Localized Access Point Association in Wireless LANs with Bounded Approximation Ratio

Mingming Lu and Jie Wu

Department of Computer Science and Engineering

Florida Atlantic University

Boca Raton, FL 33431

Abstract— **The current access point (AP) association schemes in wireless LANs, such as IEEE 802.11, cause an unbalanced load which reduces the performance of both the entire network and individual users. Intensive studies were motivated to determine efficient methods of balancing loads among APs. Previous works provided either the heuristics without theoretical analysis on the performance, or algorithms that require a centralized node or the propagation of global information. In this paper, we model the AP association problem as the many-to-one matching problem in the bipartite graph. Our objective is to maximize the total load among all APs. We propose a localized algorithm that provides a bounded approximation ratio in terms of total load and does not require the propagation of local information. We also extend the localized algorithm through iterative executions, and allow for adaptation to various environments. Extensive simulations are conducted to verify our results.**

I. INTRODUCTION

With the proliferation of wireless LANs, a wireless user can easily obtain access to multiple APs. Currently, most IEEE 802.11 protocols adopt the received signal strength indicator (RSSI) based approach in order to select an AP to affiliate with. Previous works [3], [7], [13] have shown that the RSSI-based approach can lead to poor performance in terms of the total load. To address this problem, existing works [3], [13] have proposed numerous approaches. All of them are essentially various AP access control schemes. Most of these approaches were evaluated through simulations or experiments on test beds. To the best of our knowledge, there are only two types of algorithms that provide performance guarantees. One is based on the linear programming (LP) approach [12], and the other is based on the simulated annealing technique [15].

However, the LP-based approach is centralized, and cannot adapt to the self-organized wireless LANs, where APs may be operated by different entities. Although the simulated annealing technique and the primal-dual scheme (derived from the LP) can be implemented in a localized manner, both of them require the propagation of global information and a large number of rounds in order to converge. This is not desirable in an environment with a highly-dynamic user population and APs owned by different entities. Therefore, we propose a pure, localized algorithm that provides performance guarantees and does not require the propagation of global information.

We model the AP association problem as a many-to-one matching problem in the bipartite graph, where each user can connect to only one AP, while an AP can be connected to multiple users. The objective of our problem is to maximize the network-wide throughput, i.e. the total load on all APs. Although it is well-known that optimal bipartite matching can be computed in polynomial time, the polynomial-time algorithms require either a centralized node to execute, or the propagation of global information.

The basic idea of our localized algorithm is as follows: 1) each user selects the AP with the best RSSI to send a connection request to; 2) each AP selects a set of users with the best RSSIs to establish connections with. Although this localized algorithm is simple, it has a constant approximation ratio to the optimal solution, and does not require the propagation of global information.

We first consider the case that all users have the same demand. In this case, the system objective can be reduced to the total number of users whose demands are satisfied. We theoretically prove that a constant approximation ratio of our localized algorithm exists. We also propose an iterative algorithm based on the localized algorithm to further improve the performance. We prove that a bound on the convergence rate in terms of the number of rounds exists. We then consider the case that users have different demands. We consider our localized algorithms in two more realistic environments: 1) taking into account the effect of interference by incorporating an interference model, and 2) adjusting the localized algorithm for use within a dynamic population where users can join and/or leave the network.

To sum up, the key contributions of our work are as follows:
- We fill the gap between existing centralized algorithms and existing localized heuristics by proposing a localized algorithm with a constant average approximation ratio.
- We further improve the performance of our localized algorithm by proposing an iterative localized algorithm, and prove the bound of iterations.
- We extend our localized algorithm to handle various dynamic environments and more realistic network settings.
- We evaluate our algorithms through extensive simulations.

The rest of this paper is organized as follows: Section II introduces the model and the notations, and formalizes the problem. In Section III, we propose the localized algorithms for the case where all users are homogeneous. Section IV

extends the localized solution to the case that users are heterogeneous, and extends our model to incorporate the effect of interference and dynamic environments. Section VI presents related works. Section V empirically evaluates performance through our customized simulator. Finally, Section VII concludes this work and outlines our future work.

II. PRELIMINARIES

A. The Network Model

A set of APs and a set of wireless users compose a wireless LAN. We adopt A to denote the set of APs. Each AP in A transmits with the same transmission power and can only serve users within its transmission range. The coverage area of the wireless network consists of the union of the area covered by each AP in A. We use U to denote the set of wireless users that reside in the coverage area. In order to access the wireless network, each user has to connect to an AP. We assume that the wireless users are free to move, but that they tend to stay in the same physical location for a long period of time. This assumption is backed up by recent studies of wireless user behavior [5], [17].

We model the network as a bipartite graph $(A \cup U, E)$, where $E \subseteq A \times U$ is the set of links connecting users to APs. There is a link between an AP and a user if and only if the user is within range of the AP. We call the user the AP's neighbor user and the AP the user's neighbor AP. We use u_i and a_j to denote user i and AP j, respectively.

B. Problem Formulation

To make our solutions applicable to the existing wireless LAN protocols, where users self-select their APs to access the network, our concern focuses on the localized solutions that maximize the total throughput, i.e. the total loads on all APs. The localized solutions mean that each user self-selects one of his neighbor APs to establish a connection with. His decision is based on his local information alone. Here, the local information of a user is the information about his neighbor APs.

