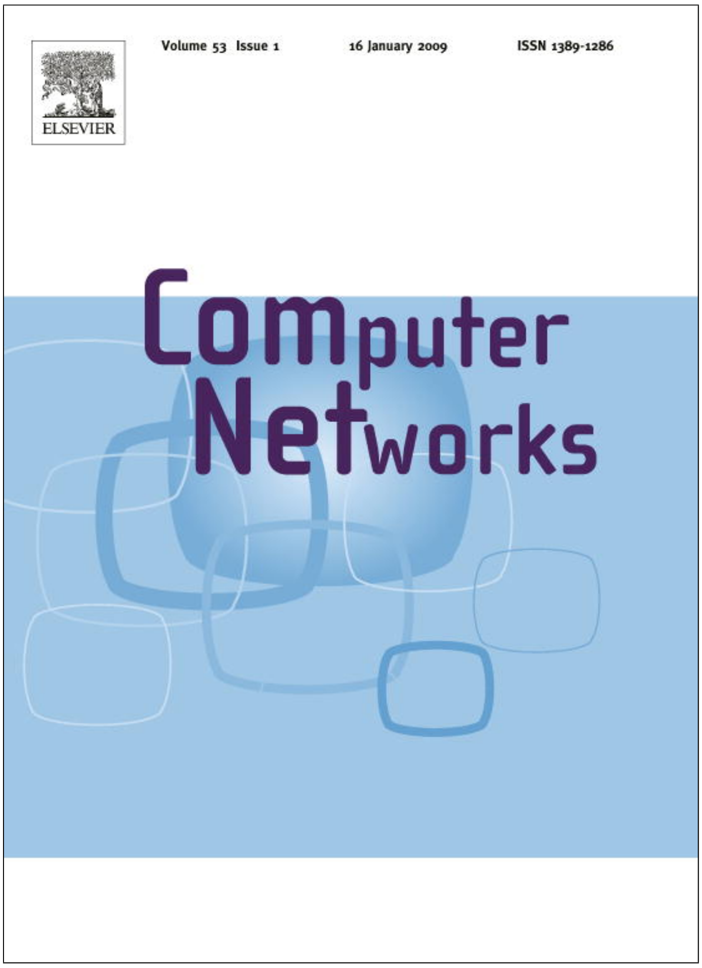

Volume 53 Issue 1
16 January 2009
ISSN 1389-1286
ELSEVIER
Computer Networks

Computer Networks 53 (2009) 45–59

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Efficient multicasting over large-scale WLANs through controlled association

Ai Chen, Dongwook Lee, Prasun Sinha *

Department of Computer Science and Engineering, The Ohio State University, 395 Dreese Laboratories, 2015 Neil Avenue, Columbus, OH 43210-1277, USA

ARTICLE INFO

Article history:
Received 15 January 2008
Received in revised form 15 September 2008
Accepted 16 September 2008
Available online 1 October 2008

Responsible Editor A. Capone

Keywords:
Wireless LAN
Association control
Algorithm/protocol design
Algorithm/protocol analysis

ABSTRACT

Support for efficient multicasting in WLANs can enable new services such as streaming of TV channels, radio channels, and visitor's information. With increasing deployments of large-scale WLANs, such services can be made available to a large number of users. However, any new multicast based services must minimally impact the existing unicast services which are currently the core services offered by most WLANs. In this paper, we leverage the flexibility of associating with different access-points (APs), which occurs often due to overlapping coverage of APs, to optimize the network's objective. Motivated by different revenue functions and network scenarios, three different optimization objectives are considered which are: maximizing the number of admitted users (MNU), balancing the load among APs (BLA), and minimizing the load of APs (MLA). We show that these problems are NP-hard and present centralized approximation algorithms and distributed approaches to solve them. These algorithms compute which AP a user should be associated with. Using simulations we evaluate their performance and compare them to a naive approach in which users associate to the AP with the best RSSI (Received Signal Strength Indicator).

1. Introduction

The goal of anytime-anywhere connectivity is becoming a reality with increasing deployments of large scale Wireless LANs. Networks deployed in industrial campuses, academic campuses, and cities are some scenarios that illustrate the scale at which WLANs are being deployed today. The city-wide network in Chaska, Minnesota[1] is one such example that provides WLAN coverage in a 15 square miles area since October 2004. A similar network is operational in Taipei[2] that consists of 2300 APs and provides coverage to 50% of the city's population, and is planned to be extended to provide coverage to 90% of the city's population in the near future.

While unicast services are essential for providing Internet access to individual users through WLANs, efficient multicast support from the network can be leveraged for distribution of live or stored multimedia data. Such services will enable distribution of multimedia rich content such as local news, visitor's information, and local TV channels.

While introducing media-rich multicast streaming in WLANs, it is critical to ensure that the multicast services use the resources efficiently and the unicast services get minimally affected.

However, the 802.11 standard [1] cannot efficiently maximize resource usage, since uncontrolled association causes multiple APs with overlapping regions to transmit the same multicast packets, thereby wasting resources for unicast services. In this paper, we study how to provide efficient multimedia multicast service to users through controlling the user-to-AP association in WLANs. Improving the efficiency of multicast services makes it feasible to introduce multicasting support while minimally impacting unicast users.

* Corresponding author. Tel.: +1 614 292 1531; fax: +1 614 292 2911.
E-mail addresses: chenai@cse.ohio-state.edu (A. Chen), leedon@cse.ohio-state.edu (D. Lee), prasun@cse.ohio-state.edu (P. Sinha).

[1] http://www.chaska.net.

[2] http://english.taipei.gov.tw.

1389-1286/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2008.09.013

Although association control has already been considered by both the research community and the industry, previous research on association control in WLANs primarily focused on unicast traffic [2–6]. The problem of unbalanced AP load under RSSI-based association is discussed in [2]. In [3,4], metrics other than RSSI are evaluated for association with APs, but only for unicast traffic. These works do not consider load-balancing between APs. Recent work [5,6] has explored the idea of association control to balance the network load and provide max–min fairness among users. A short version of this paper [7] was the first to study the problem of association control in WLANs. In a recent work [8] the joint optimization problem of packet scheduling along with association control in the context of multicasting in WLANs was studied. It uses a network model which constrains users to connect to APs in the same channel. In contrast, our model allows a user to connect to any neighboring AP by switching its channel. In addition, the objectives as well as the algorithmic methodology used are also different.

Our study of association control considers three different objective functions that are aimed to optimize typical revenue models used by service providers. The objective functions considered are as follows:

- Maximizing number of users (MNU): Under high load scenarios, it may not be feasible to satisfy all the users' multicast requests. In such cases, it may be critical to maximize the number of admitted multicast flows. If the service provider charges customers based on the duration of the multicast flows, then this objective function will be of interest.
- Balancing load among APs (BLA): In case the users' requests can be all met, it is critical to balance the multicast load so that the available fraction of time for unicast is also balanced. More precisely, the objective here is to minimize the maximum fractional time used by an AP for multicast traffic. This will lead to fairer share of the unicast bandwidth in case of uniform distribution of users across APs. If the revenue function gives higher weight to unicast traffic and the revenue function for unicast traffic per user is concave, then BLA will likely lead to improved overall revenue for the service provider.
- Minimizing load of APs (MLA): The objective is to reduce the sum of multicast load across all the APs in the WLAN. This will maximize the total available time for unicasting. Under revenue models where there is a flat rate per byte of unicast data, and a scenario with a sufficient number of users requesting unicast traffic, MLA may be the desired objective function for the service provider.

We make the following contributions in this paper. First, we show the NP-hardness of the three problems even if we restrict the problem so that multicast/broadcast packets are always transmitted at the lowest rate supported by the physical layer. Second, we reduce the three problems to other known problems and present centralized approximation algorithms. For MNU, MLA and BLA, we present approximation algorithms with approximation factors of $8, \log_{\frac{8}{7}}(n) + 1$ and $\ln(n)$ respectively. Third, we present distributed approaches to solve the problems, although we believe that in smaller WLANs (of the order of 100 APs) centralized algorithms are still feasible to execute. Note that distributed solutions are preferred in large networks due to their lower signaling complexity in the wired network.

Fourth, through simulations we study the performance of the proposed distributed and centralized solutions for the three objectives. We also compare the performance of MLA, BLA, and MNU algorithms with the optimal solutions.

The presented formulations, NP-hardness proofs, and solutions for problems MNU and BLA, also apply if we consider unicast flows in place of multicast flows. However, the problem MLA is not NP-hard if we only consider unicast flows. For MLA with unicast flows only, the total load of all APs is a summation of the load generated by each user. The latter can be minimized by selecting APs based on RSSI.

The rest of this paper is organized as follows: Section 2 summarizes relevant related work. Section 3 defines the network model, the problems, the notations, and the terminology used in the paper along with the solution framework. The centralized and distributed algorithms for MNU, BLA, and MLA are described in Sections 4–6, respectively. Section 7 presents a detailed evaluation of our approach and comparison with the RSSI-based approach using simulations. The future work is described in Section 8. Finally, Section 9 concludes the paper.

2. Related work

In this section, we outline related works in the areas of MAC layer multicast/broadcast and controlled association in wireless networks.

