An Incentive-Compatible Load Distribution Approach for Wireless Local Area Networks with Usage-Based Pricing

Bo GU†, Kyoko YAMORI††, Sugang Xu†, Members, and Yoshiaki TANAKA†††, Fellow

SUMMARY Recent studies have shown that the traffic load is often distributed unevenly among the access points. Such load imbalance results in an ineffective bandwidth utilization. The load imbalance and the consequent ineffective bandwidth utilization could be alleviated via intelligently selecting user-AP associations. In this paper, the diversity in users’ utilities is sufficiently taken into account, and a Stackelberg leader-follower game is formulated to obtain the optimal user-AP association. The effectiveness of the proposed algorithm on improving the degree of load balance is evaluated via simulations. Simulation results show that the performance of the proposed algorithm is superior to or at least comparable with the best existing algorithms.

key words: quality of service, load balancing, pricing, local area network, game theory

1. Introduction

In dense metropolitan areas, it is often the case that a user can detect several APs simultaneously. Users located in overlapping cells, however, tend to scan all available channels and associate themselves to the AP that has the strongest Received Signal Strength Indicator (RSSI). A key consequence of this behaviour is that the traffic load is often distributed unevenly among the APs [1]–[4].

To address this problem, we introduce and analyze an incentive-compatible load balancing approach, in which user-AP associations are intelligently determined based on not only the signal strength but also the load level at each AP.

In particular, we assume that APs deployed by different providers (e.g., small businesses and individuals) are able to select to join into a federated network, such as [5], [6]. A Central Aggregation Server (CAS), which is independent from each provider’s financial interests, is deployed into the federated network to keep track of the load level at each AP. Compared with setting up mutual agreements between each pair of WLAN providers, the centralized approach is more scalable [7].

The CAS attaches its brand name to the APs and ensures that a consistent product is offered among the APs deployed by different providers. The CAS also handles the billing for the service. The CAS collects the revenue from users, and then redistributes the revenue to the deployed business partners.

After receiving the service request from each user, the CAS helps to find an AP that can best accommodate the service request. Generally, the mission of CAS is to select an AP for each user from two kinds of candidate APs (if exist): (i) a candidate whose service could be provided in place (hereafter termed local candidate); and (ii) a candidate whose service could be provided for the user while requiring user’s physical roaming (hereafter termed remote candidate).

As depicted in Fig. 1, our algorithm trades off signal strength with load by suggesting a user to change the association from an heavily loaded AP with stronger signal (i) to a lightly loaded local candidate with possibly weaker signal; or (ii) to a lightly loaded remote candidate whose service is provided requiring user’s physical roaming.

We focus on learning the economic behaviours of the wireless users, and structuring a Stackelberg leader-follower game [8] to obtain the optimal user-AP association.

The remainder of this paper is organized as follows.
Section 2 introduces the related work; Sect. 3 describes the system model; Sect. 4 presents the sequence of steps involved in QoS negotiation and admission control, which forms an overview of the proposed algorithm; Sect. 5 describes the classification of APs in the federated network; Sect. 6 proposes a Stackelberg game structure to obtain the optimal AP-user association; Sect. 7 and Sect. 8 describe our simulation settings and show the simulation results. Finally, Sect. 9 concludes this paper.

2. Related Work

As studied in work [9]–[14], the load imbalance problem can be alleviated by balancing the load among the APs via intelligently selecting the user-AP association. They suggest adding load information, such as packet loss rate, throughput, retransmission probability or airtime cost [12], to the beacon frames. Users could use such information in addition to the signal strength to select the APs.

The above-mentioned approaches show nice features. However, these approaches distribute users only across available overlapping cells, hence that the load could be only locally balanced.

Exploiting user mobility for load distribution within an entire network is by no means a new idea. A seminal paper [15], describes a scenario where a user using her wireless connection at an airport gate cannot get enough bandwidth to complete her e-mail transmission, because many passengers at that gate are surfing the web. As a solution to this problem, a pervasive computing system helps to find a nearby gate with greater capacity and encourage the user to move to that gate.

