Association Control for Wireless LANs:
Pursuing Throughput Maximization and Energy Efficiency

Oyunchimeg Shagdar*, Suhua Tang†, Akio Hasegawa†, Tatsuuo Shibata†, Masayoshi Ohashi‡, and Sadao Obana‡

Email: oyunchimeg.shagdar@inria.fr, {shtang, ahase, shibata, ohashi}@atr.jp, and obana@cs.uec.ac.jp

* IMARA Project-Team, INRIA Rocquencourt
Domaine de Voluceau, B.P. 105 78153, Le Chesnay, FRANCE
† ATR Adaptive Communications Research Laboratories
Hikaridai, Keihanna Science City, Kyoto, JAPAN
‡ Graduate School of Informatics and Engineering, University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo, JAPAN

Abstract—Because the access points (APs) and the stations (STAs) of a community access network are deployed at the users’ desired places, the APs and STAs tend to concentrate in certain areas. A concentration of STAs often results in the AP(s) and STAs in that particular area suffering from severe congestion. A concentration of APs, on the other hand, may cause energy wastage. While a number of association control schemes are proposed to alleviate congestion in WLANs, the existing schemes do not necessarily maximize throughput and do not consider energy consumption. In this paper, we analytically formulate the network throughput as the multiplication of the success probability, frame transmission rate, and channel air-time ratio. The second and third components can easily be monitored and controlled based on measurements of local link and channel condition using the off-the-shelf WLAN devices. On the other hand, the first component, success probability is a function of the number of contending nodes that is extremely difficult to monitor in overlapping WLANs. Due to this reason, we extend our theoretical study and show that success probability can be indirectly maximized by controlling air-time ratio. Finally, we propose an association control scheme that aims at maximizing throughput and reducing energy consumption by taking account of the multiplication of frame transmission rate and air-time ratio. The proposed scheme is evaluated by computer simulations and testbed experiments conducted under real-world complex scenarios with UDP and TCP traffic. Both the simulations and actual implementations confirm the correctness of the theoretical work and the effectiveness of the proposed scheme.

Index Terms—association control; throughput maximization; success probability; air-time ratio; congestion alleviation; energy efficiency

I. INTRODUCTION

Due to the increasing popularity of WLAN technology, users (i.e., STAs) are often in the vicinity of one or more APs deployed at offices, school campuses, hotspot areas (e.g., cafés, train stations, airports), and individuals’ homes. Such a widespread deployment of WLAN triggered a launch of community access networks, including FON [2], which are built exploiting user-owned APs. As community access networks enable users to enjoy ubiquitous Internet access, it has the potential to play an important role in the future networking paradigm.

The APs of a community access network are deployed at the users’ desired places. Therefore, the fundamental difference of community networks from, e.g., enterprise wireless access networks is that a systematic placement of the APs is not possible. In community networks, APs are often concentrated in areas such as a residential street, and their distribution is highly non-uniform. Fig. 1 shows a FON map for an approximately 600m×600m area in Tokyo (near Akihabara station). Out of the total 14 APs installed in the overall area, 8 APs are concentrated in a small area in the upper left part of the map (approximately 1/8 of the overall area). The majority of these APs are in fact deployed in a condominium building, where residential homes, a café, conference spaces, and a fitness centre are. The remaining 6 APs are distributed in the rest 7/8 of the overall area as shown in the figure. STAs (i.e., users) are, on the other hand, expected to be concentrated in public places, such as cafés and train stations. Therefore, it is clear that STAs and APs are not necessarily concentrated
in a same area. A concentration of STAs often results in the AP(s) and STAs in that particular area suffering from severe congestion [3]-[4]. A concentration of APs, on the other hand, may cause energy wastage [5].

Congestion can be effectively alleviated by a proper association control. Indeed, a number of association control schemes are proposed for WLANs, mainly aiming at load balancing among cells under different definitions of load (e.g., load is defined as the number of nodes in [3], [6], channel condition in [7], and traffic rate in [8], [4]).

In [1], we analytically formulated the throughput maximization problem for overlapping WLANs, and showed that the existing schemes do not necessarily operate towards throughput maximization. The throughput can be expressed as the multiplication of three components: success probability, frame transmission rate, and air-time ratio (ATR). The success probability is the probability of collision-free transmission and it is a function of the number of contending nodes. The frame transmission rate is the speed of the data transmission, and it is a function of the link quality. ATR is the ratio of the channel busy time to the total time. While link quality and ATR can easily be monitored in WLAN systems, counting the number of contending nodes is extremely difficult because the nodes can be beyond the transmission range of each other. Thus in [1], we proposed an association control scheme that aims at throughput maximization by maximizing the multiplication of frame transmission rate and ATR. Success probability is indirectly maximized by controlling ATR. Moreover, because the existing association control schemes aim at load balancing among cells, they tend to use all the existing APs. However, especially in AP-concentrated areas, throughput maximization can be achieved by utilizing fewer APs. Such an association control provides a positive impact on energy efficiency since the unused APs can be in the power-saving mode. Indeed the proposed scheme does not aim at load balancing and it can reduce the number of active APs without hampering the network throughput.

This paper extends our work by a detailed theoretical study and extensive performance investigations. More specifically, the contributions of this paper are:

- An analytical study that explains why and how the success probability can be maximized by controlling air-time ratio.
- An extensive investigation of the proposed scheme through testbed evaluations targeting UDP and TCP traffic in complex real-world scenarios.