MAC layer multicast protocols: IEEE 802.11 MAC protocol implements multicast using broadcast. As the 802.11 broadcast is unreliable, several protocols [9–16] have been proposed to improve reliability. Kuri and Kasera [9] proposed a reliable multicast protocol for WLANs. Tang and Gerla [10,11] proposed an 802.11 extension for ad hoc networks that confirms that at least one receiver has received the broadcast packet. In [12], Tang and Gerla proposed the BMW (Broadcast Medium Window) protocol which implements broadcast based on unicast and exploits overhearing by receivers. In [13], Sun et al. proposed BMMM (Batch Mode Multicast MAC) protocol to implement reliable MAC layer multicast. Some MAC layer multicast/broadcast protocols, such as BPBT [14], RMAC [15], and 80211MX [16], use busy-tone to implement multicast reliability, while Chaporkar et al. [17,18] proposed algorithms for maximizing throughput for MAC layer wireless multicast using busy tones. In contrast to these approaches, our solutions are implemented above the MAC layer, and they can co-exist with such MAC layer extensions for supporting multicasting.

Association control: In 802.11 networks, user nodes often use RSSI as the key metric in selecting the AP. The problem of unbalanced AP load under RSSI based association was discussed in [2]. In [3,4], new metrics were studied to select a unicast AP instead of signal strength. Packet error rate and number of users were used in [3]. The authors showed deployability and robustness of their AP selection architecture. Association time, system load, and signal/noise ratio (SNR) together were used to initiate handoff in [4]. The authors argued that their approach can provide Quality of Service (QoS) guarantee.

Recent works [5,6] have explored the idea of association control to balance the network load and provided max–min fairness among users. The authors in [5] proved that balancing the network load is equivalent to achieving the max–min fairness, and presented algorithms that achieve a constant-factor approximation to max–min fair bandwidth allocation. In [6], an analytical model was formulated for AP selection as an optimization problem to maximize different utility functions. The authors provided the optimal association results for some simple cases. However, these works primarily focused on unicast traffic.

In a recent work [8] the joint optimization problem of packet scheduling along with association control in the context of multicasting in WLANs was studied. However, it uses a different network model in which all APs and users use one common channel. In contrast, in this work we allow users to switch to the channel of a neighboring AP for association, and assume that adequate frequency planning has addressed interference of nearby APs.

For broadcast service in wireless mesh networks, an association control based solution was proposed in [19], where minimum cost based greedy selection of an access point was shown to decrease the size of the broadcast tree. The authors also proposed the concept of multi-association, where the AP for unicast traffic and the AP for broadcast traffic are independently chosen by exploiting multiple coverages that are typical in mesh networks.

3. Preliminaries

In this section, we present the network model and outline the problems that are addressed in this paper.

3.1. Network model

We consider a WLAN with a set of users U and a set of access points A. The cardinality of U is denoted by n and the cardinality of A is denoted by m.

We use π_a to denote the maximum load AP $a \in A$ is allowed for multicast traffic. This parameter is used to limit the resource allocation for multicast traffic and provide protection to unicast traffic. Note that $0 \leqslant \pi_a \leqslant 1$. For every user $u \in U$, let $N_u = \{a : a \in A \text{ and } a \text{ is an AP } \textit{that is in communication range of user } u\}$.

We refer to an AP within the communication range of a user as a neighbor AP. We define variables $x_{u,a}$ such that $x_{u,a} = 1$ if u is associated with a; otherwise, $x_{u,a} = 0$. The maximum possible data rate on a link from an AP a to a user u is denoted by $r_{a,u}$. The set of discrete values that $r_{a,u}$ can take is denoted by R, where $|R| = \delta$.

$r_{a,u} = 0$ if u is out of the transmission range of a. We use S to denote the set of multicast sessions, and $|S| = l$. Let b_s be the data rate of multicast session $s \in S$. We use $s_u \in S$ to denote the multicast session user u subscribes to.

If an AP transmits multicast packets to its associated users, it uses the lowest rate among these users' maximum possible data rates on the links to this AP, to ensure reception by all users. We assume that MAC layer multicast/broadcast can support multi-rate transmission[3] although in the current IEEE 802.11 MAC layer standard, broadcast packets are always transmitted at the basic data rate. However, if the basic data rate is always used for multicast/broadcast, the problems MNU, BLA and MLA are still NP-hard because our NP-hardness proofs for these problems do not require multi-rate transmission, and the proposed algorithms are still applicable.

Each user node and each AP has a single radio. We assume that the radio channels of the neighboring APs are configured such that they do not interfere. Although IEEE 802.11b/g only has three non-overlapping channels, the IEEE 802.11a standard operates in the 5 GHz spectrum that supports 12 non-overlapping channels in US/Canada. The APs are connected using a wired LAN to one or more gateways that provide connectivity to the Internet. We assume that users are *quasi-static*, which means that they often tend to stay at one place for a relatively long time period before changing their location. This assumption is supported by recent studies of user mobility patterns in deployed WLANs [21,22].

Each user may request one multicast stream from the WLAN.

A user requesting a multicast stream is referred to as a multicast user, and a user requesting unicast service is referred to as a unicast user.

If a user can only be a unicast user or a multicast user, we do not need to do any other modification to the 802.11 standard except for replacing the association algorithm for multicast users.

If a user can be both a unicast as well as a multicast user, the network framework described in [19] can be applied, where the APs are synchronized through a time-synchronization protocol and each user independently selects one AP for unicast and another one for multicast services.

3.2. Problem statement

This paper focuses on algorithms for selecting the optimal AP for multicasting. We study three different objective functions and propose approximation algorithms for solving them. The suitability of the objectives depends on the user demands, user distribution, and the network providers revenue function. We first define the term *multicast load* that is used by some of the objective functions.

Definition 1. Multicast load: the multicast load of an AP is the fraction of time that the AP is busy in transmitting multicast flows; the normalized multicast load of a network is the average multicast load of an AP.

[3] Recent research [20] has implemented the multi-rate broadcast/multicast.

Maximizing number of users (MNU): When there is a heavy demand for multicast flows, all the multicast users' requests can not be met. For such scenarios, we define the goal to be maximization of the number of users that get multicast service from the network.

An example revenue model is when the unicast services have a monthly charge, but the multicast services are charged based on the time for which multicast streams are served to users.

Under this model, increasing the number of satisfied multicast users will increase the total revenue for the service provider.

Balancing load among APs (BLA): In order to reduce the impact of multicast services on unicast flows, it is critical to reduce the length of the multicast period. This can be achieved by balancing the multicast load of the APs. Specifically, the objective here is to minimize the maximum multicast load among all APs.

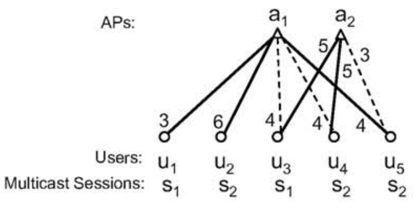
Consider a revenue model where one multicast flow is included in the basic monthly charges. Assume that the revenue function for unicast flows is concave, i.e., marginally decreasing with increasing bandwidth. Concave revenue functions are well known for achieving fairness among flows [23]. Then balancing the multicast load will typically lead to fairness among the unicast flows, and therefore a higher total revenue.

Minimizing load of APs (MLA): In order to free up the maximum amount of total time for unicast services, the total multicast load needs to be minimized. Although this can lead to an uneven distribution of the multicast load, for some revenue models this may be of interest.

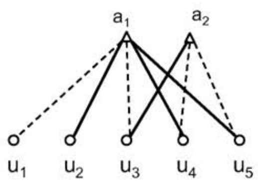
Consider a revenue model where one multicast flow is included in the basic monthly charges. However, the unicast services are charged per byte. In scenarios where there is a high demand for unicast traffic, maximizing the total amount of unicast traffic will maximize the revenue, while satisfying the multicast users.

An example: We use the example scenario shown in Fig. 1 to describe these three problems. The WLAN consists of two APs, a_1, a_2, and, 5 users, $u_1, u_2, \ldots, u_5$. Suppose that all traffic is multicast.

The maximum data rates to the five users u_1, u_2, u_3, u_4 and u_5 from a_1 are $3, 6, 4, 4$, and 4 Mbps, respectively. AP a_2 can communicate only with users u_3, u_4, and u_5, and



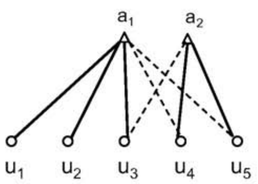
RSSI based association



MNU: optimal solution

Incoming Stream Rate for s_1: 3 Mbps, s_2: 3 Mbps

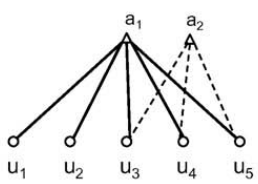
Algorithm	Satisfied Users
RSSI a_1: s_1: --, s_2: 4 a_2: s_1: 5, s_2: --	u_2, u_3, u_5
MNU a_1: s_1: --, s_2: 4 a_2: s_1: 5, s_2: --	u_2, u_3, u_4, u_5



BLA: optimal solution

Incoming Stream Rate for s_1: 1 Mbps, s_2: 1 Mbps

Algorithm	Load (a1, s1)	Load (a1, s2)	Load a1	Load (a2, s1)	Load (a2, s2)	Load a2	Total Load
RSSI a_1: s_1: 3, s_2: 4 a_2: s_1: 5, s_2: 5	1/3 (=0.33)	1/4 (=0.25)	**7/12 (=0.58)**	1/5 (=0.20)	1/5 (=0.20)	**2/5 (=0.40)**	**59/60 (=0.98)**
BLA a_1: s_1: 3, s_2: 6 a_2: s_1: --, s_2: 3	1/3 (=0.33)	1/6	**1/2 (=0.50)**	--	1/3 (=0.33)	**1/3 (=0.33)**	**5/6 (=0.83)**
MLA a_1: s_1: 3, s_2: 4 a_2: s_1: --, s_2: --	1/3 (=0.33)	1/4 (=0.25)	**7/12 (=0.58)**	--	--	**--**	**7/12 (=0.58)**



MLA: optimal solution

Fig. 1. An example network scenario: users u_1 and u_3 request for multicast session s_1, and users u_2, u_4, u_5 request for multicast session s_2. The dark lines between the users and APs represent the associations. The dotted lines represent the remaining links.

the maximum data rates are 5, 5, and 3 Mbps, respectively. Suppose that users u_1 and u_3 request multicast sessions s_1, and users u_2, u_4, and u_5 request for multicast session s_2.