We assume that there is a limitation on the capacity (transmission rate) of each AP, and the maximum capacity of each AP is fixed. Users may have different bandwidth requests (demands). We model the AP selection problem as the *many-to-one matching problem* in the bipartite graph, where each user can connect to only one AP, whereas each AP can be connected by multiple users simultaneously. The number of users that can connect to an AP simultaneously is constrained by the condition that the total demands from those users cannot exceed the capacity of the AP. The objective of our problem is to find a matching scheme so that the total *satisfied user bandwidth* is maximized. The bandwidth allocated to a user is satisfied if and only if the user connects to an AP, and the allocated bandwidth of the user is no less than his bandwidth request. In this paper, both the AP capacity and the user demand are measured in terms of bits per second. Note that existing WLAN protocols, such as IEEE 802.11e, allow users to specify their relative bandwidth demand based on access categories such as voice and video [10]. Although the

maximum transmission rate at each user may vary based on its geometric location, the maximum transmission rate achievable at the AP is fixed. The remaining capacity of each AP can be estimated according to the delay-based measurement scheme [20].

III. THE NETWORK WITH HOMOGENEOUS USER DEMAND

We first consider a special case where users have the same demand. Without loss of generality, we consider that each user has one unit of demand. Because user demands are assumed to be homogeneous, we can always scale user demand and AP capacity to make the user demand to be one unit. We also assume that the AP capacity is discretized such that one unit of AP capacity can serve one unit of user demand. Interestingly, even when users are restricted to select their APs to establish connection simultaneously in one round by exploring local information alone, we can design a localized algorithm that has a constant approximation ratio in expectation, and that is easy to implement.

A. Simultaneous Association in One Round

We assume that each user can sense the RSSIs from its neighbor APs. We also assume that each user can estimate the potential bandwidth (remaining capacity) associated with each neighbor AP by observing the delay of the beacon frame [22].

The basic idea of our localized algorithm is as follows: 1) each user selects the AP with the best RSSI to send a connection request to; 2) each AP selects a set of users with the best RSSIs to establish connections with. Although this localized algorithm is simple, it has a constant approximation ratio and does not require the propagation of global information.

More specifically, each user first scans its neighbor APs. During the scan process, each user senses the RSSI from each neighbor AP and estimates the AP's remaining capacity. For each user, the neighbor APs whose remaining capacities are larger than 0 are kept as the *available APs*. Note that we assume that the demands from all users are 1. After the scan process, each user prioritizes its available APs based on the RSSI. The higher the RSSI, the higher the priority. Then, each user tries to establish the connection to the AP with the highest priority by sending the ConnReq (connection request) message to that AP. If a user receives the acknowledgement message ConnAck (connection acknowledgement) from the AP that it sends ConnReq, the user establishes the connection and then quits the AP association process. Otherwise, the user simply quits the AP association process without establishing a connection.

At the AP side, once an AP receives at least one ConnReq from its neighbor users, it prioritizes those users according to their corresponding RSSIs. The higher the RSSI, the higher the priority. Then, the AP puts those users into1 a priority queue that is initially empty, and checks its remaining capacity. The AP assigns the bandwidth to request users one by one in order of priority from higher to lower until its remaining capacity becomes 0. Then, the AP sends ConnAck to the users that have been allocated bandwidth.

Before we formally present our localized algorithm, we need to introduce the following notations:

- A_u, user u's neighbor AP set;
- V_u, the available AP set of u;
- $S_u(a)$ the RSSI of AP a that is sensed at u;
- Q_a, the priority queue for a, which stores the users that send ConnReq to a with $S_u(a)$ as priority (the larger the $S_u(a)$, the higher the priority (u));
- $|Q_a|$, the size of Q_a;
- $C(a)$, the remaining capacity of a;
- a^*, the AP with the highest RSSI.

Algorithm 1 and Algorithm 2 present the operations at the user side and the AP side respectively.

Algorithm 1 Operations at User Side (u)

1: APScan(u);
2: Status $\leftarrow$ NotConnected;
3: $a^* \leftarrow \arg\max_{a \in V_u} S_u(a)$; /*find the best AP*/
4: Send ConnReq to a^*;
5: **if** Receive ConnAck from a^* **then**
6:     Status $\leftarrow$ Connected;
7:     Connect to a^*;

APScan(u)

1: **for** Each AP a in A_u **do**
2:     Sense signal strength $S_u(a)$;
3:     Estimate $C(a)$, the remaining capacity of a;
4:     **if** $C(a) > 0$ **then**
5:         Insert a into V_u;

Algorithm 2 Operations at AP Side (a)

Require: $Q_a = \emptyset$; /*Priority Queue is empty initially*/
1: **for** Each user u that sends ConnReq to a **do**
2:     Insert u to Q_a;
3: $k \leftarrow \min\{|Q_a|, C(a)\}$;
4: Send ConnAck to the first k users in Q_a;
5: $C(a) \leftarrow C(a) - k$;

In our one-round localized algorithm, each user selects the AP with the best RSSI. Combining with the assumptions that all APs transmit at the same power level, and that the RSSI $S_u(a)$ is proportional to the distance between AP a and user u, we can partition the covered area into numerous Voronoi regions with all APs as Voronoi points. Hence, there will be exactly one AP in each Voronoi region. Fig. 1 gives an example of a Voronoi graph with 5 APs and 7 users. Users in each Voronoi region will only select the AP in the same region because this AP has the highest RSSI among all neighbor APs. Based on the above observation, we can derive the following theorem about the approximation ratio of our one-round localized algorithm.