The remainder of the paper is organized as follows. Section II formulates the throughput maximization problem and introduces the related works. In section III, we provide a theoretical study on ATR and its relation with success probability. Section IV introduces the proposed association control scheme that aims at maximizing network throughput and reducing energy consumption. Sections V and VI evaluate the proposed scheme through computer simulations and testbed experiments. Finally, we conclude the paper in section VII.

II. PROBLEM FORMULATION AND RELATED WORK

The throughput of a node in a WLAN operating under DCF (Distributed Coordination Function) is expressed as follows [9], [10].

\[ s = \frac{P_t P_{tr} E[P]}{(1 - P_t) \sigma + P_t T} \]

(1)

Here, \( E[P] \) is the average data size, \( P_t \) is the probability that there is at least one node transmitting in the sensing range of the node, and \( P_{tr} \) is the probability of a successful transmission. \( \sigma \) is the slot time and \( T \) is the average time required for transmission of a data (includes DIFS and etc.). The numerator of (1) corresponds to the successfully transmitted payload length and the denominator is the total time. The former and the latter components of the denominator are the time that channel is idle and busy (either due to successful or unsuccessful transmissions), respectively.

The transmission of a data is successful if the frame is not collided and the frame does not contain errors (due to poor link quality). Thus, letting \( P_{sc} \) and \( P_{se} \) represent the probabilities of the former and the latter events, respectively, \( P_{sc} = P_{sc} \times P_{se} \). \( P_{sc} \) (in what follows we call \( P_{sc} \) as success probability) is expressed as

\[ P_{sc} = \frac{\tau (1 - \tau)^{n-1}}{1 - (1 - \tau)^n} \]

(2)

where \( n \) is the number of nodes, \( \tau \) is the channel access probability, which is determined by the contention window size (CW) and the probability that a node has a frame to transmit [10].

Letting \( r = P_{sc} \times E[P]/T \), we rewrite (1) as

\[ s = P_{sc} \times r \times \frac{P_{tr} T}{(1 - P_{tr}) \sigma + P_{tr} T} \]

(3)

\( E[P]/T \), the average frame transmission rate, is variable if rate adaptation is deployed at MAC (Medium Access Control protocol), and it is fixed otherwise. If rate adaptation is deployed, the better the link quality (stronger RSSI), the higher is the selected transmission rate. Furthermore, if the rate adaptation operates such that packet error rate (PER) is minimized (i.e., \( P_{sc} \) is maximized), the second component of (3), \( r \), depends mainly on the selected transmission rate, i.e., \( r \approx E[P]/T \). The last component of (3) is the ratio of time the channel is determined to be busy to the total time. Letting ATR (air-time ratio) represent the last component, the throughput of a node is the multiplication of \( P_{sc} \), \( r \), and ATR.

Finally, the throughput maximization problem for a network that consists of multiple overlapping WLANs is the maximization of

\[ S = \sum_{i=1}^{N} \sum_{j=1}^{n_i} s_{ij} \]

(4)

where \( N \) is the number of cells, \( n_i \) is the number of STAs at the \( i \)th cell, and \( s_{ij} \) is the throughput of the \( i \)th STA at the \( i \)th cell. It should be noted that the nodes that operate under the
same frequency channel and that are in each others’ sensing range share the same $P_{sc}$ and ATR regardless if they associated with a same or different APs.

In the traditional AP selection policy, a STA associates with the AP corresponding to the strongest RSSI. Thus, such a scheme takes account of only the second component of (3), $r$. However, increasing $r$ alone does not necessarily increase the total throughput, especially when STAs distribution is highly non-uniform. In such a scenario, it is possible that most of the STAs select a same AP (because they are closer to that AP). As it can be seen in (2), $P_{sc}$ decreases sharply with the increase of $n$ (because the numerator approaches zero and denominator approaches 1). Thus, in such a case, the throughput of the traditional scheme is poor because of a small $P_{sc}$ for the crowded cell(s) and likely a small ATR for the scarce neighboring cell(s). This suggests distributing STAs to the existing cells. Indeed several schemes are proposed to distribute the number of STAs among cells, and some of them take account of the link quality (RSSI in [3], and packet error rate in [6]). A drawback of these schemes is that they do not consider ATR, the channel availability.

Reference [7] proposed to balance effective channel busy-time (i.e., time the channel is busy for successful transmissions) among cells. Because it does not discriminate the time corresponding to unsuccessful transmissions and the idle time, [7] may suffer from such estimation errors. Moreover, because [7] ignores offered traffic volume, it might take a long time until load is balanced.

ATR is also the ratio of the amount of bandwidth consumed for transmissions to the total bandwidth. Thus, an increase of the offered traffic volume (injected traffic) increases ATR until it reaches its saturation value. Further increase of the offered traffic, however, results in congestion (i.e., buffer overflow) that significantly hampers communication performance. Since a cell with a small ATR can accommodate additional traffic, the overall throughput can be improved by moving STAs from the congested cell to such a non-congested cell. References [8], [4] proposed schemes that balance traffic volume among cells. A drawback of the schemes is that they do not consider the fact that the overlapping cells operating under the same frequency channel share the same channel resource. Moreover, [7], [8], [4] do not consider the link quality (the second component of (3)), and thus they may force STAs to use links with poor quality, degrading the users’ throughput. To the best of our knowledge, none of the existing schemes takes account of the overall of (3), thus they do not necessarily maximize throughput.