If the multicast data rates of s_1 and s_2 are both 3 Mbps, this WLAN cannot support all users because u_1 and u_2 can only be associated with a_1, and a_1 cannot provide multicast service to u_1 and u_2 simultaneously. The reason is that if both u_1 and u_2 are supported by a_1 then the total load on a_1 will be $\frac{3}{3}+\frac{3}{6}>1$, which is infeasible. In case of RSSI-based association, users u_2 and u_5 can receive s_2 transmitted at 4 Mbps from a_1, and user u_3 can receive s_1 transmitted at 5 Mbps from a_2. In such scenarios, the objective of maximization of number of users (MNU) is relevant. One of the optimal solutions is that u_2, u_4, and u_5 are associated with a_1 and u_3 is associated with a_2. Thus, four users can be supported instead of three with the RSSI-based association.

Suppose the data rate of s_1 and s_2 are both 1 Mbps and the objective is to balance the multicast load among APs (BLA) by minimizing the maximum multicast load among the APs. In the optimal solution u_1, u_2, and u_3 are associated with a_1, and u_4 and u_5 are associated with a_2. The load of a_1 will thus be $\frac{1}{3}+\frac{1}{6}=\frac{1}{2}$ and the load of a_2 will be $\frac{1}{3}$.

Suppose the data rate of s_1 and s_2 is 1 Mbps and the objective is to minimize the total load of all APs for multicast streams (MLA). In the optimal solution all users are associated with a_1, which results in a total AP load of $\frac{1}{3}+\frac{1}{4}=\frac{7}{12}$.

3.3. Solution framework

In this section, we present the framework for the centralized and the distributed solutions. Our centralized solutions are based on reductions of the three problems to variants of the set cover problem. We present definitions of two problems from [24] that are used in the reductions:

Definition 2. Maximum coverage with group budgets (MCG) – cost version: There are k subsets $S_1, S_2, \ldots, S_k$ of a ground set X. There are l sets $G_1, G_2, \ldots, G_l$, each G_i being a subset of $\{S_1, \ldots, S_k\}$. Each G_i is a group and the groups are disjoint from each other.

A cost $c(S_j)$ is associated with each set S_j. Further, each group G_i is given a budget B_i and the overall budget is B. The objective is to find a subset H of $\{S_1, \ldots, S_k\}$ to maximize the size of the union of sets in H under the limitation that the total cost of the sets in H is at most B, and for any group G_i, the total cost of the sets in $H \bigcap G_i$ is at most B_i.

Definition 3. Set cover with group budgets (SCG): There is a set $S = \{S_1, S_2, \ldots, S_k\}$ of subsets of a ground set X. The set S is partitioned into groups $G_1, G_2, \ldots, G_l$. A cost $c(S_j)$ is associated with each set S_j. The objective is to find a subset H of S such that all elements of X are covered by sets in H and $max_{i=1}^{l} c(H \bigcap G_i)$ is minimized.

For solving the three problems we use reductions to either MCG or SCG. The main steps of the construction are outlined below. The set of all users becomes the ground set X in the instance of MCG or SCG. Corresponding to each AP, we create $|R| \times |S|$ subsets of X in MCG, where $|R|$ is the number of discrete transmission rates that the WLAN supports and S is the set of multicast sessions. Thus, each subset represents a unique 3-tuple of an AP, a transmission rate, and a multicast session. The cost of a subset is the ratio of the corresponding multicast session's data rate and the transmission rate.

All such subsets that are related to AP a_i form the group G_i. The budget B_i for the group G_i is the fraction of the time AP a_i is allowed to spend on multicast transmissions (π_i).

For our problem, there is no overall budget limitation for the whole network, i.e., $B = \infty$, as we assume that the capacity of the wired network is not the bottleneck.

For the distributed solutions, each user periodically sends a query message to each of its neighboring APs. As the APs may be in different channels, the user may need to query on all channels. However, this step can be significantly optimized by maintaining a cache of channels or by learning about the candidate channels in the vicinity from any AP. Each AP responds with a message containing information about the ongoing multicast sessions and the corresponding data rate of such transmissions. The user also learns the maximum data rate for the link from the neighboring APs to itself. If a user is currently associated with an AP a, this user also needs to know the resulting load of a if it leaves AP a. Based on all this information, each user determines the AP to associate with. Details of how this information is used is provided in the corresponding sections that deal with the three problems.

The communication overhead on the wireless channel for the distributed solution is dependent on the number of neighboring APs for a user and the number of channels (say c). In the worst case each round of query from a user will incur mc messages, thus resulting in nmc messages considering all users. The periodicity of querying will impact the time to respond to mobility and changing channel conditions. For the centralized solutions, the same information has to be transmitted to a central network element such as a gateway. Therefore, the complexity of messages on the wireless channel will be of the same order. However, for centralized solutions there is additional overhead of messages on the wired network and additional delay due to collection, processing and dissemination of the solution at the gateway.

For our solutions to work in a real deployment, users need to cooperate to reach the common goal on the network's side. In order to enforce such behavior, the network can detect selfish users and penalize them either by reducing their quality of service or by dropping their requests. As the centralized solutions are implemented by a network entity such as the gateway, the APs can be made aware of which users will be associated with which APs to help detect selfish users. For the distributed solutions, the APs can mimic the computation of the users based on channel quality information obtained from nearby users and APs. User associations that are repeatedly deviating from the computations of the APs can used to identify selfish users.

4. Maximizing number of users (MNU)

In this section, we formulate the problem MNU, show its NP-hardness, and present centralized and distributed algorithms. We present an ILP (integer linear programming) for MNU to formalize the problem definition.

$$\text{(Problem MNU) maximize} \sum_{u \in U} \sum_{a \in A} x_{u,a} \tag{1}$$

$$\text{subject to} \quad \rho_{a,s_u} \geqslant \frac{b_{s_u}}{r_{u,a}} x_{u,a}, \quad \forall u \in U,\ \forall a \in N_u \tag{2}$$

$$\sum_{s \in S} \rho_{a,s} \leqslant \pi_a, \quad \forall a \in A \tag{3}$$

$$\sum_{a \in N_u} x_{u,a} \leqslant 1, \quad \forall u \in U \tag{4}$$

$$x_{u,a} = 0, \quad \forall u \in U,\ \forall a \notin N_u \tag{5}$$

$$x_{u,a} \in \{0,1\}, \quad \forall u \in U,\ \forall a \in A \tag{6}$$

$$\rho_{a,s} \geqslant 0, \quad \forall a \in A,\ \forall s \in S \tag{7}$$

Constraint 2 ensures that the load $\rho_{a,s}$ of a multicast session s on AP a is determined by the user u with the minimum rate $r_{u,a}$, among all users that are associated with AP a for multicast session s. Constraint 3 ensures that the total load of an AP a for multicasting is not more than its maximum allowable multicasting load. Constraint 4 forces each user to associate with at most one neighboring AP. Constraint 5 ensures that a user u can not be associated with a non-neighboring AP. The restrictions on the possible values for the variables $x_{u,a}$ and $\rho_{a,s}$, are in Constraints 6 and 7.

We show that MNU is an NP-hard problem, by showing a reduction from the decision version of bin packing problem (see Appendix B). As the decision version of the bin packing problem is NP-complete [25], MNU is NP-hard.

Note that MNU is trivially in *P*, if there is only one multicast session in a WLAN. For a single session, all APs can choose to transmit at the lowest possible rate so that the maximum allowable multicast load on each AP is not violated.

4.1. Centralized MNU

In order to solve this problem, we first construct an instance of MCG from a given instance of MNU, as outlined in Section 3.3. We illustrate the reduction through an example.

Example – MNU: If the data rate of s_1 and s_2 is 3 Mbps in the WLAN shown in Fig. 1, we can reduce the problem MNU for the WLAN to the problem MCG shown in Fig. 2. One of the optimal solutions for this problem MCG is $H = \{S_4, S_5\}$.

In [24], the authors presented a greedy algorithm for MCG as it is an NP-hard problem. Because there is no overall budget limitation for our problem, we adapt the algorithm in [24] and present the modified algorithm below.

The algorithm greedily picks up subsets with minimum cost for every additional element until either all elements have been covered or until each group's budget has been violated by the last selected subset for the group. In the pseudo-code presented in Fig. 3, H represents the set of selected subsets at any step. The set X' denotes the elements of X which have not yet been covered by the subsets in H. The statements from line 3 to line 14 are repeatedly executed until all the group budgets are exceeded or all elements of X get covered. The variable *flag* is used for this purpose in line 11. In the *for* loop, *Centralized MNU* finds a set S_j in every group G_i whose budget has not been exceeded, and S_j is the set which is the most cost-effective set in the group G_i, i.e., $\frac{|S_j \bigcap X'|}{c(S_j)} = \max_{D \in G_i} \frac{|D \bigcap X'|}{c(D)}$. Then in line 12, *Centralized MNU* finds the most cost-effective set in the sets selected in the *for* loop. This set is added into H and the elements in this set is removed from X' in line 13. Eventually, we get the output H.