Theorem 1: If both APs and users are uniformly deployed in the coverage area, the approximation ratio of our one-round localized algorithm to the optimal solution is at least $1 - \frac{1}{e}$.

Due to space limitations, we do not present the proof of Theorem 1. The proof is motivated by an observation that
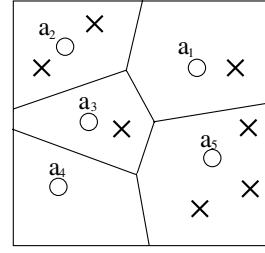

Fig. 1. The Voronoi graph with Voronoi regions centered at APs. There are 5 APs, represented by circles, and 7 users, represented by crossings.

the approximation ratio reaches its minimum when the total demand is equal to the total capacity. We regard each AP as an Voronoi point in a Voronoi region, and use probability techniques to calculate the lower bound of expected satisfied users and upper bound of optimal solution. The ratio of the above two bounds is the lower bound of our approximation ratio. The approximation ratio in Theorem 1 is ratio of the expected total satisfied demands produced by our one-round localized algorithm to the total satisfied demand produced by the optimal solution.

Under the assumption that both users and APs are uniformly distributed, the probability of overload (the number of users within a Voronoi region is larger than the capacity of the AP in the same Voronoi region) is relatively small. Hence, our localized heuristic has a high average approximation ratio.

In our one-round localized algorithm, not all users can connect to an AP. However, our algorithm does not cause conflict, i.e., the total demands from all users associated with an AP do not exceed the capacity of the AP. Therefore, the performance of our algorithm can be further improved through iterative executions.

B. Iterative Association in Multiple Rounds

The iterative association executes in multiple rounds. In each round, each unconnected user estimates the remaining capacities of his neighbor APs, counts those neighbor APs whose remaining capacities are larger than or equal to his demand as his available APs, and sends a ConnReq to the AP with the highest RSSI from the available APs. Although the iterative association is relatively simple, which executes the simultaneous association iteratively, it has a very good performance in terms of both the approximation ratio and the convergence rate. Here, the convergence rate is defined as the number of rounds. The fewer the number of rounds, the higher the convergence rate. We have the following theorem concerning the upper bound for the converge rate of our iterative localized algorithm.

Theorem 2: Our iterative association executes at most $\ln l$ rounds, where $l = \min\{m, n\}$, and m and n are the number of users and APs, respectively.

Proof: We first consider the case that $m \leq n$. Based on the analysis of Theorem 1, we observe that at least $(1 - \frac{1}{e}) \times m$ users are connected after the first round. Hence, the number of unconnected users is at most $\frac{1}{e} \times m$. Among the unconnected

users, we use d_u^i to denote the number of unconnected users that cannot be covered by any AP after the i-th round, and r_u^i to denote the number of the unconnected users except those uncovered users. Assume that the number of rounds is k. We have the following inequality:

$$r_u^i \leq r_u^i + d_u^i \leq \frac{1}{e} \times r_u^{i-1}, \quad 1 \leq i \leq k. \tag{1}$$

Combining $r_u^k \geq 1$ and Formula (1), we can conclude that $1 \leq (\frac{1}{e})^k \times m$. Note that $r_u^0 = m$. Therefore, we have $k \leq \ln m$. For the case where $m \geq n$, at least $(1 - \frac{1}{e}) \times n$ users are connected after the first round. Similarly, we can derive $k \leq \ln n$. Therefore, we have $k \leq \ln l$. ∎

In Section V, we conduct experiments to evaluate the approximation ratio and the convergence rate of our iterative algorithm. The simulation results show that the performance of the iterative algorithm approximates to the optimal solution, and its convergence rate is fast.

C. Non-Uniform User Distribution

Although our theoretical results are based on the assumption of uniform user distribution, these results also hold in stochastic distributions other than uniform distribution. A previous study [4] has shown that users tend to access network at certain popular spaces ("hot-spots") within the network. In our experiment, we adopt normal distribution to model such hot-spots. We use the user location as random variable, and the location of those hot-spots as the mean value of the normal distribution. According to the guidelines [2], APs should be distributed based on the number of users, the physical aspects of locations, etc. Thus, we adopt the normal distribution that is identical to the user distribution to deploy APs so that the ratio of users to APs are approximately the same. Through experimental evaluation, we identify that both our one-round and iterative localized algorithms still have the property of constant approximation ratio, and the convergence rate is still very fast.

IV. THE NETWORK WITH HETEROGENOUS USER DEMAND

In Section III, we assumed that users have homogeneous demands. In this section, we relax this assumption by allowing users to have heterogenous demands. We first describe the needed modifications that enable the above localized algorithms to adapt to the case of the heterogenous demand. Then, we evaluate our localized algorithms in the following practical environments: 1) the effect of interference on the AP capacity is incorporated, 2) users randomly enter or leave the network.

The modifications consist of two parts: first, for the operations at the user side (Algorithm 1), line 4 of procedure APScan(u) needs to change to "if $C(a) > B(u)$ then", where $B(u)$ is the demand of user u; second, the operation at the AP side (Algorithm 2) should change to Algorithm 3. The localized algorithms for the case of the heterogenous user demands are different from those for the case of the homogeneous user demands in that users estimate the potential bandwidths of neighbor APs, and regard the neighbor APs with higher estimated bandwidth than their demands as their

available APs. Each AP adds users into its priority queue, and pop users in the order based on the corresponding RSSIs until the demand of the last user cannot be satisfied.