To this end, we propose an association control scheme that takes account of the overall of (3). The direct target of the proposed scheme is to maximize the sum of the multiplication of the second and third components of (3), $r$ and ATR, by taking account of RSSI, ATR, and the offered traffic volume. The success probability, $P_{sc}$, is indirectly maximized by maintaining ATR smaller than a threshold value. The next section provides an analytical study on why and how $P_{sc}$ can be maximized by controlling ATR. An important difference of the proposed scheme from the previous schemes is that because both the offered traffic and channel availability are estimated for each cell, the proposed scheme does not aim at load balancing. This enables the scheme to maximize throughput by utilizing fewer APs, providing a positive impact on energy efficiency.

III. STUDY ON AIR-TIME RATIO AND SUCCESS PROBABILITY

Equation (3) shows that the throughput of a node can be expressed by the multiplication of the three components: the success probability ($P_{sc}$), the average data transmission rate ($r$), and air-time ratio (ATR). The first and second components are further expressed as functions of the number of contending nodes ($n$) and link quality (RSSI), respectively. RSSI and ATR are local link and channel condition that can easily be monitored using the existing WLAN devices. On the other hand, the nodes that contribute $P_{sc}$ can be beyond the transmission range of each other (i.e., they cannot correctly receive the packets from each other). Therefore it is an extremely difficult task to count the number of such nodes.

This section will show that the success probability ($P_{sc}$) can be indirectly maximized by controlling the air-time ratio (ATR). Before getting into the details, we will first show that it is necessary to maximize $P_{sc}$ for throughput maximization.

As (2) shows, $P_{sc}$ is a function of the number of nodes and the channel access probability ($\tau$). $\tau$ is further expressed in [10]

$$\tau = q \times \frac{1 + \sum_{k=1}^{L_{\text{retry}}} (1 - P_{sc})^k}{CW_0[1 + \sum_{k=1}^{L_{\text{retry}}} 2^k (1 - P_{sc})^k]}$$

where $q$ is the probability of having a pending data frame (to transmit), $CW_0$ is the minimum contention window size, and $L_{\text{retry}}$ is the frame retransmission limit [11]. We now calculate the success probability (using (5) and (2)) for the IEEE 802.11a system [11]. Table I shows the parameters used in the calculation. Fig. 2 shows the success probability ($P_{sc}$)
for different numbers of nodes. While the smaller the q, the larger is the $P_{sc}$, it is clear that $P_{sc}$ sharply degrades with the increase of the number of nodes (note that the vertical axis is on a logarithmic scale).

We now express ATR by $P_{sc}$. $P_{tr}$, the probability that there is at least one station transmits (see (1)) is

$$P_{tr} = 1 - (1 - \tau)^n.$$  \(\text{(6)}\)

By substituting $P_{tr}$ to the formula of ATR (the 3rd component of (3)), the following relation of ATR and $P_{sc}$ can be achieved.

$$\text{ATR} = \frac{1}{P_{sc} \left( \frac{1 - \tau}{T_{\sigma}} + \frac{\sigma + \text{DIFS} + \text{SIFS}}{T_{\sigma}} + 1 \right)}.$$  \(\text{(7)}\)

Here $T'$ is the time required for the transmission of a data frame without including the inter-frame spaces (DIFS and SIFS). Because $T'$ is more convenient to monitor than $T$ (which includes the inter-frame spaces) in the real systems, we use $T'$ in this and the following sessions. For simplification of the mathematical calculations, we target the basic access method (i.e., the RTS/CTS (Request to Send and Clear to Send) handshake is not used). In such a case, $T'$ is the sum of the average time required for the transmissions of a data frame and the corresponding acknowledgement frame (ACK) (see Fig. 3).

Substituting (7) to (3), the throughput of a node is expressed slightly differently from (3):

$$s = \frac{\tau'}{r' \times \frac{1}{\frac{1 - \tau}{T_{\sigma}} + \frac{\sigma + \text{DIFS} + \text{SIFS}}{T_{\sigma}} + 1} \frac{1}{P_{sc}}}.$$  \(\text{(8)}\)

where $r'$=E[P]/T'.

Using (8), (5), and (2) we now calculate the throughput for the IEEE 802.11a system in the IEEE 802.11a, the acknowledgement frame (ACK) and the SIGNAL field of the PLCP header (the PLCP header fields excluding the SERVICE field) are generally coded with 1/2 BPSK (6 Mbps). The PSDU (physical layer service data unit) and the SERVICE field of the PLCP header in the data frame (the shaded part in Fig. 3) are subject to the MAC rate adaptation and thus we calculate the corresponding transmission time for different data rates. Fig. 4 shows the impact of the success probability, $P_{sc}$, on the throughput performance for different data rates. First of all, the figure shows that higher the utilized data rate (i.e., better link quality), the larger is the achieved throughout, implying the necessity of considering the link quality, i.e., the second component of (3). An important observation that can be made from the figure is that a sufficient throughput can be achieved only when $P_{sc}$ is large.