Obviously, H does not obey the group budget requirements. We assume the cost of any single set S_j in any group

$X = \{u_1, u_2, u_3, u_4, u_5\}$
$S_1 = \{u_3\}$ $\quad c(S_1) = 3/4$
$S_2 = \{u_1, u_3\}$ $\quad c(S_2) = 3/3 = 1$
$S_3 = \{u_2\}$ $\quad c(S_3) = 3/6 = 1/2$
$S_4 = \{u_2, u_4, u_5\}$ $\quad c(S_4) = 3/4$
$G_1 = \{S_1, S_2, S_3, S_4\}$ $\quad B_1 = 1$
$S_5 = \{u_3\}$ $\quad c(S_5) = 3/5$
$S_6 = \{u_4\}$ $\quad c(S_6) = 3/5$
$S_7 = \{u_4, u_5\}$ $\quad c(S_7) = 3/3 = 1$
$G_2 = \{S_5, S_6, S_7\}$ $\quad B_2 = 1$

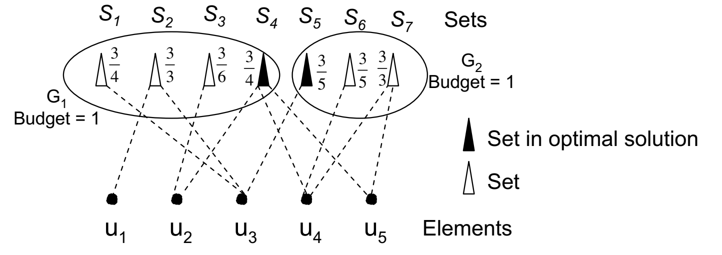


Fig. 2. The reduction from problem MNU for the WLAN in Fig. 1 to problem MCG. The data rate of s_1 and s_2 is 3 Mbps. One of the optimal solutions for this problem MCG is $H = \{S_4, S_5\}$.

Algorithm **Centralized MNU**

1. $H \leftarrow \phi, X' \leftarrow X.$
2. **repeat**
3. $\quad flag \leftarrow 0$
4. $\quad$ **for** $i = 1, 2, \dots, n$ **do**
5. $\quad\quad$ **if** $c(H \bigcap G_i) < B_i$ **then**
6. $\quad\quad\quad k \leftarrow argmax_j \frac{|S_j \bigcap X'|}{c(S_j)}\ (S_j \in G_i)$
7. $\quad\quad\quad A_i \leftarrow S_k$
8. $\quad\quad\quad flag \leftarrow 1$
9. $\quad\quad$ **else** $A_i \leftarrow \phi$
10. $\quad$ **endfor**
11. $\quad$ **if** $flag = 0$ **then Break**
12. $\quad r \leftarrow argmax_i \frac{|A_i \bigcap X'|}{c(A_i)}$
13. $\quad H \leftarrow H \bigcup A_r,\ X' \leftarrow X' - A_r$
14. $\quad$ **if** $X' = \phi$ then **break**
15. **endrepeat**
16. $H_2 \leftarrow$ sets in H that violated some group budget when added
17. $H_1 \leftarrow H - H_2$
18. $H_3 \leftarrow H_1$ or H_2 whichever covers most elements
19. **output** H_3

Fig. 3. Centralized solution for MNU.

G_i is not more than the budget of G_i. We partition H into two subsets H_1 and H_2. H_2 contains those sets S_j which when added to H caused the budget of some group G_i to be violated. $H_1 = H - H_2$. Observe that H_1 and H_2 by themselves do not violate the budget constraints and one of these two sets must be covering at least 1/2 the number of elements covered by H. Out of H_1 and H_2, we select the one which covers the most number of elements. The final solution directly maps to the solution to the problem MNU. This algorithm has an approximation factor of 8, as shown in Appendix A.

The complexity of the algorithm is polynomial. The number of sets is $nm\delta$. The *repeat* loop will be executed at most n times, as at least one user is covered in each iteration. The *for* loop is executed m times. The argmax operation in step 6 takes $nml\delta$ time. Thus, the total time complexity of the algorithm is $O(n^2m^2\delta l)$.

Example – Centralized MNU: we run *Centralized MNU* algorithm on the problem MCG shown in Fig. 2. S_4 is selected in the first round because it has the maximum value of $\frac{|S_4 \bigcap X'|}{c(S_4)} = \frac{3}{3/4} = 4$ among all $S_i (1 \leqslant i \leqslant 7)$. After that, $H = \{S_4\}$, $X' = \{u_1, u_3\}$. In the second round, S_2 is selected because it has the maximum value of $\frac{|S_2 \bigcap X'|}{c(S_2)} = \frac{2}{1} = 2$ and $c(H \bigcap G_1) = c(S_4) = 3/4 < B_1 = 1$. After that, we get output $H = \{S_2, S_4\}$ because $X' = \phi$. Now, $c(H \bigcap G_1) = c(S_2) + c(S_4) = 7/4 > B_1 = 1$. We divide H into $H_1 = \{S_4\}$ and $H_2 = \{S_2\}$. Eventually, we get output H_1 because H_1 cover more elements than H_2. Therefore, u_2, u_4, u_5 are associated with a_1 and 3 users get multicast streams.

4.2. Distributed MNU

We present a distributed algorithm to maximize the number of users.

As the total load of the APs for multicast is fixed, every user should increase the total load minimally in order to attempt increasing the total number of users. Due to the lack of global view, the distributed approach has to take decisions based only on local information obtained from the APs.

Using the framework presented in Section 3.3, each user computes the total load of its neighboring APs if it associates with it without violating the maximum multicast load for that AP. The user then associates with the neighboring AP that results in the least increase in the total load (without considering this user), and ties are broken based on RSSI. Observe that if a user changes its association, the total load of the neighboring APs will go down.

Example – Distributed MNU: consider that the data rate of s_1 and s_2 is 3 Mbps in the WLAN in Fig. 1, and users use the distributed algorithm in the order u_1, u_2, u_3, u_4, u_5. First u_1 associates with a_1. Then, u_2 can not associate with a_1 because of the load limitation of a_1. After that, u_3 associates with a_1, which results in the minimum total load 1 of

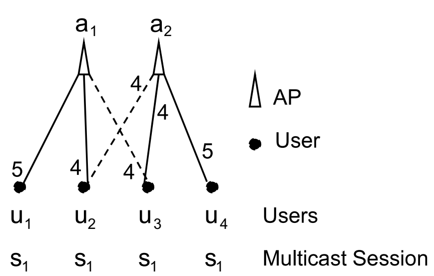


Fig. 4. Negative example of convergence for simultaneous local decisions: all users request the same multicast session. Users u_1 and u_2 are associated with a_1, and u_3 and u_4 are associated with a_2. u_2 and u_3 always make their local decisions simultaneously. The initial associations are shown using solid lines.

u_3's neighboring APs a_1, a_2. Similarly, u_4 and u_5 are associated with a_2. Eventually, four out of the five users receive their multicast service.

The algorithm *Distributed MNU* converges when the network becomes static, if at any time only one user computes its association and the resulting state is updated in all nodes before the next user computes its association. In a static network, a user's change in association will result in a reduction in total load of its neighboring APs, and therefore the total network load. As there are finite number of APs, data rates, and users, and the total network load is not increasing in each step, the solution will eventually converge.

However, if the users in an AP's transmission range make their local decisions simultaneously, the algorithm *Distributed MNU* may not converge. The example scenario is shown in Fig. 4. AP a_1 can communicate with u_1, u_2 and u_3 with the rates 5, 4 and 4 Mbps, respectively; AP a_2 can communicate with u_2, u_3 and u_5 with the rates 4, 4 and 5 Mbps, respectively. Users u_1 and u_2 are associated with a_1, and u_3 and u_4 are associated with a_2. All users request the same multicast session s_1 with the rate 1 Mbps. So, the current total load of a_1 and a_2 is $\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$. Now, u_2 and u_3 make the local decision simultaneously. If only u_2 changes its association and it associates with a_2, the total load of a_1 and a_2 is reduced to $\frac{1}{5}+\frac{1}{4}=\frac{9}{20}$. If only u_3 changes its association and it associates with a_1, the total load of a_1 and a_2 is reduced to $\frac{1}{4}+\frac{1}{5}=\frac{9}{20}$. Therefore, both u_2 and u_3 will change their associations, which actually does not change the total load. Next, if u_2 and u_3 make local decisions simultaneously again, u_2 and u_3 will be associated with a_1 and a_2 respectively, again. Therefore, the algorithm does not converge.

5. Balancing load among APs (BLA)

In this section, we prove the NP-hardness of the BLA problem and present centralized and distributed algorithms. The objective is to minimize the maximum load among the APs.

$X = \{u_1, u_2, u_3, u_4, u_5\}$
$S_1 = \{u_3\}$ $\quad c(S_1) = 1/4$
$S_2 = \{u_1, u_3\}$ $\quad c(S_2) = 1/3$
$S_3 = \{u_2\}$ $\quad c(S_3) = 1/6$
$S_4 = \{u_2, u_4, u_5\}$ $\quad c(S_4) = 1/4$
$G_1 = \{S_1, S_2, S_3, S_4\}$
$S_5 = \{u_3\}$ $\quad c(S_5) = 1/5$
$S_6 = \{u_4\}$ $\quad c(S_6) = 1/5$
$S_7 = \{u_4, u_5\}$ $\quad c(S_7) = 1/3$
$G_2 = \{S_5, S_6, S_7\}$

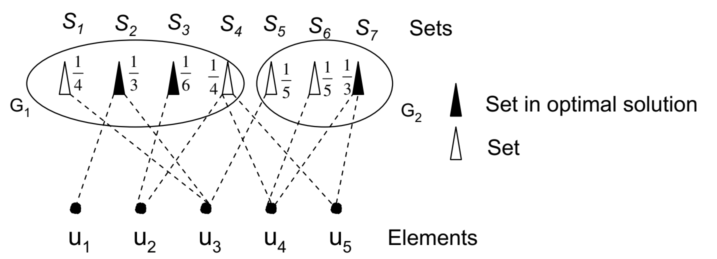


Fig. 5. The reduction from problem BLA for the WLAN in Fig. 1 to problem SCG. The data rates of s_1 and s_2 are both 1 Mbps. The optimal solution of this SCG problem is $H = \{S_2, S_3, S_7\}$.