If the value of user demands are discrete and can be split, the theoretical analysis for the homogeneous demand can be applied to the heterogenous demand by splitting the demand from a user with D units of demand into D users of unit demand. Therefore, for the splittable heterogeneous demand, there also exists a constant approximation ratio. For the non-splittable heterogenous demand, we evaluate the performance of our localized algorithm through experiment study. Through simulation, we observe that our one-round localized algorithm still has a constant approximation ratio, and the iterative algorithm still converges fast, with the performance approximating the optimal solution.

Algorithm 3 Bandwidth-based Operations at AP Side (a)

Require: $Q_a = \emptyset$; /*Priority Queue is empty initially*/
1: **for** Each user u that sends ConnReq to a **do**
2:     Insert u to Q_a;
3: $u \leftarrow \text{Pop}(Q_a)$;
4: **while** $C(a) > B(u)$ **do**
5:     Include u into the target set of ConnAck;
6:     $C(a) \leftarrow C(a) - B(u)$;
7:     $u \leftarrow \text{Pop}(Q_a)$;
8: Send ConnAck;

A. Incorporate Interference

In this subsection, we consider the effect of the interference. We adopt the interference model proposed in [14], where, for a user u_i that connects to an AP a_j, the Signal to Interference Ratio (SIR) at u_i is as follows:

$$SIR_{u_i} = \frac{S_{u_i}(a_j)}{\sum_{k \neq j} S_{u_i}(a_k)}$$

where a_k is an AP other than a_j in the covered area, and $S_{u_i}(a_j)$ is independent, exponentially distributed random variable. The mean value of $S_{u_i}(a_j)$ is $c \cdot P_{a_j}/d(a_j, u_i)^\alpha$, where P_{a_j} is the transmission power of AP a_j, $d(a_j, u_i)$ is the distance between them, c is a constant, and $\alpha = 2$ or 4. We assume that reception can occur provided the SIR exceeds a given threshold SIR^{th}.

The error probability of transmission from a_j to u_i is given by

$$Prob(SIR_{u_i} \leq SIR^{th}) = Prob(S_{u_i}(a_j) \leq SIR^{th} \sum_{k \neq j} S_{u_i}(a_k)).$$

The transmission error probability can be expressed in analytical form:

$$error_{u_i} = 1 - \prod_{k \neq j} \frac{1}{1 + SIR^{th} d(a_j, u_i)^\alpha / d(a_k, u_i)^\alpha}.$$

The above analytic expression for the transmission error probability is based on the assumption that all APs transmit at the same power, and is derived from the following result:

(a) AP transmission range dimension.

(b) User number dimension.

(c) AP number dimension.
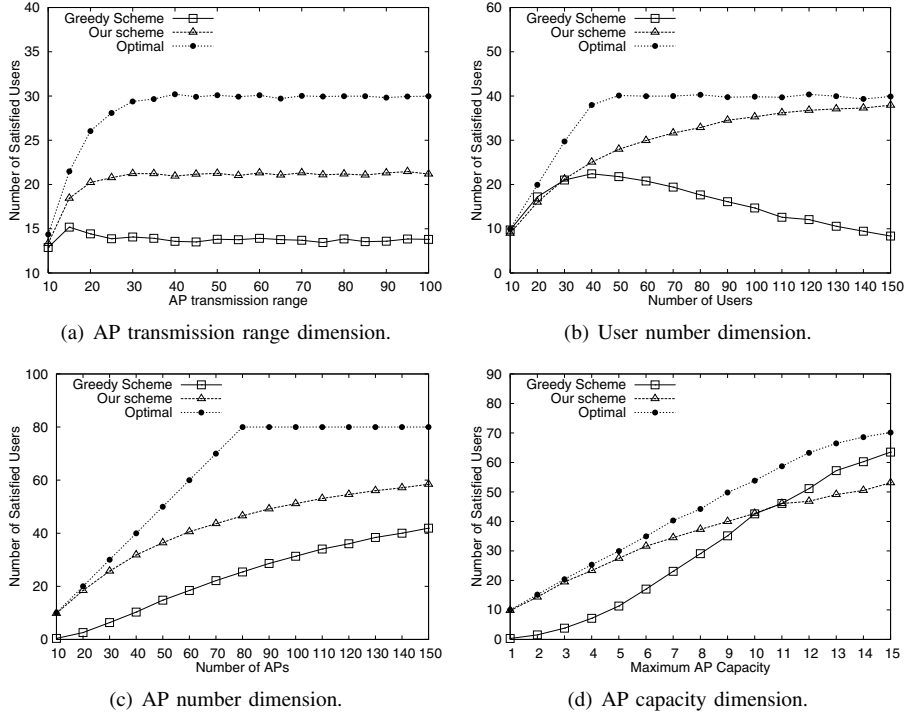
(d) AP capacity dimension.

Fig. 2. Comparison of the greedy scheme, our one-round localized algorithm, and the maximum flow based optimal solution in terms of the total satisfied demands in the homogeneous user demand model.

