Association Control for Vehicular WiFi Access: Pursuing Efficiency and Fairness

Lei Xie, *Member*, *IEEE*, Qun Li, *Member*, *IEEE*, Weizhen Mao,
Jie Wu, *Fellow*, *IEEE*, and Daoxu Chen, *Member*, *IEEE*

Abstract—Deploying road-side WiFi access points has made possible internet access in a vehicle, nevertheless it is challenging to maintain client performance at vehicular speed especially when multiple mobile users exist. This paper considers the association control problem for vehicular WiFi access in the *Drive-thru Internet scenario*. In particular, we aim to improve the efficiency and fairness for all users. We design efficient algorithms to achieve these objectives through several techniques including approximation. Our simulation results demonstrate that our algorithms can achieve significantly better performance than conventional approaches.

Index Terms—Association control, vehicular network, efficiency, fairness.

1 Introduction

Improvements in wireless technology have made it possible to deploy wireless networks spanning an entire metropolitan area. The availability of anywhere, anytime, wireless connectivity will create new categories of users. One such category is called "*Drive-thru Internet*" [1], which provides wireless access to users in moving vehicles through road-side deployed APs. These vehicular users encounter unique challenges not faced by conventional indoor users, such as dynamically changing network structure of AP-user pairing and contentions among mobile users. Unlike a wireless network comprising of static or slow moving users, vehicular users are continuously moving at high speeds, making existing AP selection and handoff algorithms unsuitable.

In order to achieve reasonable efficiency among multiple vehicular users for the above *Drive-thru Internet* scenario, several problems should be considered, e.g., rate adaptation and association control. Association control defines, while multiple users are driving along the road, how to intelligently associate vehicular users to APs and when to appropriately conduct handoffs for users to improve the overall system performance. Compared with rate adaptation [2], which adapts the modulation and coding scheme according to the quality of the radio channel, association control considers the entire network from a macrolevel perspective, which shows how to optimize system performance from a higher level viewpoint. We notice that, albeit some recent work on association control for static networks, there is little work on how to manage AP association in this type of "Vehicular Networks". We believe that a thorough theoretical study on this problem is highly necessary for the future deployment of vehicular networks. Some pitfalls can be avoided in real deployment if we have a better understanding first.

This paper aims to define a theoretical framework to analyze the performance of a vehicular network in the *Drive-thru Internet* scenario, in particular to investigate association control schemes. Considering both the long-term efficiency and fairness metrics, we propose optimized schemes to associate mobile users with APs, and approximation algorithms to reduce computation complexity of calculating optimal solutions. To the best of our knowledge, this is the first theoretical work that investigates the optimization problem for association control in vehicular networks. The contributions of this paper are summarized as follows:

1. Since the association solutions are updated frequently while the users are driving along the roads, this paper is concerned about the long-term performance in terms of efficiency and fairness, and proposes novel algorithms to achieve these long-term objectives.
2. We propose a theoretical framework for association control over vehicular networks. For the efficiency metric, the problem is transformed into an optimization problem for each snapshot over the long-term service duration. For the fairness metric, we, respectively, consider the optimization solutions for proportional fairness and max-min fairness.
3. When the involved number of mobile users and APs along the road is rather large, to reduce the computation complexity, we propose an approximation algorithm to break the large contention group into smaller subgroups, achieving a trade-off between accuracy and computation complexity.

The rest of the paper is organized as follows: We briefly present related work in Section 2. We define the performance metrics and introduce our model and assumptions in Section 3. We illustrate our overall optimization and snapshot solutions in Section 4, respectively, for efficiency

- *L. Xie and D. Chen are with the State Key Laboratory of Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing 210093, China. E-mail: lxie@nju.edu.cn, cdx@nju.edu.cn.*
- *Q. Li and W. Mao are with the Department of Computer Science, McGlothlin-Street Hall, College of William and Mary, Williamsburg, VA 23187-8795. E-mail: {liqun, wm}@cs.wm.edu.*
- *J. Wu is with the Department of Computer and Information Science, Temple University, 1805 N. Broad ST., Philadelphia, PA 19122. E-mail: jiewu@temple.edu.*

Manuscript received 14 Sept. 2009; revised 9 June 2010; accepted 15 July 2010; published online 3 Jan. 2011.
Recommended for acceptance by S. Olariu.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2009-09-0422.
Digital Object Identifier no. 10.1109/TPDS.2011.17.

and fairness. We show some major simulation results in Section 5, and we conclude the paper in Section 6.

2 Related Work

Association control and scheduling solutions for Wireless LANs (WLAN) have been intensely studied, mainly targeting the efficiency and fairness metrics. Tassiulas and Sarkar consider the max-min fair allocation of bandwidth in wireless ad hoc networks [3]. Bejerano et al. present an efficient solution to determine the user-AP association for the max-min fair bandwidth allocation [4]. Li et al. consider proportional fairness for WLANs [5]. Internet access with vehicular speeds in the IEEE 802.11 networks have been studied in recent research works. Bychkovsky et al. study the case for vehicular clients to connect to open-access residential wireless 802.11 access points in Boston [6], [7]. Giannoulis et al. address the problem of maintaining client performance at vehicular speeds within city-wide multihop 802.11 networks [8]. Ott and Kutscher report on measurements for the use of 802.11 networks in the *Drive-thru Internet scenario* [1]. Mahajan et al. deploy a modest-size test bed and analyze the fundamental characteristics of WiFi-based connectivity between base stations and vehicles in urban settings [9]. Hadaller et al. show that by exploiting wireless conditions, vehicular opportunistic access can be greatly improved [10]. Navda et al. investigate the use of directional antennas and beam steering techniques to improve performance of 802.11 links in the context of communication between a moving vehicle and roadside APs [11]. Kim et al. present novel association control algorithms that minimize the frequency of handoffs occurred to mobile devices [12]. Deshpande et al. exploit historical information to develop new handoff and data transfer strategies for improved vehicular WiFi access [13]. Wu et al. have developed a fast handoff scheme called *Proactive Scan* to reduce the handoff delay [14]. In the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17, the related works are introduced in a more comprehensive approach.