In [16], the authors focus on distributing traffic loads within the network using a centralized approach. A centralized server is deployed to monitors the bandwidth allocation in the entire network. The centralized server then helps to identify an AP where an incoming user’s QoS bound (i.e., a minimum and a maximum bound on the bandwidth) can be adequately met. However, due to the inherent contention-based medium access property, guaranteeing a certain amount of bandwidth to a user in WLANs is not a trivial matter in practice. Furthermore, in real networks, users are usually non-cooperative and selfish in the sense that they make decisions to maximize their individual utilities or payoffs. The network can be easily overtaxed by the selfish users requesting bandwidth as much as possible.

In this paper, we assume that each user contends for channel access according to some user-chosen access probability [17]. Saturation throughput of each user can be calculated according to the user-chosen access probabilities, and such saturation throughput is employed to reflect the load level at each AP. Compared to guaranteeing a certain amount of bandwidth, the access-probability-based resource allocation is easier to implement [17]. Furthermore, pricing is employed to provide incentives for selfish users to utilize the channel resource efficiently.

When interactions among users are taken into account, game theory becomes a natural modeling framework. In [18], the authors extend the scenario described in [15] and model the system as a game in order to obtain the stable system outcomes (i.e., Nash Equilibrium). The paper contributes to a deep understanding of the interaction among users. However, there are two drawbacks to overcome: (i) the load at an AP is estimated based on the number of associated users. In some deployment scenarios, data flows have bursty characteristics and generate dynamic load on the APs. Therefore, the number of existing users can not reflect the load level at each AP properly [14]; (ii) the proposed algorithm fails to produce a Nash equilibrium, when the diversity in users’ utilities is taken into account.

In this paper, the diversity in users’ utilities is sufficiently taken into account. To be specific, a Stackelberg leader-follower game is structured to analyze the interaction between the CAS and users. Given the best response of each user, the leader (i.e., CAS) can derive the private utility information of each user through backward induction, and then determine the optimal AP-user association.

3. System Model

The model we are envisioning assumes that each user communicates with a single AP directly. In order to maximize the system capacity and keep the interference to a minimum, neighboring APs (if exist) are configured to operate on different RF channels. A CAS is deployed in the federated network to maintain network status by periodically collecting the load level at each AP.

Moreover, it is assumed that users always have a packet available for transmission. Namely, the network is operated in saturation conditions. Each user contends for channel access according to some user-chosen access probability. A transmission is successful if and only if there is a single transmission attempt. Hence that QoS differentiation is achieved when users with high access probability transmit more often than those with low access probability [17].

The reason why this simple MAC model is employed is because we want to abstract out the essential features of QoS-enabled MAC, and hence can avoid being overwhelmed by the complexity of realistic MAC protocols. For window-based protocols such as IEEE 802.11, there is a direct correspondence between access probability and window size [19]. Therefore, the analysis and related results here can be extended to such scenarios easily.

Let \( u = \{\xi_1, \xi_2, ..., \xi_N\} \) be the vector of access probabilities for the user 1, 2, ..., \( N \). The saturation throughput of user \( k \) is given by \( \tau_k \) as follows.

\[
\tau_k = \xi_k \prod_{j=1,j \neq k}^{N} (1 - \xi_j)
\]  

(1)

User demand is assumed to be elastic [20], and the utility is given by \( U_k(\tau_k) \) as follows.

\[
U_k(\tau_k) = \alpha + \theta_k \ln(\tau_k)
\]  

(2)
where $\theta_k$ is a user-dependent scale factor and can be thought of as a parameter representing the priority of user $k$’s willingness to pay. $\alpha$ is a constant.