Suppose $z_1, \cdots, z_n$ are independent exponentially distributed random variables with means $Ez_i = 1/\lambda_i$, Then we have

$$Prob(z_1 \leq \sum_{i=2}^{n} z_i) = 1 - \prod_{i=2}^{n}(\frac{1}{1 + \lambda_1/\lambda_i}).$$

Due to transmission error, not all packets transmitted from AP a_j can be received by user u_i. We assume that undelivered packets incur retransmissions. As proposed in [8], for any packet, the expected number of transmissions from a_j to u_i is $\frac{1}{1-error_{u_i}}$. Since retransmissions consume bandwidth assigned by a_j, we should count the number of bits delivered to u_i every second as the bandwidth. We call this bandwidth *effective bandwidth*. If AP a_j assigns bandwidth B to user u_i, the effective bandwidth of u_i is $B(1 - error_{u_i})$. By taking interference into account, our objective turns to maximizing the total effective bandwidth.

Through simulations, we verify that the approximation ratio still holds even when the interference is taken into account.

B. Dynamic User Population

In the previous sections, we have considered the localized algorithms for the static network topology. In this subsection, we study dynamic environments where users can join and/or leave the network. When a user leaves the system, the load at the corresponding AP decreases, potentially motivating other users to change their previous selection and associate with that AP. When a user joins, he automatically selects an AP to affiliate with.

Our objective in this subsection is to discuss the impact that dynamic user joining and/or leaving has on the system performance. When users join or leave the network frequently, it is desirable to design a user association scheme without requiring many users to handoff to different APs since the overhead of handoff is non-negligible. We modify the existing localized algorithms for static user populations by requiring that a user will not switch to a different AP unless the RSSI from a new AP is improved by a threshold.

In Section V, we use simulation to investigate the behavior of the system under a dynamic user population. We adopt the Poisson process to simulate the inter-arrive time as well as the service time of users.

V. SIMULATION

A. Simulation Setup

In this section, we give an experimental evaluation of our localized algorithms in our custom simulator. Our concern focuses on two metrics: the total load and convergence rate (in terms of the number of rounds). We first consider the case of homogeneous user demand, and then consider the case of heterogeneous user demand. In both cases, we compare the performance of our one-round localized algorithm, the iterative localized algorithm, and the optimal solution. We adopt the Edmonds Karp MaxFlow algorithm [9] to compute the optimal bipartite matching.

We simulate a stationary network with APs and users randomly located in a $100m \times 100m$ area. We assume that all APs and users can be deployed in this area arbitrarily. In the simulation, we consider the following tunable parameters: (a) m, the number of users, (b) n, the number of APs, (c) r, transmission range of APs, and (d) c, the maximum capacity

(a) Performance comparison (homogeneous demand).

(b) Converge rate (homogeneous demand).

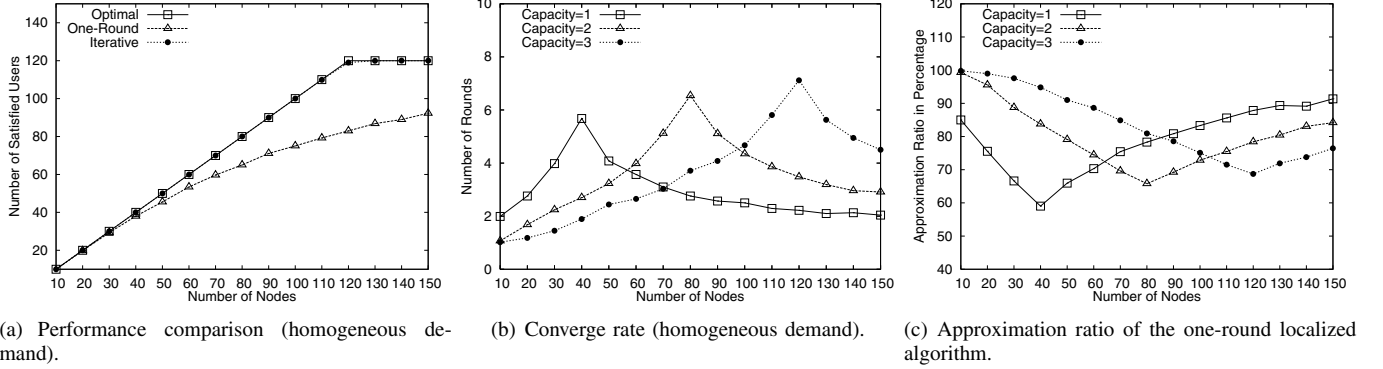(c) Approximation ratio of the one-round localized algorithm.

Fig. 3. Simulations on the performance comparison, approximation ratio for homogeneous demand.

of APs. The capacity of each AP is a random number between 1 and c. (e) d, the user demand. In the case of homogeneous user demand, $d = 1$. In the case of heterogenous user demand, d is the maximum user demands. User demand of each user is random number ranging from 1 to d.

B. Simulation Results

First, we will show that the greedy scheme, where users associate with the AP that has the best RSSI, and APs do not implement any admission control, indeed incurs poor performance by comparing with our one-round localized algorithm and the optimal algorithm when both users and APs are uniformly distributed. We also verify the approximation ratio of our one-round localized algorithm in the case of homogeneous user demand by testing four dimensions: the AP transmission range r, the number of users m, the number of APs n, and the maximum AP capacity c. The simulation result of the first dimension (r) is shown in Fig. 2(a), where $m = 40$, $n = 30$, $c = 2$, and r increases from $10m$ to $100m$ in increments of $10m$. From this result, we observe that the performances of our one-round localized algorithm and the optimal solution increase quickly, as r increases from $10m$ to $30m$, and the ratio of the performances become approximately a constant if $r \geq 40m$. The performance of the greedy scheme reaches its maximum at $r = 15m$, and remains almost the same at $r \geq 30m$. The reason for the short increments of performance is that the area is not fully covered when $10m \leq r \leq 40m$, which in turn incurs uncovered users. Therefore, we can conclude that the AP transmission range has little effect on the performance of the three algorithms when it is greater than a threshold value, which is $40m$ in our experiment.