3 Performance Models and Metrics

3.1 Models and Assumptions

In the *Drive-thru Internet* scenario, vehicular users are driving through a region covered with multiple roads, and APs are deployed along the roads nonuniformly by the service provider. Each AP has a limited coverage range and it can only serve users in its coverage area. We assume a careful frequency planning where interfering APs are assigned to orthogonal channels so that adjacent APs can fully utilize their bandwidth without causing interference to each other. Conventionally, each user on the roads may have one or more candidate APs to associate with at any time, and each time the user can only associate with exactly one AP. Furthermore, contentions for transmission may exist among users if they associate with the same AP. If a large number of users associate with the same AP, their allocated bandwidths will be greatly reduced. We assume that different users have various velocities (including speeds and directions) which may vary over the time. Thus while users are driving along the roads, at different time instants and positions, they may be contending with different users for bandwidth from different APs. Each user associates with the first AP after first entering the Wi-Fi deployment area, then goes through a series of handoffs among different APs while driving along the roads, and disconnects at the last associated AP before leaving the Wi-Fi deployment area. In this paper, we seek a series of optimized association solutions based on underlying technologies [13], [14] to conduct fast handoffs, which can limit the handoff delay within several milliseconds, so that the handoffs can be performed at a small cost.

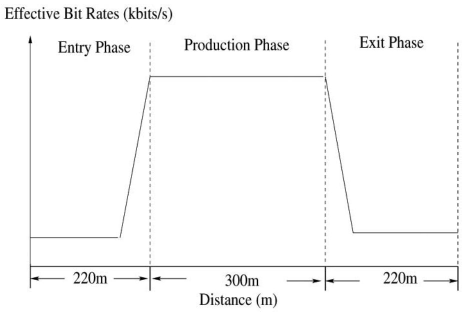


Fig. 1. Three connectivity phases.

We denote the set of APs as A indexed by $i = 1, \dots, m$ and denote the set of users as U indexed by $j = 1, \dots, n$. We consider association control over the time interval $[0, T]$. For example, 0 and T may, respectively, denote $0:00$ and $24:00$ time points of every day. For each AP-user pair (i, j), we assume that the effective bit rate $r_{i,j}(t)$ of the link between i and j at time t is known. The effective bit rate is measured over a fairly long time period and also takes into account the overhead of retransmissions due to reception errors. We use $b_j(t)$ to denote the bandwidth allocated to user j at time t. Both bit rate and bandwidth can be measured in bits per second (bit/s). For bandwidth allocation inside each AP, we use time-based fairness for scheduling. Once an AP is associated with some users, each user is assigned an equal-sized time slot regardless its effective bit rate, and is supposed to use all the allocated bandwidth. Thus, if n' users are associated with AP i at time t, then the bandwidth allocated to user j is $b_j(t) = r_{i,j}(t)/n'$.

For the effective bit rate setting in the *Drive-thru Internet* scenario, we adopt the model proposed in [1]. Fig. 1 depicts three different connectivity phases with respect to effective bit rate and relative distance between the user and AP. The entry phase and exit phase provide very weak connectivity, only the production phase provides a window of useful connectivity. As the connection is built between a user and an AP, it will maintain a constant bit rate in the production phase, which mainly depends on the AP's signal strength and the user's driving speed. Conventionally the faster the user's speed is, the lower bit rate the user can achieve. The bit rate can basically keep fixed while the user's speed does not change too much. Therefore, for each specified user we can approximately model the bit rates of APs as square waves. As Fig. 2 shows, we allow nonuniform AP deployments along any user's driving trajectory which include effective ranges, neighbor distances, and effective

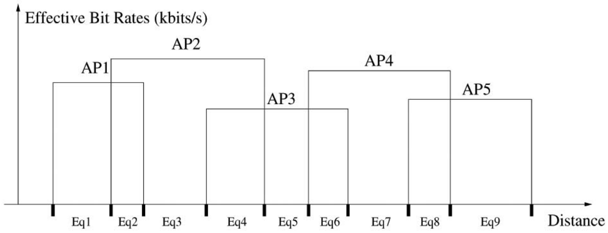


Fig. 2. Nonuniform AP deployments along user's driving trajectory.

bit rates. We then divide these regions into nonoverlapping *Equivalence Classes* as $Eq_1, Eq_2, \ldots, Eq_n$. Each Eq_i denotes a section of the roads, and within each section the candidate AP set and corresponding effective bit rates will keep fixed for the specified user.

3.2 Performance Metrics

We consider two performance metrics in this study: efficiency and fairness. Efficiency is measured with the overall throughput received by all users, and fairness is to regulate the association control so that all users have a fair distribution of bandwidth as much as possible.

For efficiency, we aim to maximize the overall throughput for all vehicular users. The throughput for any user is the average message delivery rate during the user's service period and it is usually measured in bits per second. Hence for any user j, given the service duration $[t_j, t_j + T_j]$ and the allocated bandwidth $b_j(t)$ at time $t \in [t_j, t_j + T_j]$, we can express the throughput B_j for user j as $B_j = \frac{1}{T_j}\int_{t_j}^{t_j+T_j} b_j(t)dt$. Consider the overall time interval $[0, T]$, during intervals $[0, t_j]$ and $[t_j + T_j, T]$, we actually have $b_j(t) = 0$, thus we have an equivalent uniform notion as $B_j = \frac{1}{T_j}\int_0^T b_j(t)dt$.