Now we will show the impact of $P_{sc}$ on ATR. In fact we can easily see in (7) that ATR takes on its maximum value, $\text{ATR}_{\text{max}}$, when $P_{sc}$ is 0. In this case, however, the channel is occupied with packet collisions, and obviously, the throughput is minimized (see Figure 4). On the other hand, if $P_{sc}$ is 1 (i.e., channel utilization is maximized), ATR takes on a value smaller than $\text{ATR}_{\text{max}}$, implying that the optimal value of ATR, $\text{ATR}_{\text{opt}}$ is below $\text{ATR}_{\text{max}}$ [12].

Now we will have a closer look at the relationship of ATR and $P_{sc}$. We substitute (5) to (7) and calculate ATR for the IEEE 802.11a system.
Fig. 5 shows the ATR characteristics with respect to the success probability, $P_{sc}$. Note that the solid (dashed) lines correspond to 500 bytes (1000 bytes) of data size. The figure clearly shows that ATR is maximized when $P_{sc}$ is minimized. The optimal value of ATR (that maximizes $P_{sc}$) is smaller than the maximum value of ATR. The figure also shows that the larger the data rate, the smaller is the ATR and the larger the data size, the larger is the ATR.

Fig. 5 shows the ATR characteristics with respect to $P_{sc}$ for different data sizes ($E[P]$) and data rates. The figure clearly shows that ATR is maximized when $P_{sc}$ is minimized; an increase of $P_{sc}$ results in a gradual decrease of ATR. Moreover, the larger the utilized data rate, the smaller is the ATR; the larger the data size, the larger is the ATR. The figure clearly shows that different optimal values exist for different data rate and data size. For example $ATR_{opt}$ is 0.57, 0.66, 0.88, and 0.92 for the data rate and data size combinations of (54 Mbps, 500 bytes), (54 Mbps, 1000 bytes), (6 Mbps, 500 bytes), and (6 Mbps, 1000 bytes), respectively. ($ATR_{opt}$ are shown with the round markers in the figure).

Our objective is to find a threshold value, $ATR_{th}$, which is used by our association control scheme to ensure $P_{sc}$ be sufficiently large. Because $ATR_{opt}$ takes on different values depending on the data size and the data rate, it is maybe possible to design a mechanism that finds the exact $ATR_{opt}$ and set $ATR_{th}$ to that value. However, designing such a mechanism can be very complex because it requires a precise estimation of the average data rate and the average data size. We believe that it is simple and thus realistic to set $ATR_{th}$ to a fixed value.

In WLANs, the users can be very close to the APs, i.e., the link quality (RSSI) can be good enough for a use of the highest data rate. Thus, it is safe to set $ATR_{th}$ to $ATR_{opt}$ corresponding to the highest data rate (i.e., 54 Mbps for the IEEE 802.11a). The data size, on the other hand, can be monitored or it can be set to a value, which is used by the typical applications. Thus according to Fig. 5, if the data size is 500 bytes, $ATR_{th}$ should be set to 0.57. In fact, in our previous work [1], we have empirically found that it is appropriate to set $ATR_{th}$ to 0.58 for WLANs for scenarios where data size takes on a value between 500 to 540 bytes. The analytical calculation achieved in Fig. 5 proves the correctness of the empirically-found threshold value.

As the success probability, $P_{sc}$, can be maximized by maintaining $ATR$ below the threshold value ($ATR_{th}$), the throughput maximization problem can be rewritten as

$$
\max \quad S = \sum_{i=1}^{N} \sum_{j=1}^{n_i} [r_{ij} \times ATR_{ij}] \\
\text{subject to } ATR_{ij} < ATR_{th} \quad \forall (i, j)
$$

where $N$ is the number of cells, $n_i$ is the number of STAs at the $i^{th}$ cell. $r_{ij}$ and $ATR_{ij}$ are the average frame transmission rate and ATR corresponds to the $j^{th}$ node at the $i^{th}$ cell.

IV. PROPOSED SCHEME

A user-owned AP is integrated into a community network by a common equipment/software program provided by the entity (e.g., FON [2]). Besides integrating users’ APs, the entity can play an important role for, e.g., network management for better communication quality. Hence, community networks are centrally controllable and we expect that the community members (i.e., the users) are cooperative to such a control. To this end, while a distributed mechanism can also be designed, in this paper, we propose a centralized association control scheme due to ease of management and better performance [7], [8], [5]. The proposed scheme aims at maximizing network throughput by a congestion alleviation mechanism and reducing energy consumption by a cell aggregation mechanism. For a large community network, the network can be divided into sub-networks, which are independently and separately controlled. Fig. 6 shows the network architecture.

Besides APs and STAs, the network has an information server (server) and control manager (manager). The server and manager can be physically separated or coexist in a same machine. APs and STAs periodically inform the server of information on the link quality, channel availability, and traffic volume. Periodically referring to the information, the manager triggers STAs’ handover for improved network throughput and/or energy efficiency.

A. Estimation of Channel Availability and Offered Traffic

STAs and APs measure RSSI, ATR, and the offered traffic volume, and inform the server of the information. The manager
uses the information to estimate channel availability and traffic condition for each cell, and changes STAs associations.

- Frame transmission rate

By periodically performing channel scanning, STAs are aware of the existence of the neighboring APs and the corresponding RSSI_{AP,STA}. Such background scanning is supported by the off-the-shelf wireless LAN cards [13], and some efforts have been made for fast channel scanning [14]. Transmission rate for frame payload field is estimated from RSSI_{AP,STA} and finally the frame transmission rate, r_{AP,STA} (=E[P]/T), for each pair of STA and AP is calculated.