The simulation result of the second dimension (m) is shown in Fig. 2(b), where $r = 40m$, $n = 20$, $c = 3$, and m increases from 10 to 150 in increments of 10. From this result, we observe that the performance of the greedy scheme decreases dramatically as the number of users increases. This performance decrement occurs before the entire network overloads. The fundamental reason is that the greedy scheme easily causes locally-unbalanced loads, which in turn decreases the performance of the entire network. On the contrary, the

performance of our one-round localized algorithm becomes closer to that of the optimal solution as the number of users increases. The reason is that our one-round localized algorithm is conflict-free, and the number of the empty AP-centered Voronoi regions decreases as the number of users increases.

The simulation result of the third dimension (n) is shown in Fig. 2(c), where $r = 40m$, $m = 80$, $c = 1$, and n increases from 10 to 150 in increments of 10. From this result, we observe that the increment of the number of APs increases the performances of all three algorithms. That is because the increment of the number of APs increases the network capacity. The simulation result of the fourth dimension (c) is shown in Fig. 2(d), where $r = 40m$, $m = 80$, $n = 10$, and c increases from 1 to 15 in increments of 1. From this result, we observe that the increment of the maximum AP capacity can improve the performance of all three algorithms. As the AP capacity increases, more user demands can be accommodated into the APs. Thus, greedy algorithm is working better.

In the second experiment, we compare the performance of our one-round localized algorithm, our iterative localized algorithm, and the optimal algorithm. The simulation result is illustrated in Fig. 3(a), where $n = 40$, $r = 100m$, $c = 3$, and m ranges from 10 to 150 in increments of 10. From this result, we can conclude that the total load by our iterative algorithm is very close to the optimal solution. From Fig. 3(b), where the parameters are the same as that of Fig. 3(a), except that $c = 1, 2$, and 3, we observe that the converge rate of the iterative solution reaches its minimum (the maximum number of rounds) when the total capacity is equal to the total demand (the sum of demands of all users) no matter what the value of AP capacity is. The bigger the AP capacity, the slower the minimum convergence rate associated with that AP capacity. The reason is that fewer users can establish connection when the total demand is equal to the total capacity. We also evaluate the approximation ratio of the one-round localized algorithm. The simulation result is illustrated in Fig. 3(c), where the parameters are the same as that of Fig. 3(b), we observe that all the approximation ratios for the three different capacities reach their minimum values when the total capacity is equal to the total demand for different AP capacity. The smaller the AP capacity, the lower the minimum approximation ratio associated with that
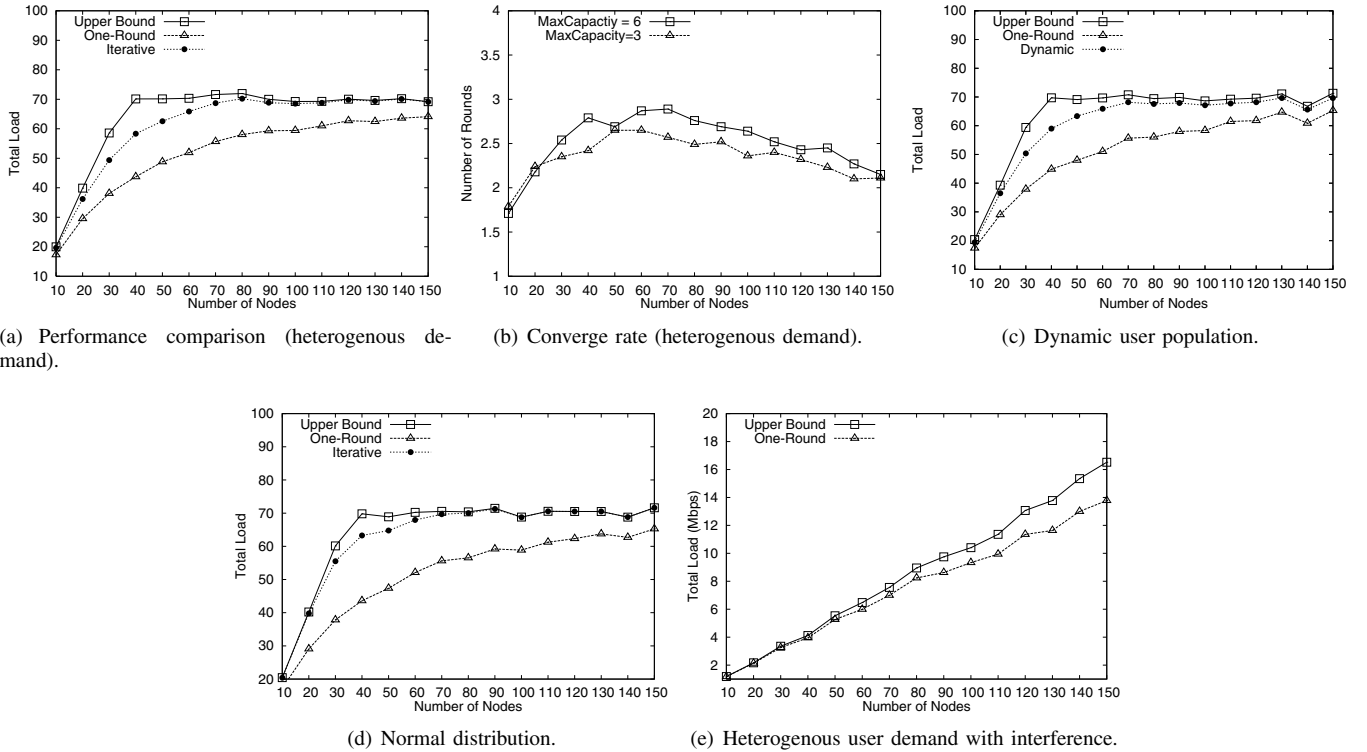
(a) Performance comparison (heterogenous demand).