Association control without considering fairness may lead to the starvation of users with poor signal strength. To consider fairness, two metrics are used frequently in literature: max-min fairness [4] and proportional fairness [5]. Suppose the throughput allocation for all n users can be denoted as a vector $\vec{B} = \langle B_1, B_2, \ldots, B_n \rangle$. For max-min fairness, an allocation $\vec{B}$ is "*max-min fair*" if and only if an increase of any throughput within the domain of feasible allocations must be at the cost of a decrease of some already smaller throughput. For proportional fairness, an allocation $\vec{B}$ is "*proportionally fair*" if and only if, for any other feasible allocation $\vec{B}'$, $\sum_{j=1}^{|\vec{B}|} \frac{B_j' - B_j}{B_j} \leq 0$. In other words, any change in the allocation must have a negative average change. It has been proved that the unique proportionally fair allocation can be obtained by maximizing $J(\vec{B}) = \sum_j \ln(B_j)$ over the set of feasible allocations [15].

Since all APs are deployed by the same organization, a centralized control scheme is possible as proposed in [16]. Therefore based on the above models and assumptions, assuming we are the service provider of the specified region, we aim to build a centralized association control system and our goal is to continuously construct optimized assignments of APs to users as they are driving along the roads, respectively, taking the efficiency and fairness metrics into consideration. We consider both offline and online settings of the optimization problem. In the offline setting, we assume that we know the mobility patterns and trajectories of vehicular users in advance, in other words, we are given the candidate AP set $A_j(t)$ for each user j at each time $t \in [0, T]$ as part of the problem input. In the online setting, each $A_j(t)$ is revealed only at time t, at which time instant we have to instantaneously select an AP from $A_j(t)$ to associate for each user j, without any knowledge of the future sets $A_j(t')$ for $t' \in [t, T]$.

4 Overall Optimization and Snapshot Solution

For the efficiency metric, with a set of vehicular users U on the roads, the objective is to maximize $\sum_{j\in U} w_j B_j$, which can be further denoted as

$$\sum_{j\in U} \frac{w_j}{T_j} \int_0^T b_j(t)dt. \tag{1}$$

Here w_j denotes priority for different users, and it is a fixed value for user j. Similarly, if we choose proportional fairness as the metric, the optimization objective is to maximize $\sum_{j\in U} w_j \ln B_j$, which can be further denoted as

$$\sum_{j\in U} w_j \ln\left(\frac{1}{T_j}\int_0^T b_j(t)dt\right). \tag{2}$$

The above two objectives are optimization metrics over the duration of service period for all users. We use the term "long-term" to denote the overall time interval the user gets service. As we aim to continuously construct optimized assignments of users to APs within this duration, we use the term "snapshot" to denote the time instant within which we have to make a decision about AP association for all users. Thus, it is necessary for us to find solutions for each snapshot to achieve the overall optimal performance.

4.1 Snapshot Optimization for Efficiency

We first prove a theorem.

Theorem 1. *For the efficiency metric, it is sufficient to maximize $\sum_{j\in U} \frac{w_j}{T_j} b_j(t)$ for each snapshot t to achieve the long-term optimization goal.*

The proof of *Theorem 1* can be found in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17. *Theorem* 1 essentially tells us that we can optimize for efficiency metric in each snapshot to achieve overall performance. In the offline setting, we already know T_j in the objective function. In the online setting, we have to estimate T_j based on the user's current speed $v_j(t)$. Suppose user j gives the driving trajectory to the centralized server through devises like GPS. Knowing the overall distance S_j and the distance $s_j(t)$ that user j has traveled at time t, we can continuously estimate T_j for user j at snapshot t using $T_j(t) = \frac{S_j - s_j(t)}{v_j(t)} + t$. In case that the vehicular user stops at traffic lights, we maintain a window of speeds for recent k snapshots, and use the average speed $\overline{v_j(t)}$ to estimate $T_j(t)$.

To describe the constraints in this problem formulation for each snapshot t, we formulate the association problem into a linear program (LP) as proposed in [5]. We use a fractional variable $p_{i,j}(t)$ to denote the fraction of time that AP i devotes to user j. For each AP i and user j, if j is associated with i, then $p_{i,j}(t)$ is a fraction between 0 and 1; if user j is not associated with i, then the fraction is 0. Since each user j is assigned to only one AP for the integral solution, there is exactly one nonzero $p_{i,j}(t)$ for each $i \in A$. We can first relax this constraint and assume that one user can associate with multiple APs for the fractional solution. Then the bandwidth $b_j(t)$ allocated to each user j can be depicted as $b_j(t) = \sum_{i \in A} r_{i,j}(t) p_{i,j}(t)$. Thus, we can obtain a fractional solution from the following linear program formulation:

$$maximize \sum_{j \in U} \frac{w_j}{T_j} b_j(t), \tag{3}$$

subject to

$$\forall j \in U \quad b_j(t) = \sum_{i \in A} r_{i,j}(t) \cdot p_{i,j}(t), \tag{4}$$

$$\forall i \in A \quad \sum_{j \in U} p_{i,j}(t) \leq 1, \tag{5}$$

$$\forall j \in U \quad \sum_{i \in A} p_{i,j}(t) \leq 1, \tag{6}$$

$$\forall i \in A, j \in U \quad 0 \leq p_{i,j}(t) \leq 1, \tag{7}$$

$$\forall j \in U \quad b_j(t) \geq C. \tag{8}$$

Constraint (4) defines $b_j(t)$, the bandwidth allocated to user j at time point t. Constraint (5) means that the overall allocated time fraction of each AP i to all users cannot be more than 1. Constraint (6) states that the overall allocated time fraction of each user j that communicates with all APs cannot be more than 1. Constraint (7) shows that the time fraction is between 0 and 1. To ensure that every user is able to maintain connectivity to the internet within the service duration, (8) guarantees that every user has a minimum bandwidth of C at any time t, where C is a constant value for the lower bound. For the *pure efficiency* goal we set $C = 0$ by default.