In most cases, commercial wireless service service providers offer users fixed-price plans (e.g., ten dollars per month for getting wireless access at three million hot-spots worldwide). The fixed-price plans do not differentiate between short-term access and long-term access. As a consequence, the network can be easily overtaken by heavy data applications with long-term usages, leading to the tragedy of the commons phenomenon [21]. Due to the possible congestion, for example, AT&T stopped offering the fixed-price plan to its new smart-phone customers in 2010.

A usage-based pricing is hence employed. The CAS proposes a price for access, which is set to be a linear function of the user-chosen access probability. The user can either accept the price and connect, or reject and leave. The service session ends at the first time the user rejects the CAS’s proposal, including three cases: (i) the user finds the slot price is too high to accept; (ii) the user’s utility decreases below the price charged due to congestion; and (iii) the user does not intend to connect any more. The overall payment charged grows proportionally with the time each user connects.

4. QoS Negotiation and Admission Control

The steps involved in QoS negotiation and admission control [16] are shown in Fig. 2.

**Step 1:** User $k$ arrives at the federated network, and detects the existence of APs via beacons periodically broadcasted in the federated network. In order to reduce network management overhead, pricing information is also contained in the beacons.

**Step 2:** The user performs authentication and indicates her own value of access probability, through sending a Service Level Specification (SLS) packet [16]. As shown in Fig. 3, the SLS packet contains: username, password, access probability, and a list of APs (i.e., $AP_{List_k}$), which includes the APs that are within communication range of the user.

**Step 3:** The CAS selects a candidate AP (if exists) that can best accommodate the user’s service request. The details of candidate-AP-selection procedure are described later in Sect. 5 and Sect. 6.

**Step 4:** The CAS informs the user of the candidate AP through sending a Service Level Acknowledgement (SLAck) packet. As shown in Fig. 3, the SLAck packet contains: a service type field indicating if the service is provided in place or if roaming is required; the AP which provides the service; the physical coordinates of the AP; user $k$’s network access key; and the estimated saturation throughput.

**Step 5:** The user determines either to associate to the candidate AP, or to reject and leave.

Note that our negotiation process mainly differs from the one described in [16] in the following aspects:

- Pricing information is added into the beacon packet.

This information is used to provide incentives for users to use the channel resources effectively.

- Each user specifies a value of access probability instead of the QoS bound.
- The CAS informs the user of the estimated saturation throughput instead of the permissible bandwidth bound.

5. Classification of APs in the Federated Network

The CAS maintains all per-cell state for the entire network, it can monitor and control the use of the wireless bandwidth efficiently. Knowledge of the system state enables the CAS to easily identify heavily loaded APs and hence distribute traffic loads from a heavily loaded AP to a lightly loaded AP.

For ease of understanding, an example is illustrated in Fig. 4. A user is within hearing range of three APs operating on different RF channels (channel a, b and c). The user’s wireless adapter associates, by default, to the AP from which it senses the strongest signal temporarily.
the locations of APs: three categories according to SNR (Signal Noise Ratio) and CAS classifies all of the APs in the federated network into their alternatives, have clear preferences, and take action de-
wireless users are rational in the sense that they are aware of wireless users. Under an assumption that the CAS and the equilibrium of bandwidth sharing between the CAS and the gain the highest profit [22].

competition in quantity, given the best response of each fol-
6. Stackelberg Game and Profit Maximization

In Economics, the Stackelberg game is used to analyze the competition in an oligopoly market. In such a market, a leader firm commits a strategy first and then other follower firms move sequentially. The equilibrium of the game can be obtained by backward induction. For the case of oligopoly competition in quantity, given the best response of each follower, the leader can choose the optimal supply quantity to gain the highest profit [22].

This Stackelberg game structure is applied to obtain the equilibrium of bandwidth sharing between the CAS and the wireless users. Under an assumption that the CAS and the wireless users are rational in the sense that they are aware of their alternatives, have clear preferences, and take action de-

After receiving a service request from the user, the CAS classifies all of the APs in the candidate-AP-selection procedure. Classification of APs in the candidate-AP-selection procedure.