(b) Converge rate (heterogenous demand).

(c) Dynamic user population.

(d) Normal distribution.

(e) Heterogenous user demand with interference.

Fig. 4. Simulations on the performance comparison, approximation ratio for heterogenous demand, as well as simulations on dynamic user population, normal user distribution, and the effect of interference.

AP capacity. Therefore, when $c = 1$ and the total capacity is equal to the total demand, the approximation ratio is minimum, which is consistent with Theorem 1.

In the third experiment, we consider the case where users have heterogenous demands. We first compare the performance of our one-round localized algorithm, our iterative localized algorithm, and the upper bound of the optimal solution, which is the minimum of the total capacity and the total demand because the optimal solution cannot be larger than the total capacity or the total demand. The simulation result is illustrated in Fig. 4(a), where $n = 20$, $r = 40m$, $c = 6$, maximum user demand $d = 3$, and m ranges from 10 to 150 in increments of 10. This result testifies that our localized algorithms maintain a good approximation in the case of heterogenous demand. We then evaluate the convergence rate of the iterative localized algorithm. Fig. 4(b) shows the simulation result, which illustrates the fast convergence rate in terms of the number of rounds. In the same parameter setting as that of Fig. 4(b), we also conduct the experiment on the dynamic user population. We adopt the Poisson process to simulate the inter-arrive time of users. The simulation result shown in Fig. 4(c) illustrates that our result holds for dynamic user population either. In the same parameter setting, we conduct the experiment on the normal distribution on user deployment. The simulation result shown in Fig. 4(d) illustrates that the constant approximation ratio property of our localized algorithms does not limited to the uniform user distribution.

We also evaluate the performance under the environment where interference is taken into account. The simulation result is shown in Fig. 4(e), where the area size is $1000m \times 1000m$, $n = 10$, $r = 250m$, $c = 11Mbps$, maximum user demand $d = 1Mbps$, and m ranges from 10 to 150 in increments of 10. The simulation result illustrates that our localized algorithm approximates to the optimal solution in performance when the interference is considered.

VI. RELATED WORK

Currently, in the IEEE 802.11 based wireless networks, each user greedily connects to the AP with the strongest signal strength. Various works [3], [7], [13], identified the poor performance of this approach through analysis and simulation. Recent studies [5], [6], [17] on wireless networks have shown that the user load is often distributed unevenly among APs because of the current AP selection strategy. Numerous works [3], [19], [21] initialized the study on the load balance among APs. This motivated intensive study on more efficient methods to balance the load and achieve different objectives. As far as we know, there are only two types of algorithms that provide performance guarantees. One is based on the linear programming (LP) approach [7], [12], and the other is based on the simulated annealing technique [15].

Bejerano et al. [7] proposed a user-AP association model that considers both load balance and fairness. The associated problem is NP-hard. By utilizing the existing LP-based methods, they designed approximation algorithms with a constant approximation ratio. However, their LP-based approach is centralized. Bahl et al. [12] also proposed a LP-based scheme to handle client congestion in wireless LANs. They modeled

the AP association problem as a weighted matching problem in bipartite graph, and adopted the power control at the AP side to balance the loads among APs. Their major contributions are their solid analysis of the worst case bounds, and the avoiding of the modification at the user side. Although their primal-dual method (derived from their LP formulation) is distributed, it requires the propagation of global information and a large number of rounds in order to converge. This implies two drawbacks: 1) high communication complexity when the number of APs and users is large; 2) inapplicability to wireless LANs, where APs are owned by different entities. Kauffmann et al. [15] proposed a distributed solution based on simulated annealing techniques with performance guaranteed. But it has the same drawbacks as the above primal-dual method.

Vasudevan et al. [22] proposed a localized measurement-based method for users to associate APs in IEEE 802.11-based wireless networks. In their scheme, users make decisions as opposed to APs. Each user estimates the potential bandwidth of neighbor APs by passive measurement and selects the AP with the highest potential bandwidth to connect. Papanikos and Logothetis [19] also proposed a load balancing heuristic that helps users to select the AP to associate with based on the number of users already associated with the AP and the mean RSSI. Haidar et al. [11] proposed an AP association scheme based on power management in order to minimize the maximum congested AP. However, none of above localized schemes provided a bounded approximation ratio.