For completeness, we describe briefly in the following how to find the integral solution based on the fractional solution. After we obtain $p_{i,j}(t)$ for each user-AP pair, we can further calculate the fractional assignment $x_{i,j}(t) = \frac{r_{i,j}(t) \cdot p_{i,j}(t)}{b_j(t)}$, which reflects the fraction of user j's total bandwidth that it expects to get from AP i. Apparently $0 \leq x_{i,j}(t) \leq 1$. We can view the assignment as a bipartite graph. Then, the final integral solution is a set of binary variables $\hat{x}_{i,j}(t)$ for all user-AP pairs, where $\hat{x}_{i,j}(t)$ is equal to 1 if user j is associated with AP i and 0 otherwise. We use the rounding algorithm proposed by Shmoy and Tardos [17] to calculate the integral solution $\hat{x}_{i,j}(t)$. Readers can refer to [5] for detailed description.

Since we have obtained the optimized strategy for association control over each snapshot, we need to consider the handoff strategy for efficiency. Continuously, computing the optimized association solution for each snapshot is definitely not an appropriate solution, as it incurs too much computing and communication cost. Without loss of generality, we assume that the boundaries of an AP's effective range will not coincide with the others. Fig. 3 shows an example of the vehicular scenario, where a set of users $U_1, U_2, \ldots, U_n$ are driving through various *Equivalence Classes* over the time span $[0, T]$. As the time intervals for each user to drive through the *Equivalence Classes* may overlap with each other, hence for ease of analysis we can further divide the overall time span $[0, T]$ into smaller time intervals according to the boundaries of *Equivalence Classes* over the time span. We denote these time intervals as $[T'_0, T'_1], [T'_1, T'_2], \ldots, [T'_{L-1}, T'_L]$, where $T'_0 = 0$ and $T'_L = T$. We rely on the following theorem to devise an efficient handoff strategy for efficiency metric.

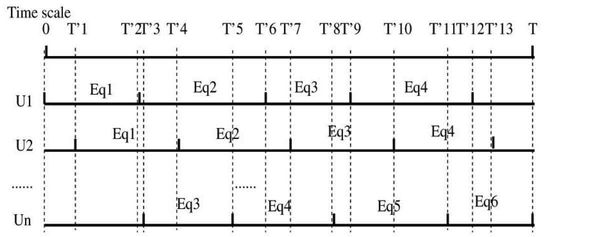


Fig. 3. Time line for users to drive through the *Equivalence Classes*.

Theorem 2. *For optimal association control to maximize the efficiency metric, handoffs to new association solutions for users only happen when at least one user is crossing the boundaries of* Equivalence Classes. *At each boundary the user will meet with one of the following cases: 1) new candidate AP is detected; 2) original optimal AP is lost; and 3) original candidate AP is lost. For cases 1) and 2), new association control is necessary. For case 3), new association control is not needed, so the original optimized solution holds.*

The proof of *Theorem 2* can be found in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17. According to *Theorem* 2, for the efficiency metric we only have to compute the optimized association solution each time when one or more users cross the boundary of *Equivalence Classes*. We can further prevent unnecessary computation by checking the special patterns of adjacent *Equivalence Classes*.

4.2 Online Algorithm for Proportional Fairness

In the above section, we have demonstrated that for the efficiency metric we can transform the long-term overall optimization into the snapshot optimization. However, for proportional fairness, as each snapshot decision for the optimal solution may depend on its former and future situations, we cannot simply conduct this transformation.

We know that the exact optimal solution can only be achieved with information obtained over the whole time span $[0, T]$ in advance. However, in practice we cannot precisely know the users' future mobility trajectory, thus no information about which users will be contending for specified APs in the future can be obtained beforehand. In this section, according to the online setting described in the end of Section 3, we design an online algorithm. Our solution relies on the following theorem.

Theorem 3. *Maximizing the long-term objective function*

$$\sum_{j\in U} w_j \ln\left(\epsilon + \int_0^T b_j(t)dt\right), \tag{9}$$

is consistent with maximizing the long-term objective function

$$\int_0^T \sum_{j\in U} \frac{w_j}{\epsilon + \int_0^t b_j(t)dt} b_j(t)dt.$$

Here, $\int_0^t b_j(t)dt$ denotes the accumulated bandwidth in time span $[0, t]$, w_j denotes the original fixed weight as priority for each user j, and $\epsilon > 0$ is a small constant number.

The proof of *Theorem 3* can be found in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17. Recall that the original long-term goal is to maximize

$$\sum_{j\in U} w_j \ln\left(\int_0^T b_j(t)dt\right). \tag{10}$$

The only difference between the above two objective functions (9) and (10) is ϵ, which may have an impact on the corresponding optimal solution. However, as long as we set ϵ small enough ($\epsilon \to 0$) in (9), the long-term goal in (9) becomes very near to $\sum_{j\in U} w_j \ln(\int_0^T b_j(t)dt)$.

In order to maximize

$$f_O = \int_0^T \sum_{j\in U} \frac{w_j}{\epsilon + \int_0^t b_j(t)dt} b_j(t)dt,$$

we use the following heuristic snapshot objective

$$f'_O(t) = \sum_{j\in U} \frac{w_j}{\epsilon + \int_0^t b_j(t)dt} b_j(t),$$

along with the constraint depicted in (4)-(8) to approximate the long-term optimization solution. The intuition is that maximizing $f'_O(t)$ at each t contributes to the maximization of f_O. We thus propose an algorithm based on the dynamic weight

$$W_j(t) = \frac{w_j}{\epsilon + \int_0^t b_j(t)dt}.$$

Since it is possible that $\int_0^t b_j(t)dt = 0$, we let $\epsilon > 0$ to prevent $W_j(t)$ from equal to $+\infty$. This online algorithm called Dynamic Weight based Online Algorithm (DWOA) is illustrated in Algorithm 1. Here, Line 3 takes care of the fairness metric by setting $W_j(t)$ inversely proportional to the accumulated bandwidth. Line 4 considers the efficiency metric by attempting to maximize the sum of the weighted bandwidths. We update the association solution for every Δt time interval. Conventionally the less Δt we use, the better solution we can obtain, but the drawback is that it may cause too many handoffs. In the supplementary file, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17, we further provide the performance analysis of DWOA and introduce the offline optimization for proportional fairness.