Fig. 4

• First-choice AP (represented by a triangle filled with black color): the one whose transmission range can cover the user, as well as SNR is above a certain threshold;
• Second-choice AP (represented by a triangle drawn in solid line): the one whose transmission range can cover the user, but SNR is below a certain threshold; or the one whose service is provided requiring user’s physical roaming within a certain distance d;
• Invalid AP (represented by a triangle drawn in dotted line): the one who is out of the certain distance d that the user could roam.

The rationale behind the maximum roaming distance d is to reduce the possibility of asking the user to roam a long distance for service.

6. Stackelberg Game and Profit Maximization

Let Nl represent the number of existing users in the cell of AP l. The access probabilities chosen by the existing users of AP l and the incoming user Nl + 1, are denoted by ξf,l, ∀j ∈ {1, 2, ..., Nl} and ξf,l+1, respectively. Furthermore, the prices charged are set to be βξf,l, ∀j ∈ {1, 2, ..., Nl + 1}. Supposing that the incoming user is associated to AP l, the payoff for the incoming user is given by

\[ \phi_{Nl+1}(\xi_{f,l+1}) = \alpha + \theta_{Nl+1} \ln \left( \frac{\xi_{f,l+1}}{\beta} \right) - \beta \xi_{f,l+1} \] (3)

Therefore, the optimal access probability can be obtained by differentiating the profit function and then setting it to zero. We find that the unique optimal solution is given by

\[ \xi_{f,l+1}^* = \frac{\theta_{Nl+1}}{\beta} \] (4)

Based on the best response of each user (i.e., ξf,l+1 = \xi_{f,l+1}^* = \theta_{Nl+1}/\beta and \xi_{f,l} = \xi_{f,l}^* = \theta_j/\beta, ∀j ∈ {1, 2, ..., Nl}), the CAS can induce the \( \theta \)-value of each user as follows:

\[ \theta_{Nl+1} = \beta \xi_{f,l+1} \] (5)
\[ \theta_j = \beta \xi_{f,l}^*, ∀j ∈ {1, 2, ..., Nl} \] (6)

The utility of each existing user decreases with the admission of new users. In case that the utility decreases below the price charged, the existing user may reject the price and leave. This imposes the CAS a capacity constraint on its payoff maximization problem. Intuitively, the CAS should associate the incoming user to a lightly loaded AP in order to achieve the highest payoff.

The payoff of the CAS is obtained from the net revenue gained from existing users and the incoming user. When the incoming user is associated to AP l, the payoff of the CAS can be defined by

\[ \psi(l) = \arg \max_{l \in L} \left\{ \beta \xi_{f,l+1}^{*} + \sum_{l>1} \beta \xi_{f,l}^{*} + \sum_{m \neq 1} \sum_{k=1}^{Nl} \beta \xi_{f,k}^{*} \right\} \] (7)

where

- \( L \) is the set of APs in the federated network.
• $N_m$ represents the number of existing users in the cell of AP $m$, $\forall m \in L$.
• The first part inside the square brackets of Eq. (7), i.e., $\beta \xi_{l}^{(i)}$, represents the revenue gained from the incoming user.
• The second part inside the square brackets of Eq. (7) represents the revenue gained from the existing users of AP $l$.
• The last part inside the square brackets of Eq. (7) represents the revenue gained from the other APs in the federated network.

The payoff is maximized when $\exists l \in L,$ and

$$\alpha + \beta \xi_{k}^{(i)} \ln \xi_{k}^{(i)} \prod_{j=1, j \neq k}^{N_l+1} \left(1 - \xi_{j}^{(i)}\right) > \beta \xi_{l}^{(i)}.$$  

$$\forall k \in \{1, 2, ..., N_l + 1\} \quad (8)$$

From the set of first-choice APs, the CAS selects a local candidate where the user-chosen access probabilities can satisfy Eq. (8). If there are more than one first-choice AP that can accommodate the service request, the CAS chooses the one that could provide the maximum saturation throughput (i.e., $\arg \max_{m \in L} \left[\prod_{j=1}^{N_l+1} (1 - \xi_{j}^{(m)})\right]$). In this way, the admission control procedure tries to associate each user to the most lightly loaded AP, hence maximizes the total utilization in the network. If there is no such local candidate, the CAS selects a remote candidate from the second-choice APs in the same way. The process flow is summarized as shown in Fig. 5.