- ATR

Because ATR value does not largely change over the space (e.g., it is around the same in the transmission range), only APs measure ATR on their operating channel. Such a measurement can easily be made using the existing WLAN cards [7],[5],[15]. We empirically found that the appropriate values for ATR_{th} are 0.58.

- Potential Throughput

The manager estimates the maximum achievable throughput for each pair of STA and its neighboring AP (which is not the STAs currently associated AP). Let PT_{STA,AP} (potential throughput) represent the maximum achievable throughput for such a STA and an AP. Equation (3) shows that the potential throughput is a function of P_{sc}, the estimated transmission rate (r_{AP,STA}), and ATR_{AP}. As discussed in the previous section, however, P_{sc} can be maximized by ensuring ATR below ATR_{th}. This enables PT to be estimated only from r_{AP,STA} and ATR_{AP}. Thus the manager calculates PT for each STA and its neighboring AP as

\[ PT_{STA,AP} = \begin{cases} 
0 & \text{ATR}_{AP} \geq ATR_{th} \\
\left(ATR_{th} - ATR_{AP}\right) \times r_{AP,STA} & \text{otherwise} 
\end{cases} \tag{10} \]

The upper equation is to not move the STA to the AP, at which ATR_{AP} exceeds ATR_{th}. Otherwise, the AP is a candidate destination AP for the STA, and the maximum achievable throughput at the candidate cell is expressed as the lower equation. In our previous work [15], we confirmed that such an estimation of PT can be achieved with a high accuracy using the existing WLAN cards.

- Offered Traffic

A STA can be moved to a neighboring AP, if it does not cause congestion at the neighboring cell. The condition can be verified by comparing the offered traffic volume for the STA and PT_{STA,AP} (to be discussed later). Letting EnqueueRate_{A,B} be the rate at which packets destined to node B are inserted into the IP queue at node A, the offered traffic volume for a STA is defined as

\[ \text{OfferedRate}_{STA} = \text{QueueRate}_{STA,AP} + \text{QueueRate}_{AP,STA} \tag{11} \]

If the WLAN is the bottleneck link of the end-to-end route, the OfferedRate is approximately equal to the traffic generation rate.

B. Congestion Alleviation

A cell is considered to be congested if

\[ ATR > ATR_{th}. \tag{12} \]

If the condition (12) is met for a cell, the manager checks if the aggregate offered rate exceeds the aggregate packet throughput for that cell, i.e.,

\[ \alpha \sum_{STAs} \text{OfferedRate}_{STA} > \sum_{STAs} \text{PacketThroughput}_{STA} \tag{13} \]

Here \( \alpha \ (<1) \) is a system parameter to absorb rate fluctuation. PacketThroughput_{STA} is the sum of the rates at which packets are successfully transmitted on the uplink and downlink for the STA. The condition (13) indicates that one or more nodes belonging to the cell are experiencing buffer overflow. It is possible that a cell satisfies (12) but not (13), in a case, when the cell does not have much traffic but the channel is congested due to the transmission activities at the overlapping cells.

A cell that satisfies both (12) and (13) becomes a target cell of congestion control. Letting AP_t be the AP of the target cell, the manager selects STAs to move from AP_t to its neighboring cells. The association control is made based on the following policies:

1) The number of moving STAs should be as small as possible.
2) A STA should be moved only if it will not cause congestion at the destination cell.
3) Among the candidate destination APs, the STA should be moved to the AP corresponding to the strongest RSSI.

The more the number of handovers, the larger is the control overhead. Thus policy 1 is to minimize the number of moving STAs. To support this objective, we define load of a STA, as Load_{STA} = OfferedRate_{STA}/TxRate_{STA,AP}. Here TxRate_{STA,AP} is the data rate used for transmissions of frame payload fields between the STA and AP (the shaded part in Fig. 3).

Obviously, the larger the offered traffic and/or the lower the transmission rate, the heavier loaded is the STA for AP_t. For policy 1, heavier loaded STAs are preferred to be moved (from AP_t) over lighter loaded STAs. For policy 2, a STA is moved to a neighboring AP, AP_d, only if the condition (14) is met.

\[ \text{OfferedRate}_{STA} < \text{PT}_{STA,AP_d} \tag{14} \]

In other words, the STA is moved to AP_d only if the destination cell can accommodate the offered traffic volume for the STA. Finally, among the candidate destination APs (i.e., the APs that satisfy (14) for the STA), the AP corresponding to the strongest RSSI is selected as the destination AP for the STA.

When a STA, STAm, is selected to be moved from AP_t, the manager updates the aggregate offered rate (see (13)) for the target cell by decrementing it by OfferedRate_{STAm}. Moreover ATR for the destination cell and its overlapping cells (which operate using the same channel) is incremented by OfferedRate_{STAm}/r_{AP_d,STAm}. After updating the values, the
manager checks if the target cell still satisfies (13). If it does, the manager searches the next STA to move from APt.

- Discussion on TCP traffic

TCP reacts to congestion and controls its rate based on AIMD (Additive Increase Multiplicative Decrease) algorithm. However such a rate change occurs in the order of RTT (round-trip time), i.e., in order of milliseconds, while the control period at the manager is in order of seconds. Therefore we expect that the proposed scheme does not react to the AIMD-based rate fluctuation, but only to the gradual change of the average rate. Hence, the proposed scheme and TCP can stably coexist. Moreover because TCP adjusts its traffic rate, OfferedRate\textsubscript{STA} for STA might be largely changed due to the STAs movement. However, it should be reminded that a STA is moved to a neighboring cell only if the destination cell can accommodate the current OfferedRate for the STA (see (14)). Thus, we expect that TCP throughput will be increased or maintained after the STAs movement. Furthermore, since some amount of channel resource is released at the previous cell of the moving STA, STAs in that cell can now increase their rate.