Algorithm **Centralized BLA**

(1) Guess the optimal value B^* and assume there is an optimal cover H^* such that $max_{i=1}^{l} c(H^* \bigcap G_i) \leq B^*$

(2) Create an instance of MCG by having a budget of B^* on each group G_i. Run *Centralized MNU*, which covers at least $\frac{1}{8}$th the elements in X. Remove the covered elements from X and all subsets of X, and run *Centralized MNU* again. Iterating *Centralized MNU* $log_{8/7}n+1$ times results in a solution that covers all elements. Let $H = \{$*all sets S_i added when repeating Centralized MNU*$\}$.

(3) Output H.

Fig. 6. Algorithm Centralized BLA.

We present a reduction from Minimum Makespan Scheduling problem [26] to the BLA problem, which is described in the Appendix B, to prove the NP-hardness of BLA.

Based on the reduction, BLA is NP-hard because the Minimum Makespan scheduling problem is NP-hard. Note that BLA is in *P* if there is only one multicast session. As there are constant number of discrete transmission rates, each of these transmission rates can be checked in sequence for feasibility of being the maximum transmission rate. For a given value of the transmission rate, all APs are assigned the same rate (as the optimization function only concerns the maximum). Among all the transmission rates the highest rate (when assigned to all APs) that provides service to all users, is the solution.

The ILP for BLA is similar to the ILP for MNU. The objective function and the Constraints 4 and 5 are different, which are described below:

$$\text{(Problem BLA) minimize} \quad Z \tag{8}$$

$$\sum_{a \in N_u} x_{u,a} = 1, \quad \forall u \in U \tag{9}$$

$$\sum_{s \in S} \rho_{a,s} \leqslant Z, \quad \forall a \in A \tag{10}$$

$$(3)\ (6)\ \text{and}\ (7) \tag{11}$$

Constraint 9 ensures that every user u is associated with exactly one of the neighboring APs. Constraint 10 forces Z to be higher than the load of all APs. The objective function minimizes Z which leads to minimization of the maximum load.

5.1. Centralized BLA

We solve BLA by reducing it to an instance of the SCG problem using the construction presented earlier in Section 3.3.

Example – BLA: If the data rate of s_1 and s_2 is 1 Mbps in the WLAN in Fig. 1, we can reduce the problem BLA for the WLAN to problem SCG shown in Fig. 5. The optimal solution of this SCG problem is $H = \{S_2, S_3, S_7\}$.

SCG is also NP-hard. In [24], the authors gave an algorithm for the cardinality version of SCG based on the greedy algorithm for MCG. Our algorithm is similar (Fig. 6). It has an approximation factor of $(\log_{8/7} n + 1)$, as shown in the Appendix A.

To implement the algorithm *Centralized BLA*, there is an issue of how to guess B^*.

Let the maximum cost among all subsets of X in all groups be $c_{\max}$. B^* also should be less than 1. Therefore, we can try several (a constant number) values of B^* between $c_{\max}$ and 1 to get the best result.

The algorithm requires $\log_{8/7} n + 1$ executions of the solution for MNU. So the complexity is $O(n^2 m^2 l \delta \log(n))$.

Example – Centralized BLA: We run *Centralized BLA* algorithm on the problem SCG shown in Fig. 5. Let $B^* = 1/2$, and create an instance of problem MCG. Then run *Centralized MNU*, and get the output $\{S_4\}$. After that, remove u_2, u_4, u_5 from every $S_i (1 \leqslant i \leqslant 7)$ and create a new instance of MCG. Run *Centralized MNU* again, and get output $\{S_2\}$. Therefore, all users are associated with a_1.

5.2. Distributed BLA

As the objective is to balance the load among APs, a user should attempt to minimize the maximum load of the neighboring APs. The following is the distributed algorithm for BLA.

Using the distributed solution's framework presented in Section 3.3, each user calculates the resulting sequence of loads of neighboring APs for each choice of association with a neighboring AP. Each load sequence is sorted in

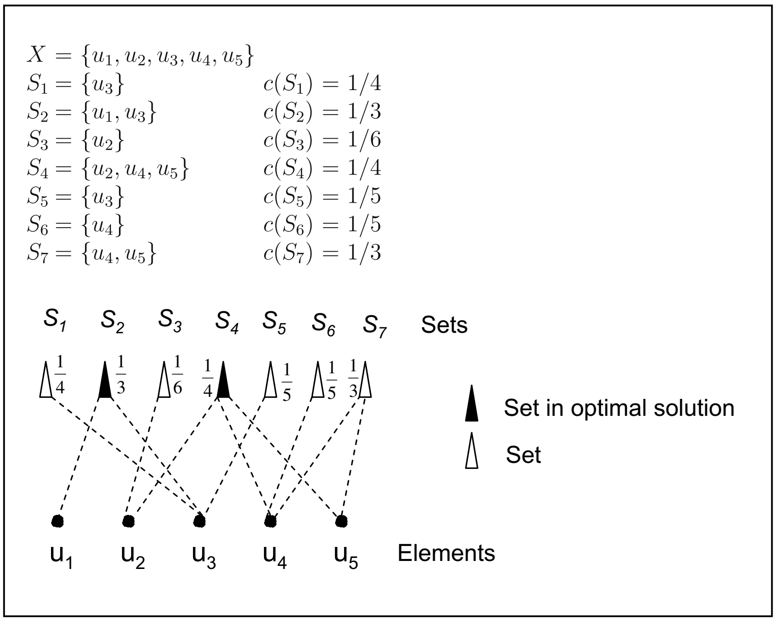


Fig. 7. The reduction from MLA problem for the WLAN in Fig. 1 to the set cover problem. The data rate of s_1 and s_2 is 1 Mbps. The optimal solution of this set cover problem is $H = \{S_2, S_4\}$.

non-increasing order, and the user joins the AP with the lexicographically smallest sorted load sequence.

Example – Distributed BLA: Assume that the data rates of s_1 and s_2 are both 1 Mbps in the WLAN in Fig. 1, and users run the distributed algorithm in the order u_1, u_2, u_3, u_4, u_5. First u_1 and u_2 are associated with a_1. After that, u_3 makes the decision. If u_3 is associated with a_1, it's neighboring APs' load vector in non-increasing order is (1/2, 0); if u_3 is associated with a_2, the load vector is (1/2, 1/5). Therefore, u_3 is associated with a_1. Next, if u_4 is associated with a_1, its neighboring APs' load vector with non-increasing order is (7/12, 0); if u_4 is associated with a_2, the load vector is $(1/2, 1/5)$. Hence, u_4 is associated with a_2. Similarly, u_5 is associated with a_2. Eventually, the load of a_1 is $1/2$ and the load of a_2 is $1/3$, which is also the optimal solution.

The distributed algorithm for BLA converges when the network is static if in an AP's transmission range, only one user runs the algorithm at a given time and all information is updated before the next user computes its association. This follows from the observation that each action by a user reduces the vector of neighboring APs' loads and therefore the global vector.

However, if the users in an AP's transmission range make their local decisions simultaneously, the distributed algorithm for BLA may not converge. The example scenario is same as the scenario for the distributed algorithm for MNU shown in Fig. 4.

6. Minimizing the load of APs (MLA)

In this section, we prove the NP-hardness of MLA and describe our centralized and distributed algorithms. The objective of MLA is to reduce the total network load. The constraints of the ILP for MLA are same to the constraints of the ILP-BLA except that Inequality 10 is not needed.

The MLA problem is formally defined as follows:

$$\text{(Problem MLA) minimize} \quad \sum_{a \in A} \sum_{s \in S} \rho_{a,s} \tag{12}$$

$$\sum_{a \in N_u} x_{u,a} = 1, \quad \forall u \in U \tag{13}$$

$$(2)\ (3)\ (6) \text{ and } (7) \tag{14}$$

We show that MLA is an NP-hard problem, by showing a reduction from the Set Cover problem, which is described in the Appendix B. MLA is NP-hard as set cover problem is NP-hard [25].

6.1. Centralized MLA

In order to solve MLA, we reduce it to the set cover problem using the construction presented in Section 3.3 except that there are no groups since we are concerned with the total multicast load of a network. The load on each AP is not of concern.

Example – MLA: If the data rates for s_1 and s_2 are 1 Mbps in the WLAN in Fig. 1, we can reduce problem MLA for the WLAN to the set cover problem shown in Fig. 7. The optimal solution of this set cover problem is $H = \{S_2, S_4\}$.

We use the greedy solution [26] to the set cover problem in our simulations. However, it should be mentioned that the layer algorithm, which is bounded by a constant, can also be used if for any user the number of APs that it can associate with is bounded by a constant [26]. The cost version of the greedy set cover algorithm is shown in Fig. 8, which can be directly used to solve MLA after reducing it to an instance of the set cover problem. As the number of sets is $m\delta l$, the time complexity of the algorithm is $O(nm\delta l)$.