7. Evaluation Scenario

We consider the uplink of random access MAC where each user contends for channel access according to some user-chosen access probability. A transmission is successful if and only if there is a single transmission attempt — there is no carrier sensing, and we do not model explicit back-off.

APs are deployed in a 300 m × 300 m square area as shown in Fig. 6. Each AP locates at the central point of a 150 m × 100 m square area, which is termed home area hereafter. Each user arrives according to a Poisson process and stays for a time, which is exponentially distributed. Furthermore, two APs (i.e., AP 3 and AP 4) are selected to be heavily loaded APs. The arrival rate of users’ requests at the home area of the heavily loaded AP is twice higher than the arrival rates at the other home areas. Other detailed simulation settings are summarized as shown in Table 1.

Algorithms termed no load balancing (NLB), network directed roaming (NDR) [16], and distributed myopic se-
Table 2

<table>
<thead>
<tr>
<th>Alternative APs</th>
<th>The proposed algorithm</th>
<th>The NLB algorithm</th>
<th>The NDR algorithm</th>
<th>The DMS algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal AP</td>
<td>the AP which provides the maximum saturation throughput</td>
<td>the AP with the strongest signal strength</td>
<td>the AP with the shortest roaming distance</td>
<td>the AP with the lowest cost [18]</td>
</tr>
</tbody>
</table>

Fig. 7  Balance index of the network as a function of the simulation time.

8. Simulation Results

An important element in a load balancing system is the function used to evaluate the balance of the system. Comparing the throughput at different APs is not feasible, because the varying nature of the traffic generated by the stations always makes these values different. To explore the effect achieved by distributing load across the network, we adapt the concept of balance index firstly introduced in [23] and then used in [16]. Let $B_i$ be the total throughput of AP $i$, the balance index is defined as:

$$\gamma = \frac{\left(\sum B_i\right)^2}{n \sum B_i^2}$$

where $n$ is the number of APs in the federated network. The balance index is 1 when all APs have exactly the same throughput, and tends to $1/n$ when the traffic load is severely unbalanced. The main target for a load balancing system is to maximize the balance index.

Figure 7 shows the balance index as a function of the simulation time. The curves show that the proposed algorithm has a larger performance gain even compared to the NDR algorithm and the DMS algorithm. This is not surprising because compared to the number of users in service, the permissible saturation throughput reflects the load level at each AP more properly. Hence, the load can be distributed across the federated network more evenly.

Figure 8(a) and Fig. 8(b) show the blocking probability and the total throughput as a function of the simulation time, respectively. The curves show the effect of the proposed algorithm in terms of improving QoS. For instance, at the time of 30 min, the average blocking probabilities are 0.155, 0.067, and 0.065 in the NLB algorithm, the NDR algorithm, and the DMS algorithm, respectively, and decrease to 0.062...
in our proposed algorithm. Besides, the total throughput of six cells is 9.92 Mbps, 11.74 Mbps, and 11.87 Mbps in the NLB, the NDR, and the DMS algorithm, respectively, and increases to 12.01 Mbps in our proposed algorithm.

Figure 9(a) and Fig. 9(b) show the number of users in service and total revenue as a function of the simulation time, respectively. The curves show the effect of the proposed algorithm in terms of increasing the total revenue of the federated network.