In this work, we focused on the localized solution that provides a bounded approximation ratio. We modeled the user-AP association problem as a many-to-one matching problem in a bipartite graph. We adopted the number of users connected to an AP as the load capacity of this AP by assuming all users have the same minimum quality requirements. Mittal, Belding, and Suri [18] also adopted the same metric to model the load capacity of APs. Although Kelsen [16] proposed a local algorithm for parallel matching in expander graphs and proved that the proposed algorithm can find the maximum matching with high probability in sub-linear running time, it is unpractical in IEEE 802.11 based wireless LANs because it requires complicated communications between neighbor nodes and a large number of rounds to converge.

Akella et al. [1] pointed out that the current deployment of APs is unplanned, i.e., most APs are deployed by users in a spontaneous manner, resulting in highly variable AP densities. This reduces the performance of end-users. To model the deployment of APs in the real world, we randomly deploy APs in an area so that the AP densities are variable.

VII. CONCLUSIONS

In this work, we model the AP selection in IEEE 802.11 wireless networks as the many-to-one matching problem in the bipartite graph. Our objective is to maximize the total load among all APs. We propose a localized algorithm that provides a bounded approximation ratio in terms of average load. We also extend the localized algorithm through iterative executions, adapted to various environments. Extensive simulations are conducted to verify our results.

REFERENCES

[1] A. Akella, G. Judd, S. Seshan, and P. Steenkiste. Self-management in chaotic wireless deployments. In *Proceedings of ACM MobiCom'05*, pages 185–199, 2005.

[2] B. Alexander and S. Snow. Preparing for wireless LANs: Secrets to successful wireless deployment. Packet Magazine. Cisco Systems. http://www.cisco.com/warp/public/784/ packet/apr02/p36 cover.html.

[3] A. Balachandran, P. Bahl, and G. Voelker. Hot-spot congestion relief and service guarantees in public-area wireless networks. *SIGCOMM Computer Communication Review*, 32(1), 2002.

[4] A. Balachandran, P. Bahl, and G. M. Voelker. Hot-spot congestion relief in public-area wireless networks. In *Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications (WM-CSA'02)*, page 70, 2002.

[5] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan. Characterizing user behavior and network performance in a public wireless LAN. In *Proceedings of the ACM SIGMETRICS'02*, pages 195–205, 2002.

[6] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of MobiSys'03*, 2003.

[7] Y. Bejerano, S. Han, and L. Li. Fairness and load balancing in wireless lans using association control. In *Proceedings of ACM MobiCom'04*, 2004.

[8] S. Biswas and R. Morris. ExOR: Opportunistic multi-hop routing for wireless networks. In *Proceedings of ACM SIGCOMM'05*, pages 133–144, 2005.

[9] T. Cormen, C. Stein, R. Rivest, and C. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2001.

[10] D. Gao, J. Cai, and K. N. Ngan. Admission control in ieee 802.11e wireless lans. *IEEE Network, special issue on Wireless Local Area Networking: QoS provision and Resource Management*, 19(4):6–13, 2005.

[11] M. Haidar, R. Akl, H. Al-Rizzo, Y. Chan, and R. Adada. Optimal load distribution in large scale WLAN networks utilizing a power management algorithm. In *Proceedings of IEEE Sarnoff Symposium*, 2007.

[12] M. T. Hajiaghayi, S. V. Mirrokni, A. Saberi, P. Bahl, K. Jain, and L. Qiu. Cell breathing in wireless LANs: Algorithms and evaluation. *IEEE Transactions on Mobile Computing*, 6(2):164–178, 2007.

[13] G. Judd and P. Steenkiste. Fixing 801.11 access point selection. In *Poster in Proceedings of ACM MobiCom'02*, 2002.

[14] S. Kandukuri and S. Boyd. Optimal power control in interference limited wireless channels with outage probability fading specications. *IEEE Transactions on Wireless Communications*, 1(1):46–55, 2002.

[15] B. Kauffmann, F. Baccelli, A. Chaintreau, V. Mhatre, K. Papagiannaki, and C. Diot. Measurement-based self organization of interfering 802.11 wireless access networks. In *Proceedings of INFOCOM'07*, 2007.

[16] P. Kelsen. Fast parallel matching in expander graphs. In *Proceedings of the fifth annual ACM symposium on Parallel algorithms and architectures (SPAA'93)*, pages 293–299, 1993.

[17] D. Kotz and K. Essien. Analysis of a campus-wide wireless network. *Wireless Network*, 11(1-2):115–133, 2005.

[18] K. Mittal, E. Belding, and S. Suri. A game-theoretic analysis of wireless access point selection by mobile users. Technical report, University of California, Santa Barbara, 2006.

[19] I. Papanikos and M. Logothetis. A study on dynamic load balance for IEEE 802.11b wireless LAN. In *Proceedings of the 8th International Conference on Advances in Communications and Control (COMCON'01)*, 2001.

[20] S. Shah, K. Chen, and K. Nahrstedt. Available bandwidth estimation in IEEE 802.11-based wireless networks. In *Proceedings of the 1st ISMA/CAIDA Workshop on Bandwidth Estimation (BEst'03)*, 2003.

[21] T. -C. Tsai and C. -F. Lien. IEEE 802.11 hot spot load balance and qos-maintained seamless roaming. In *Proceedings of the 2003 National Computer Symposium (NCS'03)*, 2003.

[22] S. Vasudevan, K. Papagiannaki, C. Diot, J. Kurose, and D. Towsley. Facilitating access point selection in IEEE 802.11 wireless networks. In *Proceedings of the 5th ACM SIGCOMM conference on Internet measurement (IMC'05)*, pages 293–298, 2005.