Example – Centralized MLA: We run *CostSC* algorithm on the set cover problem corresponding to Fig. 7. S_4 is selected in the first round because it has the maximum value of $\frac{|S_4 \bigcap X'|}{c(S_4)} = \frac{3}{1/4} = 12$ among all $S_i (1 \leqslant i \leqslant 7)$. After that, $H = \{S_4\}$, $X' = \{u_1, u_3\}$. In the second round, S_2 is selected because it has the maximum value of $\frac{|S_2 \bigcap X'|}{c(S_2)} = \frac{2}{1/3} = 6$. After that, we get output $H = \{S_2, S_4\}$ because $X' = \phi$. Therefore, all users are associated with AP a_1, which is also the optimal solution. The algorithm *CostSC* is a $(\ln n + 1)$-approximation algorithm for the set cover problem as proven in [26] and stated in the theorem in Appendix A.

6.2. Distributed MLA

Because the objective of MLA is to minimize the total load of the APs in the network, intuitively, a user should be associated with the AP which increases the total load minimally. Therefore, we use the same distributed algorithm for MLA as the one for MNU.

Example – Distributed MLA: Consider that the data rate of s_1 and s_2 is 1 Mbps in the WLAN in Fig. 1, and users use the distributed algorithm in the order u_1, u_2, u_3, u_4, u_5. First u_1, u_2 is associated with a_1. After that, u_3 is associated with a_1 because the total load of u_3's neighboring APs a_1 and a_2 is $\frac{1}{3} + \frac{1}{6} = \frac{1}{2}$ if u_3 is associated with a_1 and the total load is $\frac{1}{3} + \frac{1}{6} + \frac{1}{5} = \frac{7}{10}$ if u_3 is associated with a_2. Similarly, u_4, u_5

Algorithm **CostSC**

1. $H \leftarrow \phi, X' \leftarrow X$.
2. **while** $X' \neq \phi$ do
3. $\quad A \leftarrow S_i$ s.t. $\frac{|S_i \bigcap X'|}{c(S_i)} = max_{D \in S} \frac{|D \bigcap X'|}{c(D)}$
4. $\quad H \leftarrow H \bigcup A, X' \leftarrow X' - A$
5. **endwhile**
6. **output** H

Fig. 8. Algorithm CostSC [26].

are associated with a_1. Eventually, all users are associated with AP a_1, which is also the optimal solution.

7. Performance evaluation

In this section we report on performance studies of the proposed association algorithms for multicast using simulations in the Network Simulator ns2 [27].[4] The simulation results show the average performance of our algorithms, while our analysis in the previous sections only shows the performance of our algorithms in the worst cases. We compare the performance of the three algorithms, MLA, BLA, and MNU, with the RSSI-based association algorithm (RSSI). We have simulated the proposed algorithms over 1.2 km^2 area with up to 200 APs and 400 users randomly located in the area. The radio propagation range of both AP and user is 200 m. The transmission rates and their distance thresholds are shown in Table 1. The users collect information of neighbor APs using active scanning [28]. Every user joins one multicast session. The APs operate in IEEE 802.11a infrastructure mode. We use 0.9 as the load limitation of multicast for every AP. Unless otherwise specified, we use five multicast sessions. Each user selects one of the multicast sessions at random. We assume nearby APs use different channels and there is no interference among their transmission with users. These simulation settings are used for all algorithms unless mentioned otherwise. We depict the average of 40 random scenarios in the figures, along with the 95% confidence interval. In order to clearly depict the graphs with the confidence intervals, we slightly offset the curves for our algorithms (other than RSSI) on the x-axis. The x-axis values shown for the RSSI curves are the true x-axis values.

Minimizing load of APs: Fig. 9 shows the normalized load with respect to the number of users, APs, and sessions, respectively. Figs. 9a and c show that as the number of users and the number of sessions increase, the total AP load increases, because of increased demand for multicast traffic. In Fig. 9a we observe that for higher number of users, the improvement in normalized AP load is higher. In case of RSSI, with increasing number of users, a large number of APs end up transmitting a large number of sessions at low data rates as the multicast transmission rate caters to the associated user with the lowest data rate.

The normalized multicast load for the centralized MLA performs 25% better than that of RSSI for 400 users (Fig. 9a). For low user-to-AP ratio (50 users, 200 APs), our algorithms perform similar to RSSI, as there are few opportunities for taking advantage of simultaneous reception of multicast packets by multiple users. The normalized AP load, however, has an inverse relationship to the number of APs as shown in Fig. 9b. The reason is that the resulting increased density of APs allows for higher transmission rate between APs and users.

The distributed algorithm performs similar to the centralized algorithm.

Due to poor frequency planning, high density of deployment, or scarcity of available channels, interference from nearby APs may reduce the maximum possible load of an AP. For example, in dense scenarios with as many as 200 APs, an AP may observe a large number of interfering APs in its vicinity which will lead to interference. One way to address interference is to modify the definition of "load" to include the fractional time that an AP's channel is busy due to interference from nearby APs. However, this model does not differentiate between interference at a user location and at the AP. Another approach is to consider the joint optimization problem of frequency planning and load optimization. Extending our work while considering interference is left as future work. In a related work [8], a different WLAN model that considers only binary interference between APs is considered. However, in that work a different objective of determining a schedule for transmissions from APs to multicast users is studied. For simplicity, in this evaluation we assume that there is no interference from nearby APs, or alternately there are an unlimited number of channels.

Balancing load among APs: Fig. 10 shows the maximum load among APs with respect to the number of users, APs, and sessions, respectively. The centralized BLA algorithm has 26% (50 users) to 49% (400 users) lower maximum load than RSSI (Fig. 10a). Moreover, unlike the RSSI algorithm, for the distributed and centralized BLA algorithms, the maximum load increases slowly with the number of users and sessions (Figs. 10a and c). Fig. 10b shows that the maximum load decreases as the number of APs increases, since the multicast load can be shared by more APs. We observe that the centralized and distributed BLA algorithms have similar performance.

Maximizing number of users: Fig. 11 shows the number of satisfied users with respect to the multicast load limitation. As the multicast load limitation increases, the number of satisfied users increases as well.

The centralized algorithm increases the number of admitted users by 4% (load 0.16) to 40% (load 0.04). The distributed algorithm increases the number of admitted users by 4% (load 0.16) to 18% (load 0.08). For higher multicast load limit per AP, fewer users will be constrained and hence the scope for improvement by association control is reduced.

Overloaded hotspots: In Fig. 12, we evaluate the performance for busy hotspots such as airports and malls, where there are few APs but a large number of users. The user-to-AP ratio is of the order of 10–100. In such cases the reduction in normalized AP load and the maximum AP load is not significant compared to sparser scenarios shown in previous graphs. Due to the large number of users with varying data rates to different APs, and varying multicast demands, each AP needs to transmit a large number of sessions at low data rates, leaving little room for improvement.

Optimal solutions: In Fig. 13, we evaluate the optimality of MLA, BLA, and MNU algorithms by comparing them with

Table 1
Transmission rate vs. distance threshold [29]

Rate (Mbps)	6	12	18	24	36	48	54
Distance threshold (m)	200	145	105	85	60	40	35

[4] The simulation source code can be downloaded from http://www.cse.ohio-state.edu~chenai/ICDCS07/.

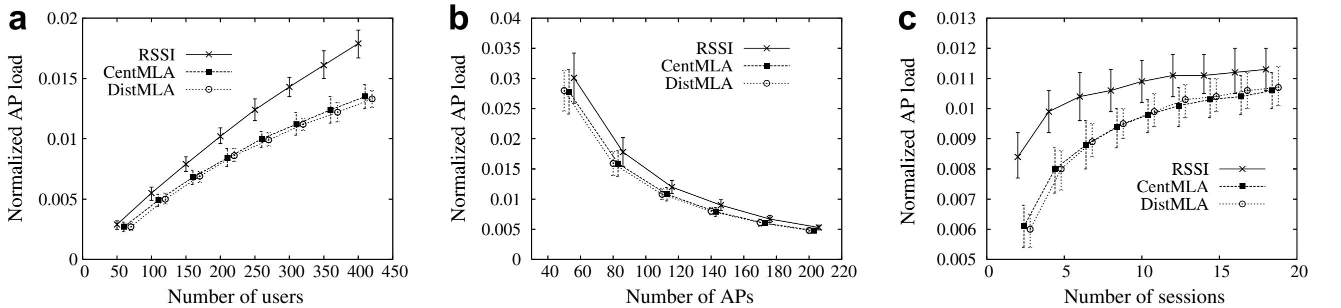


Fig. 9. Normalized AP load for multicast sessions. (a) Varies the number of users for 200 APs, (b) changes the number of APs with 100 users, and (c) changes the number of session with 200 APs and 200 users.

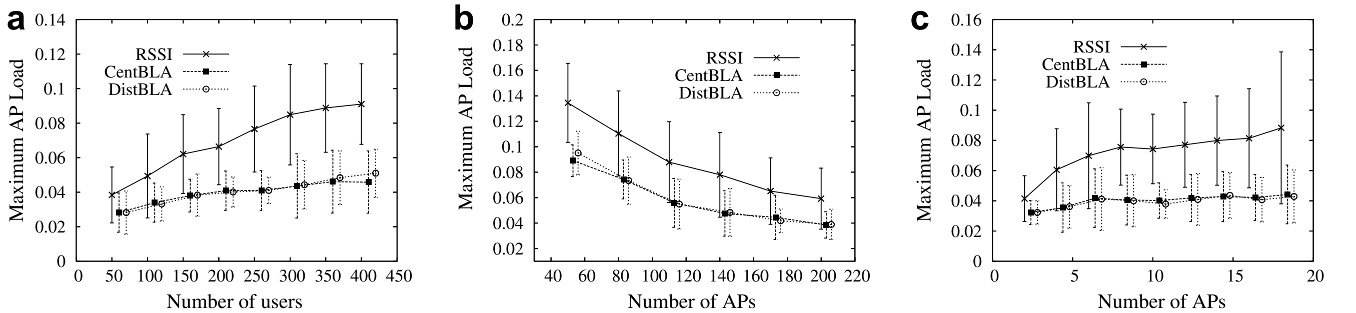


Fig. 10. Maximum load among APs for multicast sessions. (a) Varies the number of users for 200 APs, (b) varies the number of APs with 100 users, and (c) varies the number of sessions with 200 APs and 200 users.