C. Cell Aggregation

Since both of the offered traffic volume and the channel availability are known for each cell, load balancing among cells is not necessary. Indeed, if all the associated STAs of an AP can be moved to its neighboring cells without hampering the network throughput, such STA movements should be encouraged for energy efficiency, since the AP can now be in the power-saving mode. Our scheme can provide such an association control based on the following policies:

1) The target AP is an AP that is associated with preferably few STAs, which all can be moved to the neighboring cells.
2) To suppress channel interference, the target AP should have overlapping cells that operate using the same frequency channel.
3) Policy 2 described in the previous subsection.
4) Policy 3 described in the previous subsection.

The manager triggers a handover only if destination APs are found for all the STAs of the target AP. A detailed protocol design for actually putting APs in the power-saving mode is left for our future work. The main concern of such a protocol design is to ensure newly arriving STAs are covered by the network. For such a control, Wake-on-WLAN technology [16] can be used.

D. Changing STA’s Association

To change a STA’s association, the manager sends a control frame to the STA, indicating the destination AP and the corresponding channel information. Such a network directed association control can be supported by the upcoming IEEE 802.11v [17], which enables APs to explicitly request STAs to re-associate with an alternate AP.

V. SIMULATION EVALUATION

Using Scenargie network simulator [18], we investigate the proposed scheme with and without cell aggregation capability. The performance of the proposed scheme is compared against:

- Strongest RSSI: The traditional AP selection scheme.
- LB(NumofSTAs): A load balancing scheme [3], which takes account of the number of STAs and RSSI.
- LB(Traffic): A load balancing scheme [8], where load of a cell is defined by traffic rate.

In each scheme, STAs initially associate with APs based on the strongest RSSI policy. In the proposed scheme, STAs inform the server of the measured information using a 160-bytes packet. The manager checks the collected information in every second. 20-bytes of packets are used for handover requests and replies between the manager and moving STAs. \( \alpha \) (see (13)) is set to 0.98.

The performances of the schemes are investigated targeting the real-world scenario depicted in Fig. 1, where 14 APs are non-uniformly distributed in a 600×600 m\(^2\) area. 8 APs are concentrated in the small area (around the condominium), the remaining 6 APs are installed in the rest of the overall area as shown in the figure. The network operates using the IEEE 802.11a [11], using 3 orthogonal frequency channels. The frequency channels are allocated (to the APs) following a simple graph coloring technique.

STAs (users) can be anywhere, but it is natural to expect that users are especially attracted to the public places specifically, the train station, condominium building (for the café), McDonalds shown in the target area. Thus, in our simulations, 40 STAs are distributed in the circle-shaped areas centered at the 1) train station, 2) condominium building, 3) McDonalds, and 4) the center of the map, with a radius of 100, 10, 10 and 300 meters, respectively (see Fig. 1). The first three areas are to create users concentration in the public places, while the last area is for uniform distribution of users in the overall area. Table II shows the simulation scenarios, which have different number of STAs in each circle-shaped area. The larger the scenario number, the more uniform is the STAs’ distribution.

The simulation evaluations are conducted for TCP and UDP traffic. In TCP simulations, STAs upload 5 MB of file using FTP/TCP-SACK. In UDP simulations, STAs have uplink and downlink CBR traffic generated at a random rate in the range of [0 Mbps, 1.2 Mbps]. The data size is set to 512 bytes for
both TCP and UDP traffic.

Fig. 7 compares the network throughput for the TCP and UDP traffic. As the figures show, the performance of Strongest RSSI scheme is inferior to the remaining schemes and the proposed scheme shows the best performance. The proposed scheme with the cell aggregation mechanism achieves around the same throughputs as the scheme without cell aggregation, especially for UDP traffic. The load balancing schemes have lower throughput than the proposed scheme, conceivably because they do not take account of the overall of (3).

The figures also show that, in general, the throughput tends to be smaller when STAs’ distribution is less uniform (e.g., the throughput of S7 is smaller than that of S5). As discussed in Section II, the reason is clear for the strongest RSSI scheme (many STAs select the same AP). The reason for the remaining schemes is as follows. When STAs are highly concentrated around a particular AP, the schemes have to move some of the STAs to farther APs. This reduces the frame transmission rate, r, for the moving STAs, resulting in lower throughput compared to that of a scenario where STAs’ distribution is more uniform. Nevertheless, compared with the strongest RSSI scheme, the overall throughput is improved due to an increase of the 1st and 3rd components of (3). Finally, S1 (where STAs’ distribution is uniform), however, does not show the largest throughput due to the non-uniform AP distribution.