Algorithm 1 DWOA: Dynamic Weight based Online Algorithm

1: $t = 0$
2: **while** $t < T$ **do**
3: For each user j, calculate $W_j(t) = \frac{w_j}{\epsilon + \int_0^t b_j(t)dt}$ for snapshot at t.
4: For snapshot at t, set the object function as to maximize $\sum_{j\in U} W_j(t) b_j(t)$, calculate and apply the solution for association.
5: $t = t + \Delta t$.
6: **end while**

4.3 Online Algorithm for Max-Min Fairness

Since max-min fairness is also a frequently used metric for fairness, for completeness, in this section we consider the approach to achieve long-term max-min fairness. The intuition of max-min fairness for the *Drive-thru Internet* scenario is to *maximize* the throughput allocated to those users that receive the *minimum* throughput. As it has been proved by Bejerano et al. [4] that the problem of finding a max-min fair integral association is NP-hard, thus we consider an online algorithm to approximately achieve the max-min fairness.

Assume at each snapshot t, each user j has his accumulated bandwidth $\int_0^t b_j(t)dt$ and current service duration $T_j(t)$. We define a user j to be *saturated* when $\sum_{i\in A} p_{i,j} = 1$ or $\forall i \in A_j, \sum_{j'\in U} p_{i,j'} = 1$. In other words, a user j is saturated only when j has used all his time fraction to connect to APs or no remaining time fraction of his candidate APs can be further allocated to j. Algorithm 2 illustrates the online algorithm to achieve long-term max-min fairness. We update the association solution for every Δt time interval. For each snapshot t, we sort the users according to nondecreasing order of

$$y_j = \frac{\int_0^t b_j(t)dt}{w_j \cdot T_j(t)},$$

which is the current allocated throughput normalized by the weight, and we denote the reordered users as $1, 2, \ldots, j, \ldots, n$. Then we try to allocate fractional resource $p_{i,j}(t)(i \in A)$ to each user j in a progressive filling approach. We start from user 1 and try to allocate resource to user 1 as much as possible until

$$y'_1 = \frac{\int_0^t b_1(t)dt + \sum_{i\in A} r_{i,j}(t) \cdot p_{i,j}(t) \cdot \Delta t}{w_1 \cdot T_1(t)} = y_2$$

or user 1 is saturated with available resources. We further allocate resource to user 1 (if user 1 is not yet saturated, otherwise we stop allocation for user 1) and user 2 as much as possible until any of the users is saturated or $y'_1 = y'_2 = y_3$. We continue this progressive filling procedure until all available resources have been allocated or all of the users are saturated. For each round we use parameter J to denote the indices for the involved users $1, 2, \ldots, J$ in progressive filling.

Algorithm 2 Using progressive filling method to achieve max-min fairness

$t = 0$
while $t < T$ **do**
  For each snapshot t, sort the users according to non-decreasing order of $\frac{\int_0^t b_j(t)dt}{w_j \cdot T_j(t)}$. Denote the re-ordered users as $1, 2, ..., j, ..., n$.
  Initialize $p_{i,j}(t) = 0$ for all i, j pairs, initialize the saturated user set $S = \emptyset$.
  $J = 1$
  while $J \leq n$ **do**
    Call Algorithm 3 to progressively allocate additional fractional resource $p_{i,j}(t)$ to each nonsaturated user j ($j \leq J$), update the saturated user set S.
    $J = J + 1$
  end while
  Calculate the integral solution by using rounding algorithm.
  Update $\int_0^t b_j(t)dt$ and $T_j(t)$ for each user j.
  $t = t + \Delta t$
end while

Algorithm 3 is the optimized allocation algorithm for each progressive filling round. During each round we utilize the linear program **LP(**U'**)** to iteratively determine the saturated users while trying to maximize the throughput of these saturated users, as shown in the following formulation:

$$maximize \quad y, \tag{11}$$

subject to

$$\forall j \in U' \quad y_j = \frac{\int_0^t b_j(t)dt + \sum_{i \in A} r_{i,j}(t) \cdot p'_{i,j}(t) \cdot \Delta t}{w_j \cdot (T_j(t) + \Delta t)}$$

$$\forall j \in U' \quad y_j \geq y, \tag{12}$$

$$\forall j \in U' \quad y_j \leq \frac{\int_0^t b_J(t)dt}{w_j \cdot (T_J(t) + \Delta t)}, \tag{13}$$

$$\forall i \in A \quad \sum_{j \in U} (p_{i,j}(t) + p'_{i,j}(t)) \leq 1, \tag{14}$$

$$\forall j \in U' \quad \sum_{i \in A} (p_{i,j}(t) + p'_{i,j}(t)) \leq 1, \tag{15}$$

$$\forall i \in A, j \in U' \quad 0 \leq p_{i,j}(t) + p'_{i,j}(t) \leq 1. \tag{16}$$

In the linear programming **LP(**U'**)**, U' denotes the nonsaturated user set, we, respectively, use $p'_{i,j}(t)$ and $p_{i,j}(t)$ to denote the new allocated time fraction and the already allocated time fraction. The objective is to maximize y, which is the minimum value of y_j for each user j inside the nonsaturated user set U'. Constraint (12) depicts that y is the minimum value of y_j. Constraint (13) depicts that y_j should not be allocated more than the value of

$$\frac{\int_0^t b_J(t)dt}{w_j \cdot (T_J(t) + \Delta t)},$$

since we are considering the progressive filling approach for users in $1, 2, \ldots, J$ for each round. Constraint (14) means that the overall allocated time fraction of each AP i to all users cannot be more than 1. Constraint (15) states that the overall allocated time fraction of each user j that communicates with all APs cannot be more than 1. Constraint (16) shows that the time fraction is between 0 and 1.