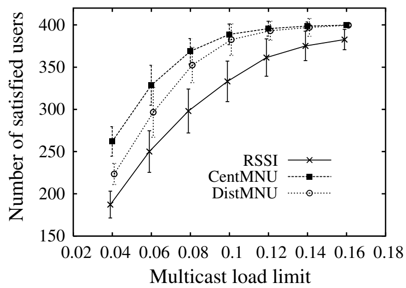


Fig. 11. The number of satisfied users with respect to multicast load limit (budget) with 400 users, 100 APs, and 18 multicast sessions.

the corresponding optimum solutions obtained by using an ILP solver.

Note that MLA, BLA, and MNU are NP-hard problems. The solver takes an excessive amount of time for finding the optimal solutions. So, we limit our evaluation only to small networks. We observe that our algorithms improve the performance under all scenarios studied when compared to RSSI. However, there is still a gap between the optimum algorithm and our algorithms.

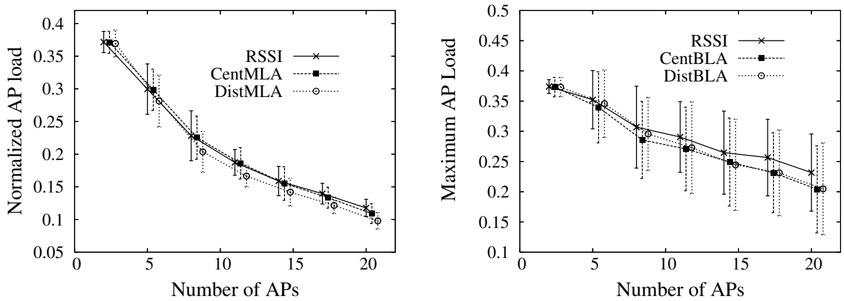
Although the centralized algorithms have global information, they are approximation algorithms. The distributed algorithms leverage the local information for making decisions. The comparable performance of the centralized and distributed algorithms in most scenarios studied implies the localized nature of the problem.

In summary, we observe that our centralized and distributed algorithms bring about the most improvement in performance for medium user-to-AP ratio (on the order of 1–2). For very low user-to-AP ratios (0.25–0.5) and for very high user-to-AP ratios (10–100), the improvement is not significant. In the former case the sparsity of users presents few opportunities for leveraging "multicasting" for conserving network resources. Whereas in the latter case, due to the large number of users with varying data rates to different APs, and varying multicast demands, each AP needs to transmit a large number of sessions at low data rates, leaving little room for improvement. The average performance is better than RSSI-based association under all scenarios studied. The centralized algorithm and the distributed algorithms have comparable performance, which indicates the "mostly local" nature of the problem. In small network scenarios, we observe that our algorithms are much closer to optimum than RSSI based association, although there is still room for improvement.

8. Discussions and future work

In this section, we outline some open issues that we are current investigating.

Distributed convergence: We have shown that in some cases, simultaneous association decisions by multiple nodes may not necessarily result in global optimization of the objective functions. We are currently working on local coordination mechanisms to guarantee optimization of the global objectives at each step. An idea that we are currently exploring uses explicit *locks* from neighboring APs before committing to a change in association. We are



(a) Normalized AP Load (b) Maximum AP Load

Fig. 12. Normalized AP load for fewer number of APs. Number of sessions = 5. Two hundred users are randomly located in 600 m^2 area. (a) Normalized AP load (MLA) (b) Maximum AP load (BLA).

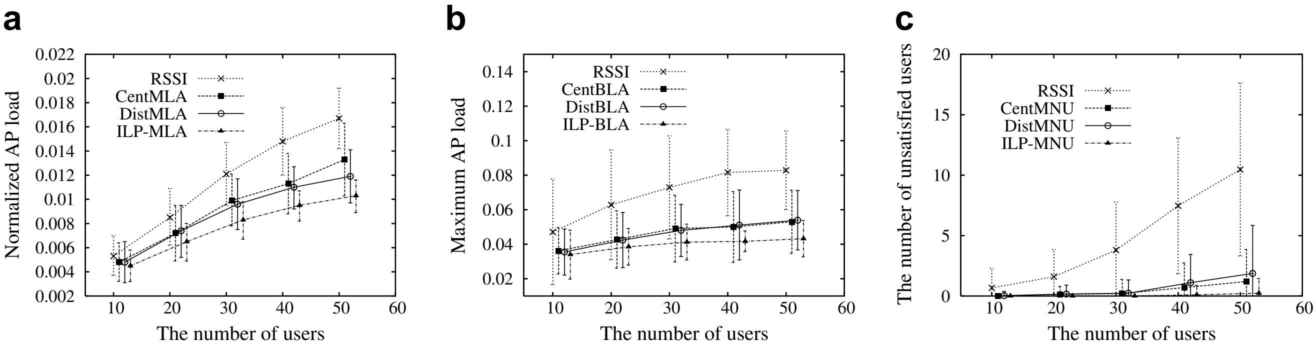


Fig. 13. The optimal solution of MLA, BLA, and MNU with respect to the number of users with 30 APs randomly located in 600 m^2 area: (a) total AP load, (b) maximum AP load among all APs, and (c) the number of unsatisfied users with multicast budget 0.042. Number of sessions is 3.

exploring issues such as deadlocks, communication overhead and delayed association for such approaches.

Adaptive power control: Adaptive power control provides an additional degree of flexibility that has not been explored in this work. We are currently working on approximation algorithms based on a generalized network model that allows nodes to choose from a finite set of discrete power levels.

Explicit interference modeling: The approximation algorithms need to be modified to explicitly account for interference from neighboring users and APs. In addition, along with such explicit interference models, it is necessary for the nodes to dynamically maintain a list of interfering nodes. To keep track of interfering sources, naive solutions proposed in the literature use explicit beaconing at high power levels, which adds extra overhead. We are currently investigating extensions to our work that use explicit interference modeling along with practical ways to keep track of interfering sources.

9. Conclusion

Multicast services must be deployed with minimal impact to existing unicast services in WLANs. The problem of enabling multicast based streaming services in large-scale WLANs has not received attention in the past. Motivated by recent reports of dense deployments of APs in WLANs, we study techniques for exploiting overlapping coverage from neighboring APs to optimize performance. Three objective functions motivated by different revenue functions and network scenarios are studied: maximizing the number of users (MNU), balancing the load among APs (BLA) and minimizing the load of APs (MLA). We show that these problems are NP-hard. We present centralized approximation algorithms and distributed algorithms for these problems. Using simulations we evaluate the performance of these protocols and find that compared with the RSSI-based approach, the number of users, the maximum load, and the normalized load can be improved when the user-to-AP ratio is of the order of 1–2. For very low user-to-AP ratios (0.25–0.5) and for very high user-to-AP ratios (10–100), the improvement is not significant. The centralized approximation algorithms and distributed algorithms have comparable performance in most scenarios studied. Due to the lower signaling complexity and comparable performance, the distributed algorithms are a clear choice for large networks.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants CNS-0546630 (CAREER Award), and CNS-0403342. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix A. The proofs of approximation factors

Theorem 9.1. *The approximation factors for the centralized algorithms are as follows: (a) The algorithm Centralized MNU is an 8-approximation algorithm for problem MNU with no total budget limitation. (b) The algorithm Centralized BLA is an* $(\log_{8/7} n + 1)$*-approximation algorithm for BLA problem. (c) The algorithm CostSC is an* $(\ln n + 1)$*-approximation algorithm for set cover problem.*

Proof. (a) Define $X(H)$ as the number of the elements covered by the subsets of X in H. Let OPT be some fixed optimal solution to the given problem instance. In [24], it was proved that $X(H) \geqslant \frac{1}{4}X(\mathrm{OPT})$. As either H_1 or H_2 must contain at least half the elements covered by H, *Centralized MNU* is an 8-approximation algorithm for problem MNU.

(b) In each iteration of running *Centralized MNU*, the total cost of the sets added from any group G_i is bounded by B^*. Because we iterate *Centralized MNU* $\log_{8/7} n + 1$ times in algorithm *Centralized BLA*, the total cost of the sets added from any group G_i is bounded by $(\log_{8/7} n + 1)B^*$ when all elements in X are covered. Therefore, *Centralized BLA* is an $(\log_{8/7} n + 1)$-approximation algorithm for BLA.

(c) It is proven in [26]. □

Appendix B. The proofs of NP-hardness

This section presents the proofs for NP-hardness for the three problems discussed in the paper. For problems MNU and BLA, the proofs consider unicast users only, which is a special case of multicasting. For unicast users only, the problem MLA can be solved by associating based on RSSI. However, for multicast users, problem MLA is also NP-hard.

B.1. Reducing the bin packing problem to the MNU problem

The decision version of bin packing problem is defined as follows.