Fig. 8 compares the number of active APs for each scheme. The figures show that the load balancing schemes tend to use all the existing APs. The proposed scheme without cell aggregation utilizes around the same number of APs as that of Strongest RSSI scheme. The proposed scheme with cell aggregation utilizes the smallest number of APs.
VI. TESTBED EVALUATION

We implemented the proposed scheme in a wireless testbed and evaluated its throughput performance. Computers with Cent OS 5.5 (kernel 2.6.25-17) are used for STAs, APs, and the manager (note that the manager and server coexist on a same machine). The APs, the manager, and a source computer (acting as a corresponding node (CN)) are connected to a 100 Mbps Ethernet. The APs and STAs are equipped with IEEE 802.11a WLAN cards made by NEC (Aterm WL54AG). We modified the Atheros device driver to enable measurements of ATR and PacketThroughput (see (13)). The packet transmission rate from the kernel to the device driver is monitored to measure EnqueueRate (see (11)). TCP is used for information collection at the server and for handover requests and replies.

As the testbed environment, we selected an active office area, which has a complicated room and obstacle structure and intermittent movement of people, is used as the test environment. The test scenario consists of 4 APs and 12 STAs. STAs are initially associated with the APs using the strongest RSSI method.

Figure 9. An active office area, which has a complicated room and obstacle structure and intermittent movement of people, is used as the test environment. The test scenario consists of 4 APs and 12 STAs. STAs are initially associated with the APs using the strongest RSSI method.

To see the impact of the proposed scheme, the manager is initially disabled and it is activated after around 40 seconds. Unless otherwise noted, the offered traffic volume for individual cells is initially set to be largely different. Moreover, in order to create traffic dynamics, additional traffic is injected to some cells after around 30 seconds. To simplify experimentation, the CN (corresponding node) generates traffic for individual STAs (i.e., downlink traffic), enabling the offered traffic volume to be adjusted only at CN. Three orthogonal frequency channels of the IEEE 802.11a are allocated such that AP1 and AP3 share the same channel and AP2 and AP4 use the remaining two channels. The testbed evaluations made for three scenarios: the first scenario is to evaluate congestion alleviation method for TCP and UDP traffic, the second scenario is to evaluate congestion alleviation method for UDP traffic, and the last scenario is to evaluate cell aggregation method.

A. Congestion alleviation for UDP traffic

CBR/UDP traffic is generated for individual STAs. Initial setting of the aggregate offered traffic at cells 1, 2, 3 and 4 are 5, 20, 18, and 3 Mbps, respectively. At 30 seconds, the offered traffic rate for cells 2 and 3 are further increased by 10 Mbps.
in Fig. 12 is that ATR can be tuned well to the desired level, $\text{ATR}_\text{th}$. Fig. 13 shows that $\alpha \times \sum_{\text{STA}_i} \text{OfferedRate}_{\text{STA}_i} - \sum_{\text{STA}_i} \text{PacketThroughput}_{\text{STA}_i}$, at each cell. The figure shows that the parameter reacts well to buffer overflow.

Figs. 10 and 11 show the time series plots of the throughput and packet delivery performance for each cell. Because cells 2 and 3 were initially very loaded, the additional traffic injected to the cells (at around 30 seconds) results in only a slight increase of the total throughput. As the offered traffic volume requires a larger resource, as Fig. 11 shows, the cells could not accommodate the overall offered traffic, and hence the packet delivery ratio drops down to 60%. Upon activation of the manager (at around 40 seconds), STAs are moved from congested cells (cells 2 and 3) to non-congested neighboring cells (cells 1 and 4), and the packet delivery ratio is maximized by the association control.

Figs. 12 and 13 show the time series plots of ATR and $\alpha \times \sum_{\text{STA}_i} \text{OfferedRate}_{\text{STA}_i} - \sum_{\text{STA}_i} \text{PacketThroughput}_{\text{STA}_i}$, the parameters used to trigger congestion control (see (12) and (13)). Two important observations can be made from Fig. 12. Firstly, before the congestion (before 30 seconds), ATR at Cell2 already exceeds the threshold value, but as it can be seen in Figs. 11 and 13, there was no buffer overflow. This teaches us that ATR above $\text{ATR}_\text{th}$ does not necessarily indicates buffer overflow. The second point can be observed in Figs. 12 and 13. ATRA measured at each APs. The figure shows that ATR can be well tuned by the association control to the desired value.

Figs. 10-13 show that unfortunately, it took approximately 6 seconds to complete the handover (without depending on the number of moving STAs). The 6 seconds are used for 1) MAC layer disassociation/association, 2) IP route advertisement, 3) IP duplicate address detection, 4) Mobile IP binding update, and 5) Mobile IP binding acknowledgement. Among the operations, there was a software bug corresponding to 2) and we confirmed that by fixing this bug, handover time can be reduced down to 3 seconds. We are now working on this issue.

B. Congestion control for TCP and UDP traffic

The CN generates TCP traffic to the STAs at cells 1, 3, and 4. For Cell2, TCP traffic is generated for STA5, and UDP traffic is generated for STA4 and STA6. The initial settings of the maximum offered traffic volume (i.e., the sum of the CBR rate and the maximum rate of TCP) for the cells 1, 2, 3, and 4 are 15, 27, 12, and 15 Mbps, respectively. At around 30 seconds, the UDP traffic for STA4 is further increased by 5 Mbps. Fig. 14 shows the time series plots of the aggregate throughput and throughput of individual cells. Upon activation of the manager, STA6 is moved from Cell2 to Cell3. This improves the aggregate throughput by approximately 10 Mbps.