Algorithm 3 Calculate optimized allocations iteratively while maximizing the throughput of the saturated users

$U' = \{1, 2, ..., J\} - S.$
while $U' \neq \emptyset$ **do**
  Conduct the linear programming **LP(**U'**)** to figure out the optimal parameters for $p'_{i,j}(t)$.
  Update the following parameters

$$\forall i \in A, j \in U \quad p_{i,j}(t) = p_{i,j}(t) + p'_{i,j}(t)$$

$$\forall j \in U, \quad \int_0^t b_j(t)dt = \int_0^t b_j(t)dt + \sum_{i \in A} r_{i,j}(t) \cdot p'_{i,j}(t) \cdot \Delta t$$

  for each user j **do**
    if $\sum_{i \in A} p_{i,j} = 1$ or $\forall i \in A_j, \sum_{j \in U} p_{i,j} = 1$ **then**
      Add j into S.
    end if
  end for
  $U' = U' - S$
end while

In each iteration of Algorithm 3, we attempt to determine the optimal allocations for the users which are still nonsaturated, and check current users if they are already saturated according to the definition. Then we remove the saturated users from the target set U'. In this way we achieve the *"max-min fair"* allocation in an approximate approach by using the progressive filling method.

In the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17, we further introduce the group-based methodology to reduce association control complexity, and discuss the practical utilization in realistic settings.

5 PERFORMANCE EVALUATIONS

We have implemented a simulator to simulate the Drive-thru Internet scenario. In order to simulate the realistic settings about the road topologies and the vehicles' moving traces, we use the realistic traffic generator Simulation of Urban MObility (SUMO) [18] to construct the large road network and generate vehicle traffic. In the simulation, we build the road topologies based on the road networks in a rectangular region (3,500 m $\times$ 3,000 m) in Washington DC., which is imported from the TIGER database [19]. We build the roads with lanes, and the number of lanes of each road ranges from 1 to 6. Hence, the major roads take the majority of traffic flow as there are more lanes on the major roads than the other roads. Traffic lights are deployed at the cross of roads. We generate vehicle traffics over the road networks according to the parameters illustrated in Table 1, where $accel$ and $decel$, respectively, denote the acceleration and deceleration ability of vehicles, $sigma$ denotes the driver imperfection (between 0 and 1), $length$ and $speed$, respectively, denote the vehicle length and average speed, $density$ denotes the average density of vehicles on the roads. The average speed $speed = 15$ m/s and we observe that the 95 percent confidence interval for vehicles' speed is

TABLE 1
Parameters for Generating Vehicle Traffics

parameters	default value	parameters	default value
$accel$	$0.8(m/s^2)$	$length$	5 (m)
$decel$	$4.5(m/s^2)$	$speed$	15 (m/s)
$sigma$	0.5	$depart$	0 (s)
$period$	30 (s)	$repno$	100
$density$	22 (users/km)		

TABLE 2
Abbreviations

Abbreviation	Full Name
SSF	Strongest Signal First
CUB	Connect Until Broken
OPT-E(offline)	Offline optimization for efficiency
OPT-E(online)	Online optimization for efficiency
OPT-PF(offline)	Offline optimization for proportional fairness
OPT-PF(online)	Online optimization for proportional fairness
OPT-MM	Online optimization for max-min fairness

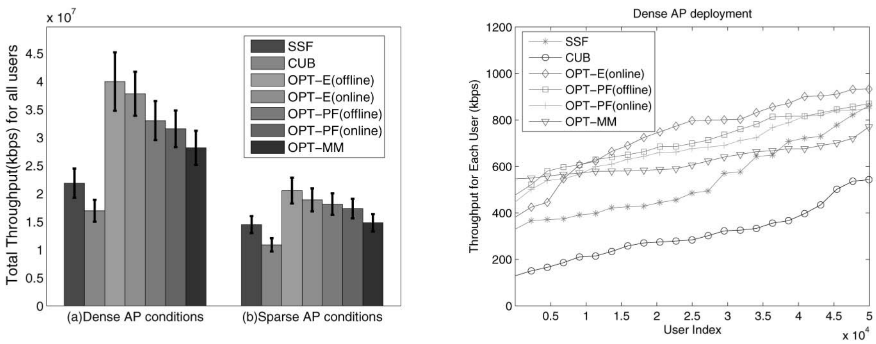
$(5\ \mathrm{m/s}, 20\ \mathrm{m/s})$. We simulate 500 various routes over the road network for the vehicles, for each route we randomly pick the origin position and the destination position and hence specify the trip for the route. We use $depart$ to denote the time at which the vehicles are emitted into the network, and we use $period$ to denote the average time interval after which another vehicular user with the same route shall be emitted and $repno$ to denote the number of vehicles to emit which share the same route. We use random seeds to generate the time intervals between emitted vehicles with the same route. As we set $period = 30\ \mathrm{s}$, thus on average every 30 s the vehicles are emitted, and we set the 95 percent confidence interval of the time interval as $(20\ \mathrm{s}, 40\ \mathrm{s})$. Therefore, the application scenario involves about 50,000 vehicular users and lasts about 50 minutes. According to the above settings, the average density $density = 22$ users/km, i.e., the average number of vehicles per kilometer of road is 22. We observe that the 95 percent confidence interval for the density is (14 users/km, 60 users/km).