9. Conclusions and Future Work

In this paper, a Stackelberg game structure is applied to obtain the equilibrium of bandwidth sharing between the CAS and the wireless users. The game is composed of three steps: (i) the CAS predefines a pricing scheme; (ii) users choose their access probabilities to optimize their payoffs, namely, best response strategies; (iii) the CAS uses the best response information to determine the AP-user association, which in turn maximizes its total revenue. In order to exploit users’ mobility for load balancing, a remote AP can also be selected by the CAS. The remote AP with better QoS guarantee (i.e., saturation throughput) encourages the users to connect. Hence, the load could be distributed across the federated network dynamically.

The performance of the proposed algorithm is evaluated via simulations. The observed results show that the proposed algorithm achieves greater or at least comparable balance and overall utilization, comparing to the best existing algorithms.

Several assumptions used in the proposed algorithm may be considered unrealistic or complex (e.g., the ability of the network to determine a user’s location, and the ability to direct the user to the locations of APs with the most available capacity). It is worth pointing out that the main scope of this paper is to demonstrate the feasibility of an incentive-compatible load distribution approach for WLANs. In the future, we would like to develop a more practical management system based on the theoretical foundation presented in this study.

References

Bo Gu received the B.E. degree in computer science from Tianjin University, Tianjin, China, in 2004, and M.E. degree from Peking University, Beijing, China, in 2007, respectively. He is currently studying towards his Ph.D. degree at Global Information and Telecommunication Studies, Waseda University. From 2007 to 2011, he was a research engineer at Sony Digital Network Applications, Japan. He is currently a research associate at Global Information and Telecommunication Studies, Waseda University. His research interests include resource and mobility management of wireless networks, QoS provisioning, and next generation wireless networks. He received the IEICE Young Researcher's Award in 2011, and the IEICE Communication Quality Conference Premium Award in 2012.

Kyoko Yamori received the B.A. degree in business administration in 1995, and the 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 Business Administration, Asahi University. And 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 Networks Research Award in 2006, and the Ericsson Young Scientist Award in 2006.

Sugang Xu received his B.E. and M.E. degrees in computer engineering from Beijing Polytechnic University, Beijing, China, in 1994 and 1997, respectively, and Ph.D. degree in information and communication engineering at the University of Tokyo, Tokyo, Japan, in 2002. He joined the Global Information and Telecommunication Institute, Waseda University in 2002, as a research associate there. Since 2005, he joined National Institute of Information and Communications Technology (NICT), Tokyo, Japan, as an expert researcher. He is also a visiting researcher at Waseda University. His research interests include algorithms, network architectures, photonic network control, optical grid network systems, parallel and distributed processing. He is a member of IEEE.

Yoshiaki Tanaka received the B.E., M.E., and D.E. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1974, 1976, and 1979, respectively. He became a staff at the Department of Electrical Engineering, the University of Tokyo, in 1979, and has been engaged in teaching and researching in the fields of telecommunication networks, switching systems, and network security. He was a guest professor at the Department of Communication Systems, Lund Institute of Technology, Sweden, from 1986 to 1987. He was also a visiting researcher at the Institute for Posts and Telecommunications Policy, from 1988 to 1991, and at the Institute for Monetary and Economic Studies, Bank of Japan, from 1994 to 1996. He is presently a professor at Global Information and Telecommunication Institute, Waseda University, and a visiting professor at National Institute of Informatics. He received the IEEE Outstanding Student Award in 1977, the Niwa Memorial Prize in 1980, the IEICE Achievements Award in 1980, the Okawa Publication Prize in 1994, the TAF Telecom System Technology Award in 1995 and in 2006, the IEICE Information Network Research Award in 1996, in 2001, in 2004, and in 2006, the IEICE Communications Society Activity Testimonial in 1997 and in 1998, the IEICE Switching System Research Award in 2001, the IEICE Best Paper Award in 2005, the IEICE Network System Research Award in 2006 and in 2008, the IEICE Communications Society Activity Award in 2008, the Commendation by Minister for Internal Affairs and Communications in 2009, and the APNOMS Best Paper Award in 2009.