Fig. 15 gives a closer look at the time series plots of the individual throughput at Cell2 and Cell3. The offered CBR rate and the maximum offered TCP rate for each STA are also noted in the figure. Until 30 seconds, the UDP throughput achieved for STA4 and STA6 (in Cell2) are 10 Mbps, which are equal to the offered rate. On the other hand, the throughput
of TCP traffic (for STA5) is only 2 Mbps, which is much lower than the maximum offered rate (7 Mbps). The additional traffic injected for STA4 (at around 30 seconds) results in a large throughput degradation for both STA6 and STA5. The throughput degradation is especially significant for the TCP traffic for STA5 (down to 0 Mbps). This clearly shows severe unfairness between TCP and UDP traffic. At around 40 seconds, the manager is activated, then STA6 is moved to cell 3. After approximately 5 seconds of handover delay (due to the issues mentioned in the previous subsection), all the traffic at both the cells could be transmitted at their maximum rate, resulting in 10 Mbps of throughput improvement.

In the above mentioned scenario, a STA with UDP traffic is moved from the congested cell. Fig. 16 shows results achieved from an experiment where STAs with TCP traffic are moved. In the experiment, the target scenario consists of two cells: one has three STAs (STA1-STA3) and the other has one STA (STA4). The manager is activated at around 70 seconds. STA1 and STA4 have UDP traffic with 25 and 10 Mbps of offered rates, respectively. STA2 and STA3 have TCP traffic with maximum 6 Mbps of offered rate. As the figure shows, until activation of the manager, due to the large traffic volume offered to STA1, the coexisting TCP traffic significantly reduced their traffic rate down to 0 Mbps. Upon activation of the manager, the STAs with TCP traffic are moved from Cell1 to Cell2. The association control results in the TCPs increase their traffic rate at the destination cell.

The results achieved in Figs. 10-16 also show that monitoring/estimation of traffic, channel, and link condition made at Layer1-Layer 3 are realized with sufficient accuracy. However because TCP controls its rate at Layer 4, the OfferedTraffic estimated at Layer 3 is the result of the TCP’s rate adaptation. This implies, on the one hand, that the proposed congestion alleviation scheme does not negatively affect the TCP’s congestion control, and the two congestion control schemes can stably coexist. On the other hand, however, if the proposed scheme is aware of the TCP’s maximum rate, further throughput improvement is possible. The following is a list of some approaches for this challenging task.

- Enhancement of the proposed scheme by tuning the control parameter (e.g., by increasing $\alpha$).
- Exploit TCP flow characteristics for association control. As Figs. 15 and 16 show, if TCP is transmitting at its maximum rate (i.e., at RWND), the throughput does not fluctuate with time. On the other hand, if TCP is not transmitting at its maximum rate, the throughput tends to fluctuate largely. By monitoring such characteristics, we believe that it is possible to “guess” that TCP is not transmitting at its maximum rate.
- Cross layer approach that enables direct information exchange between Layer 4 and Layer 3.

C. Cell aggregation

We evaluate the performance of the cell aggregation technique. The testbed scenario is the same as shown in Fig. 9. Each cell has STAs with UDP traffic whose aggregate offered rate is 9 Mbps, thus congestion is not an issue. Fig. 17 shows the evaluation result. As the figure shows, upon activation of the manager (at around 30 seconds), the STAs at cells 1 and 4 are moved to cells 2 and 3, enabling AP1 and AP4 be in the power-saving mode. Although the throughput is degraded during the disconnection period, as soon as the STAs are associated with the destination APs, the network achieves its maximum throughput utilizing the half of the total number of APs.

VII. Conclusion

In community WLANs, APs and STAs (i.e., users) tend to concentrate on different areas. A concentration of STAs often
results in the APs and STAs in that particular area suffer from congestion. A concentration of APs, on the other hand, causes energy wastage. In this paper, we proposed an association control scheme that aims at maximizing throughput by congestion alleviation method and reducing energy consumption by cell aggregation method. We first analytically formulated that the throughput of a node belonging to a WLAN can be expressed as the multiplication of three components: success probability, frame transmission rate, and air-time ratio. The success probability is the probability of collision-free transmission. The frame transmission rate is the speed of the data transmission, and it is a function of the link quality. The air-time ratio is the ratio of the channel busy time to the total time. The frame transmission rate and air-time ratio can easily be monitored and estimated. The success probability, on the other hand, is a function of the number of contending nodes, which are extremely difficult to monitor in the real systems. Due to this reason, we extended our theoretical study and showed that success probability can be maximized by controlling air-time ratio. Finally, we proposed our association control scheme for throughput maximization and energy efficiency by taking account of the multiplication of frame transmission rate and air-time ratio. The performance of the proposed scheme is investigated by extensive evaluations using computer simulations and testbed experiments. Both the simulation and testbed experiments strongly indicate the correctness of the theoretical work and the effectiveness of the proposed scheme for realistic network scenario with both UDP and TCP traffic. The testbed experiments prove that the proposed scheme and TCP can stably coexist for congestion control. We plan to further enhance the proposed scheme for improved TCP throughput. In addition to simulation and testbed evaluations, we intend to evaluate the scheme using statistical analysis. Our future work also includes a study on energy efficiency induced by the cell aggregation mechanism.

ACKNOWLEDGMENT

This research was performed under research contract of "Research and Development for Reliability Improvement by The Dynamic Utilization of Heterogeneous Radio Systems", for the Ministry of Internal Affairs and Communications, Japan.

REFERENCES