We conduct the performance evaluation based on the settings of large road network and vehicle traffic generated by SUMO. We randomly place the APs inside the specified region and adopt the experiment results from [1] to simulate the effective bit rates of APs. We set their peak bit rates within the range from 4,000 to 5,000 kbps for vehicular users. To sufficiently evaluate the performance of various association control strategies, we consider two kinds of situations for AP deployment: the dense AP deployment and the sparse AP deployment. For the dense AP deployment, we randomly deploy 500 APs and make sure that at any location of the roads the user is within effective range of at least one AP. For the sparse AP deployment, we randomly deploy 150 APs and the user is not guaranteed within the effective range of at least one AP at any location of the roads. In the following part, we conduct the performance evaluation to show how the optimal solutions work under the two different situations. To obtain each simulation result, we take the average value of 50 simulation runs. In the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2011.17, we illustrate more detail simulation results in a comprehensive approach.

5.1 Efficiency and Fairness

In this section, we evaluate the performance in terms of efficiency and fairness. In order to illustrate the performance gains of our optimized solutions, we compare with two heuristic strategies. The first strategy is *Strongest Signal First*, which always associates a user with the AP yielding the strongest received signal strength at all times. The second strategy is *Connect Until Broken*, which maintains a connection with a user and an AP until the user considers the link to be broken. Upon disconnection, the user will be associated with a new AP which yields the largest signal strength. When calculating the optimized solution, we solve the linear program and convex program using MATLAB. In the rest of this paper, we use the abbreviations as shown in Table 2 to denote the specified solutions. For the ease of comparison, we set $w_j = 1$ for each user. We set $\epsilon = 1$ kbits for *OPT-PF(online)*, and, respectively, set $C = 200$ kbps and $C = 0$ kbps in the dense AP situation and sparse AP situation for both *OPT-E(offline)* and *OPT-E(online)*.

Fig. 4a depicts the total throughput of all users achieved by various solutions. As in each run of simulation the generated traffic mobility has some variances, in order to show the statistical performance results, so we provide the



(a) (b)

Fig. 4. Simulation results for efficiency and fairness. (a) Total throughput (kbps) for all users. (b) Per-user throughput comparison with dense AP deployment.

90 percent confidence interval for the total throughput. Note that in both the dense AP situation and sparse AP situation, *OPT-E(offline)* achieves the largest overall throughput, while *CUB* achieves the smallest throughput. Due to some nonpredictable issues, *OPT-E(online)* achieves a little smaller value for the overall throughput than *OPT-E(offline)*. All solutions in the sparse AP situation achieves a fairly smaller overall throughput compared to the dense AP situation, as fewer APs are available for association to provide sufficient throughput for users. We observed that in both situations the optimized solutions outperform the two heuristic solutions. In the dense AP situation *OPT-E(online)*, respectively, achieves 72.9 and 122.9 percent more throughput than *SSF* and *CUB*, while in the sparse AP situation *OPT-E(online)*, respectively, achieves 30.6 and 73.7 percent more throughput than *SSF* and *CUB*. Since users have more candidate APs to associate with in the dense AP situation, there exist more opportunities for an optimized solution to achieve more performance gains.

In order to show the performance comparison in terms of fairness, Fig. 4b illustrates per-user throughput comparison in the dense AP situation. The X-axis is the user index and the Y-axis is users' throughput in kbps. The users are sorted by their throughput in increasing order. The throughput of the user with the same x index actually indicates the average throughput of the xth lowest throughput user (users allocated the xth lowest bandwidth). In the dense AP situation, we observe that the optimized solutions *OPT-E(online)*, *OPT-PF(offline)*, *OPT-PF(online)* and *OPT-MM* all outperform the two heuristic solutions *SSF* and *CUB*. For instance, the median indexed user's bandwidth value of *OPT-E(online)* is, respectively, 64 percent higher than *SSF* and 181 percent higher than *CUB*. *OPT-PF(offline)*, *OPT-PF(online)* and *OPT-MM* have better performance in fairness than *OPT-E(online)*, since the users with lower indices have higher throughput in *OPT-PF(offline)*, *OPT-PF(online)* and *OPT-MM* compared to *OPT-E(online)*. Among the three solutions, *OPT-MM* achieves the best performance in fairness, as the users with lower indices are higher than all the other solutions, inferring that more fairness is achieved among the users. The performance gains of the above optimized solutions are between 200 and 400 kbps in throughput for each user.

6 Conclusion

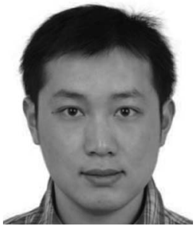
In this paper, we conduct a theoretical study on association control over the *Drive-thru Internet scenario*. We, respectively, consider efficiency and fairness as the optimization metrics. Due to issues concerning both technology and privacy, the data needed to compute the optimal solutions are currently not easy to gather. Hence, our present research work intends to be a theoretical effort to determine the upper bounds to what can be achieved in reality.

Acknowledgments

The authors would like to thank the reviewers for their careful reading and comments. This work is partially supported by the National Natural Science Foundation of China under Grant Nos. 61073028 and 61021062; the National Basic Research Program of China (973) under Grant No. 2009CB320705; the Technology Support Program of Jiangsu Province under Grant No. BE2010179; the US National Science Foundation (NSF) under Grant Nos. CNS-0721443 and CNS-0831904, CAREER Award CNS-0747108, CNS-0434533, CNS-0531410, and CNS-0626240.