Definition 4. Bin packing: Given a finite set of natural numbers $U = \{g_1, g_2, \ldots, g_m\}$, and two positive integers B and K, decide if there is a partition of U into disjoint sets $U_1, U_2, \ldots, U_K$ such that $\sum_{g_i \in U_j} g_i \leqslant B$ for any $1 \leqslant j \leqslant K$.

Theorem 9.2. *The decision version of bin packing problem can be reduced to the MNU problem.*

Proof. We present a reduction from an arbitrary instance of the bin packing problem. We construct a WLAN with K APs. The maximum load every AP can support for multicast traffic is B. The WLAN supports m multicast sessions $s_1, s_2, \ldots, s_m$, where session s_i creates a load of g_i when transmitted at unit data rate. Corresponding to each session s_i we add one user requesting session s_i. So, there are m users in the WLAN. Each user has a link of unit data rate to all APs. If the maximum number of users this WLAN can support for multicast traffic is m, the bin packing problem has a positive answer; otherwise, it has a negative answer. Note that the maximum value of the load of an AP is 1 according to Definition 1 while B can be any natural number. However, we can make them less than 1 by dividing the maximum load B of every AP and the load g_i of every multicast session s_i by a large enough number. □

B.2. Reducing the Minimum Makespan scheduling problem to the BLA problem

Definition 5. Minimum Makespan scheduling: Given processing times for n jobs, $p_1, p_2, \ldots, p_n$, and an integer m, find an assignment of the jobs to m identical machines so that the completion time, also called the makespan, is minimized.

Theorem 9.3. *The Minimum Makespan scheduling problem can be reduced to the BLA problem.*

Proof. We construct a WLAN with m APs. Every AP only provides one transmission rate to users. The multicast sessions supported by this WLAN are $s_1, s_2, \ldots, s_n$. All APs can provide service to all users. The load requirement for a multicast session s_i is p_i $(1 \leqslant i \leqslant n)$. The objective is to minimize the maximum value of an AP's load among all APs under the limitation that all users get multicast service. □

B.3. Reducing the set cover problem to the MLA problem

We first define the Set Cover problem, and then present a reduction from the set cover problem to MLA problem to show that MLA is an NP-hard problem.

Definition 6. Set cover: There are m subsets $S_1, S_2, \ldots, S_m$ of a ground set $X = \{u_1, u_2, \ldots, u_n\}$. A cost $c(S_j)$ is associate with each set S_j. The objective is to find a subset H of $S = \{S_1, \ldots, S_m\}$ which covers all elements of X and has the minimum total cost. If the cost of every subset S_j is a same value c, the set cover problem is a cardinality version.

Theorem 9.4. *The cardinality version of set cover problem is reducible to MLA.*

Proof. We construct a WLAN with m APs, $a_1, a_2, \ldots, a_m$, and n users, $u_1, u_2, \ldots, u_n$. In this WLAN, all users request for the same multicast stream session with load requirement c. AP a_j can provide service to users in subset $S_j (1 \leqslant j \leqslant m)$. The link between the AP and the user has a unit data rate. The objective is to minimize the total load of all APs under the limitation that all users receive multicast service. □

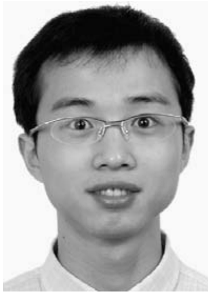
References

[1] IEEE 801.11, 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, http://standards.ieee.org/getieee802/802.11.html, 2007.

[2] G. Judd, P. Steenkiste, Fixing 802.11 access point selection, in: ACM SIGCOMM 2002 Poster Session, 2002.

[3] Y. Fukuda, Y. Oie, Decentralized access point selection architecture for wireless LANs, IEEE Transactions on Communications E90-B (9) (2007) 2513–2523.

[4] L. Wang, Z. Niu, Y. Zhu, H. Deng, M. Yano, Integration of SNR, load and time in handoff initiation for wireless LAN, in: Proceedings of PIMRC, Vol. 3, 2003, pp. 2032–2036.
[5] Y. Bejerano, S.-J. Han, L. Li, Fairness and load balancing in wireless LANs using association control, in: Proceedings of Mobicom, 2004.
[6] A. Kumar, V. Kumar, Optimal association of stations and APs in an IEEE 802.11 WLAN, in: Proceedings of the National Conference on Communications, IIT Kharagpur, 2005.
[7] A. Chen, D. Lee, P. Sinha, Optimizing multicast performance in large scale WLANs, in: IEEE ICDCS, Toronto, Canada, 2007, pp. 17–24.
[8] Y. Bejerano, D. Lee, P. Sinha, L. Zhang, Approximation algorithms for scheduling real-time multicast flows in wireless lans, in: Proceedings of IEEE INFOCOM Mini-Conference, Phoenix, AZ, 2008, pp. 151–155.
[9] J. Kuri, S.K. Kasera, Reliable multicast in multi-access wireless LANs, in: Proceedings of INFOCOM, 1999, pp. 760–767.
[10] K. Tang, M. Gerla, MAC layer broadcast support in 802.11 wireless networks, in: Proceedings of IEEE MILCOM, 2000, pp. 544–548.
[11] K. Tang, M. Gerla, Random access MAC for efficient broadcast support in ad hoc networks, in: Proceedings of WCNC, 2000, pp. 454–459.
[12] K. Tang, M. Gerla, MAC reliable broadcast in ad hoc networks, in: Proceedings of IEEE MILCOM, 2001, pp. 1008–1013.
[13] M.T. Sun, L. Huang, A. Arora, T.H. Lai, Reliable MAC Layer Multicast in IEEE 802.11 Wireless Networks, in: Proc. of ICPP, 2002, pp. 527–536.
[14] C.Y. Chiu, E.H. Wu, G.H. Chen, A reliable and efficient MAC layer broadcast (multicast) protocol for mobile ad hoc networks, in: Proceedings of Global Telecommunications Conference, 2004, pp. 2802–2807.
[15] W. Si, C. Li, RMAC: a reliable multicast MAC protocol for wireless ad hoc networks, in: Proceedings of ICPP, 2004, pp. 494–501.
[16] S. Gupta, V. Shankar, S. Lalwani, Reliable multicast MAC protocol for wireless LANs, in: Proceedings IEEE ICC, 2003, pp. 93–97.
[17] P. Chaporkar, A. Bhat, S. Sarkar, An adaptive strategy for maximizing throughput in MAC layer wireless multicast, in: Proceedings of MobiHoc, Roppongi, Japan, 2004, pp. 256–267.
[18] P. Chaporkar, S. Sarkar, Stochastic control techniques for throughput optimal wireless multicast, in: Proceedings of Control and Decision Conference (CDC), Maui, Hawaii, 2003, pp. 1598–1603.
[19] D. Lee, G. Chandrasekaran, P. Sinha, Optimizing broadcast load in mesh networks using dual-association, in: 1st IEEE Workshop on Wireless Mesh Networks, Santa Clara, CA, USA, 2005.
[20] C.T. Chou, A. Misra, Low latency multimedia broadcast in multi-rate wireless meshes, in: Proceedings of the First IEEE Workshop on Wireless Mesh Networks, 2005, pp. 54–63.
[21] A. Balachandran, G. Voelker, P. Bahl, V. Rangan, Characterizing user behavior and network performance in a public wireless LAN, in: Proceedings of ACM SIGMETRICS, Marina Del Rey, 2002, pp. 195–205.
[22] D. Kotz, K. Essien, Analysis of a campus-wide wireless network, in: Proceedings of ACM Mobicom, Marina Del Rey, 2002, pp. 115–133.
[23] F.P. Kelly, A. Maulloo, D. Tan, Rate control in communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society 49 (1998) 237–252.
[24] C. Chekuri, A. Kumar, Maximum coverage problem with group budget constraints and applications, in: Proceedings of the APPROX, 2004.
[25] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman Publishing Company, 1979.
[26] V. Vazirani, Approximation Algorithms, Springer-Verlag, 2001.
[27] ns2: Network Simulator, < http://www.isi.edu/nsnam/ns/>.
[28] I. Ramani, S. Savage, SyncScan: practical fast handoff for 802.11 infrastructure networks, in: Proceedings of IEEE INFOCOM, Miami, FL, USA, 2005.
[29] M.H. Manshaei, T. Turletti, Simulation-based performance analysis of 802.11a wireless LAN, in: Proceedings of International Symposium of Telecommunications (IST), Isfahan, Iran, 2003.



Ai Chen received his B.S. degree and M.S. degree in Electronic Engineering from the Tsinghua University, China in 2001 and 2004, respectively. He is now working toward his Ph.D. degree in Computer Science and Engineering of the Ohio State University, USA. His research interests include wireless networking and sensor networking.

Dongwook Lee received his Ph.D. degree in Information and Communication from Kwangju Institute of Science and Technology, Korea in 2004. He is a postdoctoral researcher in the Ohio State University since 2004. His research interests include wireless networks and mesh networks.

Prasun Sinha received his Ph.D. from University of Illinois, Urbana-Champaign in 2001, MS from Michigan State University in 1997, and B.Tech. from IIT Delhi in 1995. He worked at Bell Labs, Lucent Technologies as a Member of Technical Staff from 2001 to 2003. Since 2003 he is an Assistant Professor in Department of Computer Science and Engineering at Ohio State University. His research focuses on design of network protocols for sensor networks and mesh networks. He served on the program committees of various conferences including INFOCOM (2004–2007) and MOBICOM (2004–2005). He has won several awards including Ray Ozzie Fellowship (UIUC, 2000), Mavis Memorial Scholarship (UIUC, 1999), and Distinguished Academic Achievement Award (MSU, 1997). He received the prestigious NSF CAREER award in 2006.