1-g(n))^{n-1}. \qquad (\mathrm{A}\cdot 1) \end{aligned}$$

Let t_{col} denote the average time of a DIFS and a time from when an STA starts to transmit a data frame until when it resumes channel sensing after collision occurs. When the node number is assigned such as $T_{i,\mathrm{DATA}} \le T_{j,\mathrm{DATA}}$ for $1 \le i < j \le n = |\mathcal{T}_{\mathrm{bl}}|$, we estimate the value of t_{col} as

$$t_{\mathrm{col}} = \frac{1}{p_{\mathrm{col}}} \sum_{r=2}^{n} g(n)^r (1-g(n))^{n-r} \sum_{1 \le k_1 < \cdots < k_r \le n} \{T_{\mathrm{DIFS}}$$

$$
\left. + \sum_{l=k_1,\ldots,k_r} \frac{\max\{T_{l,\mathrm{DATA}} + \delta_{\mathrm{ACK}}, T_{k_r,\mathrm{DATA}} + \tau\}}{r} \right\}
$$
$$
\approx \frac{1}{p_{\mathrm{col}}} \sum_{r=2}^{n} g(n)^r (1 - g(n))^{n-r}
$$
$$
\times \sum_{k=r}^{n} \binom{k-1}{r-1} (T_{\mathrm{DIFS}} + T_{k,\mathrm{DATA}} + \tau), \qquad (\mathrm{A}\cdot 2)
$$

where δ_{ACK} denotes an ACK timeout. In the above, we ignore the period that roughly corresponds to ACK timeout to simplify the calculation because some STAs with shorter frame transmission time can find that the channel is busy when their ACK timeout expires in the case of heterogeneous frame-length.

Let $p_{\mathrm{tag|col}}$ denote the conditional probability that a certain node, which we call "tagged node," transmits a data frame under the condition that collision occurs in a slot. Taking account of Eq. (9) and performing some algebra, we have

$$
p_{\mathrm{tag|col}} = \frac{g(n) \sum_{r=1}^{n-1} \binom{n-1}{r} g(n)^r (1 - g(n))^{n-1-r}}{p_{\mathrm{col}}}
$$
$$
= \frac{g(n)\gamma(n)}{p_{\mathrm{col}}}. \qquad (\mathrm{A}\cdot 3)
$$

Since each node encounters collisions $(R(n) - S(n))$ times on average in a polling period, collisions occur $(R(n) - S(n))/p_{\mathrm{tag|col}}$ times on average in a polling period regardless of the tagged node.

Therefore, from Eqs. (A·2) and (A·3), we have

$$
T_{\mathrm{col}}(\mathcal{T}_{\mathrm{bl}}) = \frac{R(n) - S(n)}{p_{\mathrm{tag|col}}} t_{\mathrm{col}}. \qquad (\mathrm{A}\cdot 4)
$$

By substituting Eqs. (4), (5), (A·2), and (A·3) into Eq. (A·4) and taking account of $(R(n) - S(n))/\gamma(n) = R(n)$, we finally have Eq. (18).

Takahiro Iwami received B.E. and M.E. degrees in engineering from Kobe University, Hyogo Japan, in 2012 and 2014, respectively. His interests include resource allocation in wireless networks. From April 2014, he joined NTT Data Corporation.

Irda Roslan received B.Sc. and M.Sc. degrees in Computer Science from Universiti Teknologi Malaysia (UTM), Malaysia in 2006 and 2008, respectively. She started her career as an IT lecturer in Cosmopoint College of Technology in Malacca, Malaysia in early 2008. She then joined the Department Of Computer Systems and Communication, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia in September 2008. She has been pursuing her Ph.D. in System Informatics at Kobe University since October 2012. Her research interests include computer systems and wireless networks.

Yumi Takaki received B.E. of science and technology from Saga University, Saga Japan, in 1989. From April 1989, she joined FUJITSU LIMITED. She then joined Hitachi Zosen Corporation. In 1999, she joined Kobe University, Hyogo Japan. She became an Assistant at the Graduate School of Engineering, Kobe University in 2007. Since April 2010, she has worked as an Assistant at the Graduate School of System Informatics, Kobe University. Her current interests include the research of wireless ad-hoc networks and network simulation techniques.

Kyoko Yamori received a B.A. degree in business administration in 1995, and M.A. and Ph.D. degrees in the graduate program in information management science from Asahi University, Gifu, Japan in 1997 and 2000, respectively. She is presently an associate professor at Department of Management Information, Asahi University. She is also a visiting associate professor at Global Information and Telecommunication Institute, Waseda University. She received the IEICE Switching System Research Award in 2001, the IEICE Young Researcher's Award in 2005, the IEICE Best Paper Award in 2005, the IEICE Information Network Research Award in 2006, the Ericsson Young Research Scientist Award in 2006, the IEICE Network System System Research Award in 2011, the Communication Quality Research Award in 2011, and the IEICE Communications Society Activity Award in 2011.

Chikara Ohta received his B.E., M.E. and Ph.D. (Engineering) degrees in Communication Engineering from Osaka University, Osaka Japan, in 1990, 1992 and 1995, respectively. From April 1995, he worked as an Assistant Professor at the Department of Computer Science at the Faculty of Engineering of Gunma University in Gunma, Japan. In October 1996, he joined the Department of Information Science and Intelligent Systems at the Faculty of Engineering, University of Tokushima, Tokushima, Japan, as a Lecturer. He became an Associate Professor at the same facility in March 2001. Since November 2002, he has worked as an Associate Professor of the Department of Computer and Systems Engineering in the Faculty of Engineering at Kobe University, Kobe, Japan. From March 2003 to February 2004, he was a visiting scholar at the University of Massachusetts at Amherst, USA. Since April 2010, he has worked as an Associate Professor of the Graduate School of System Informatics at same University, and he is a professor of the same graduate school from January 2015. His current interests include sensor networks, content delivery networks and resource control of wireless networks. He is a member of IPSJ, IEEE, and SIGCOMM.

Hisashi Tamaki received his B.E., M.E., and Ph.D. degrees, all in electrical engineering, from Kyoto University, Kyoto Japan in 1985, 1987 and 1993, respectively. From 1990, he was a Research Associate with the Department of Electrical Engineering, Kyoto University. Since 1995, he worked as a lecturer at the Department of Electrical and Electronics Engineering, Kobe University, Kobe Japan. He became an Associate Professor at the same faculty in 1999. From 2006, he worked as a Professor. Since 2010, he joined the Graduate School of System Informatics, Kobe University. He is a member of IEEE, IEE, SICE and ISCIE.