References

[1] J. Ott and D. Kutscher, "Drive-thru Internet: IEEE 802.11b for 'Automobile' Users," *Proc. IEEE INFOCOM*, 2004.
[2] J. Camp and E. Knightly, "Modulation Rate Adaptation in Urban and Vehicular Environments: Cross-Layer Implementation and Experimental Evaluation," *Proc. ACM MOBICOM*, 2008.
[3] L. Tassiulas and S. Sarkar, "Maxmin Fair Scheduling in Wireless Ad Hoc Networks," *IEEE J. Selected Areas in Comm.*, vol. 23, no. 1, pp. 163-173, Jan. 2005.
[4] Y. Bejerano, S.-J. Han, and L. Li, "Fairness and Load Balancing in Wireless LANs Using Association Control," *IEEE/ACM Trans. Networking*, vol. 15, no. 3, pp. 560-573, June 2007.
[5] L. Li, M. Pal, and Y.R. Yang, "Proportional Fairness in Multi-Rate Wireless LANs," *Proc. IEEE INFOCOM*, 2008.
[6] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden, "A Measurement Study of Vehicular Internet Access Using In Situ Wi-Fi Networks," *Proc. ACM MOBICOM*, 2006.
[7] J. Eriksson, H. Balakrishnan, and S. Madden, "Cabernet: Vehicular Content Delivery Using Wi-Fi," *Proc. ACM MOBICOM*, 2008.
[8] A. Giannoulis, M. Fiore, and E.W. Knightly, "Supporting Vehicular Mobility in Urban Multi-Hop Wireless Networks," *Proc. ACM Mobile Systems, Applications, and Services (MOBISYS)*, 2008.
[9] R. Mahajan, J. Zahorjan, and B. Zill, "Understanding WiFi-Based Connectivity," *Proc. ACM Internet Measurement Conf. (IMC)*, 2007.
[10] D. Hadaller, S. Keshav, T. Brecht, and S. Agarwal, "Vehicular Opportunistic Communication under the Microscope," *Proc. ACM Mobile Systems, Applications, and Services (MOBISYS)*, 2007.
[11] V. Navda, A.P. Subramanian, K. Dhanasekaran, A. Timm-Giel, and S.R. Das, "Mobisteer: Using Steerable Beam Directional Antenna for Vehicular Network Access," *Proc. ACM MOBISYS*, 2007.
[12] M. Kim, Z. Liu, S. Parthasarathy, D. Pendarakis, and H. Yang, "Association Control in Mobile Wireless Networks," *Proc. IEEE INFOCOM*, 2008.
[13] P. Deshpande, A. Kashyap, C. Sung, and S. Das, "Predictive Methods for Improved Vehicular Wifi Access," *Proc. ACM Mobile Systems, Applications, and Services (MOBISYS)*, 2009.
[14] H. Wu, K. Tan, Y. Zhang, and Q. Zhang, "Proactive Scan: Fast Handoff with Smart Triggers for 802.11 Wireless Lan," *Proc. INFOCOM*, 2007.
[15] T. Nandagopal, T.-E. Kim, X. Gao, and V. Bharghavan, "Achieving Mac Layer Fairness in Wireless Packet Networks," *Proc. ACM MOBICOM*, pp. 87-98, 2000.
[16] R. Murty, J. Padhye, A. Wolman, and B. Zill, "Designing High Performance Enterprise Wi-Fi Networks," *Proc. ACM Networked Systems Design and Implementation (NSDI)*, 2008.
[17] D.B. Shmoys and E. Tardos, "An Approximation Algorithm for the Generalized Assignment Problem," *Math. Programming*, vol. 62, no. 3, pp. 461-474, 1993.
[18] "Sumo," http://sourceforge.net/apps/mediawiki/sumo/, 2011.
[19] "Tiger," www.census.gov/geo/www/tiger/, 2011.



Lei Xie received the BS and PhD degrees from Nanjing University in 2004 and 2010, respectively, all in computer science. He is an assistant professor in the Department of Computer Science and Technology, Nanjing University, China. His research interests include sensor networks, RFID systems, vehicular networks, cognitive radios, and high performance computing. He is a member of the IEEE and the IEEE Computer Society.

Qun Li received the PhD degree in computer science from Dartmouth College. He is an associate professor in the Department of Computer Science at the College of William and Mary. His research interests include wireless networks, sensor networks, RFID, and pervasive computing systems. He received the NSF Career award in 2008. He is a member of the IEEE and the IEEE Computer Society.

Weizhen Mao received the PhD degree in computer science from Princeton University under the supervision of Andrew C.-C. Yao. Her current research interests include the design and analysis of online and approximation algorithms for problems arising from application areas, such as multicore job scheduling, wireless and sensor networks. She is currently an associate professor in the Department of Computer Science at the College of William and Mary.

Jie Wu is the chair and a professor in the Department of Computer and Information Sciences, Temple University. Prior to joining Temple University, he was a program director at National Science Foundation. His research interests include wireless networks and mobile computing, routing protocols, fault-tolerant computing, and interconnection networks. He has published more than 500 papers in various journals and conference proceedings. He serves in the editorial board of the *IEEE Transactions on Computers*, *IEEE Transactions on Mobile Computing*, and *Journal of Parallel and Distributed Computing*. He is program cochair for IEEE INFOCOM 2011. He was also general cochair for IEEE MASS 2006, IEEE IPDPS 2008, ACM WiMD 2009, and IEEE/ACM DCOSS 2009. He also served as panel chair for ACM MobiCom 2009. He has served as an IEEE computer society distinguished visitor. Currently, he is the chair of the IEEE Technical Committee on Distributed Processing (TCDP) and ACM distinguished speaker. He is a fellow of the IEEE.

Daoxu Chen received the BS degree in computer science from Nanjing University, Nanjing, China, in 1982. He was a visiting researcher with Purdue University, West Lafayette, IN, and the City University of Hong Kong, Kowloon, Hong Kong. He is currently a professor with the Department of Computer Science and Technology, Nanjing University, Nanjing, China. His research interests include distributed computing, parallel processing, and computer networks. He is a member of the Association for Computing Machinery and a Senior Member of the China Computer Federation. He is a member of the IEEE and the IEEE Computer Society.