Ad Hoc Networks 12 (2014) 115–129

Contents lists available at SciVerse ScienceDirect

Ad Hoc Networks

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

Throughput optimization in wireless local networks with inter-AP interference via a joint-association control, rate control, and contention resolution

Ka-Lok Hung, Brahim Bensaou *

Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 22 June 2011
Accepted 24 August 2011
Available online 29 September 2011

Keywords:
AP association
Rate control
Contention resolution
Hidden terminals
Mathematical programming

ABSTRACT

The dense deployment of wireless access points (APs) either in wireless local area networks (WLANs) or in wireless mesh networks facilitates greatly ubiquitous Internet access, however, due to the limited number of orthogonal frequency channels allotted to the IEEE802.11-based networks, this also induces the inevitable problem of inter-AP interference. In this paper, we study the problem of determining the optimal association in multi-cell or extended wireless networks in the presence of hidden terminals and inter-AP collisions. Unlike most previous work in this area, which deals with networks without inter-AP interference, we first reveal that association control alone is not sufficient to achieve fair throughput allocation and load balancing across APs, then advocate a solution based on the joint association control, rate control and fair contention resolution as a means to improving network performance. Based on this, we formulate a cross-layer association control, throughput optimization and contention resolution problem whose objective is to allocate downlink throughput according to the proportional fairness principle. As the problem turns out to be a non-convex mixed integer programming problem, known to be NP-hard, we relax it first into a continuous problem, then transform the resultant into a convex problem and finally propose a distributed algorithm to solve it. We then design a simple yet effective approximation algorithm to recover an optimal solution that fulfills the discrete integral association constraints. The algorithm yields the optimal association, the maximum achievable rate for each downlink flow as well as each AP's optimal average backoff time. Using these results as settings in a network, we can achieve the optimal operation point without any scheduling. Numerical experiments and simulation results show that our algorithm converges rapidly and works effectively.

1. Introduction

IEEE 802.11 wireless local area networks (WLANs) have gained tremendous popularity in recent years, so much so that innumerable APs have been deployed and often without planning. Typically, a client station in WLANs must select and associate with one AP (among those available) before it can start communicating and/or accessing the Internet via this AP. Today, the most commonly used AP selection metric or criterion is the strength of the signal received from the AP as indicated by the received signal strength indicator or RSSI. However, it is well known that RSSI-based association may create uneven network loading between basic service sets (BSSs) in extended networks, and fails to achieve the best performance.

* Corresponding author.
E-mail addresses: ether@cse.ust.hk (K.-L. Hung), brahim@cse.ust.hk (B. Bensaou).

1570-8705/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2011.08.004

To ensure a good quality of network coverage, APs are usually deployed such that they cover partially overlapped areas. However, due to the limited number of orthogonal channels in the ISM frequency band, inter-AP interference is inevitable and the hidden terminal problem arises [35] between APs, especially in indoor environments, where walls an/or obstacles impose a denser deployment of the APs. Hidden node collisions deeply impact network performance in multi-cell WLANs [14]. During the AP selection process, it is desirable that possible hidden node collisions are taken into account, to, for example, intelligently associate clients to one of their accessible APs such that the number of hidden nodes is reduced. However, although association control can be shown to improve the network performance by balancing the load between APs and reducing possible hidden node collisions, as we will show in body of the paper, it may fail to address the throughput unfairness that results from spatial constraints on the APs and the clients in some cases. APs which are not affected by hidden node collisions gain spatial advantage and obtain most of the channel air time. Nearby BSSs may thus suffer from starvation. To overcome this problem, BSSs which are at an advantage should control their transmission rate and channel access probability to relinquish some access opportunities to other BSSs. Therefore, it is essential to deploy association control jointly with rate control and contention resolution for fair spatial bandwidth distribution in the network in general.

In this paper, we propose a cross-layer approach to optimize downlink throughput and fairness performance in multi-cell WLANs with a fully distributed algorithm. We study how to jointly control the association, transmission rate and contention resolution by formulating the problem as an optimization problem. Unlike previous recent association optimization studies [4,19,25,36] which mainly focus on providing load balancing and fairness from a theoretic point of view, without considering MAC layer contention, we include inter-AP interference and the MAC layer interaction of CSMA/CA networks in our problem formulation. Practical issues like random backoff, hidden terminals, frame retransmission are also considered. To the best of our knowledge, such practical cross-layer optimization for multi-cell WLANs has never been studied before.

To tackle the problem, we first formulate the proportional-fair [17] throughput allocation problem as a non-convex integer programming problem which is NP-hard. To cope with the non-convexity and integral constraints of the problem, we relax the integer programming problem into a continuous problem, then transform the resulting problem into a convex one by solving a series of geometric programs. The obtained solution yields a set of fractional associations, maximum transmission rate for each downlink flow, as well as the optimal average backoff time of each AP. Based on this solution, we propose a simple yet accurate rounding procedure to round the fractional association into an integral one.

The rest of the paper is organized as follows. In Section 2, we motivate our study and introduce the challenges that we have to address. We briefly describe some related work in Section 3. We describe the system model in Section 4, and formulate our problem in Section 5. In Section 6, we present a distributed fractional algorithm and describe the associated rounding algorithm in Section 7. In Section 8, we evaluate our algorithm via numerical experiments and simulations. We finally conclude the paper in Section 9.

2. Motivation

Our goal here is to motivate the need for both rate control and contention resolution together with association control to achieve optimal network performance. We will notably illustrate, via examples from WLANs as well as wireless Mesh networks, how association control alone may not be sufficient.

Consider the two example networks shown in Fig. 1a and c. Each network comprises four APs and several clients. Dashed circles around the APs represent their respective transmission ranges and dashed lines represent possible association between APs and clients. To maximize spatial reuse, APs are placed in a way that they are not synchronized by the distributed coordination function (DCF) [1]. Without loss of generality, we assume the APs to operate on the same frequency channel and thus inter-AP interference exists. Clients receive saturated downlink traffic from remote sources via their respective APs. A client is vulnerable to hidden node collisions if it is located somewhere in the overlap region between coverage areas of two or more APs.

Bejerano et al. [4] and Li et al. [25] illustrated that association control can improve network performance by balancing the load among a set of non-interfering APs. In this section, we would like to highlight the extra complexity and challenges that arise when inter-AP interference is present.

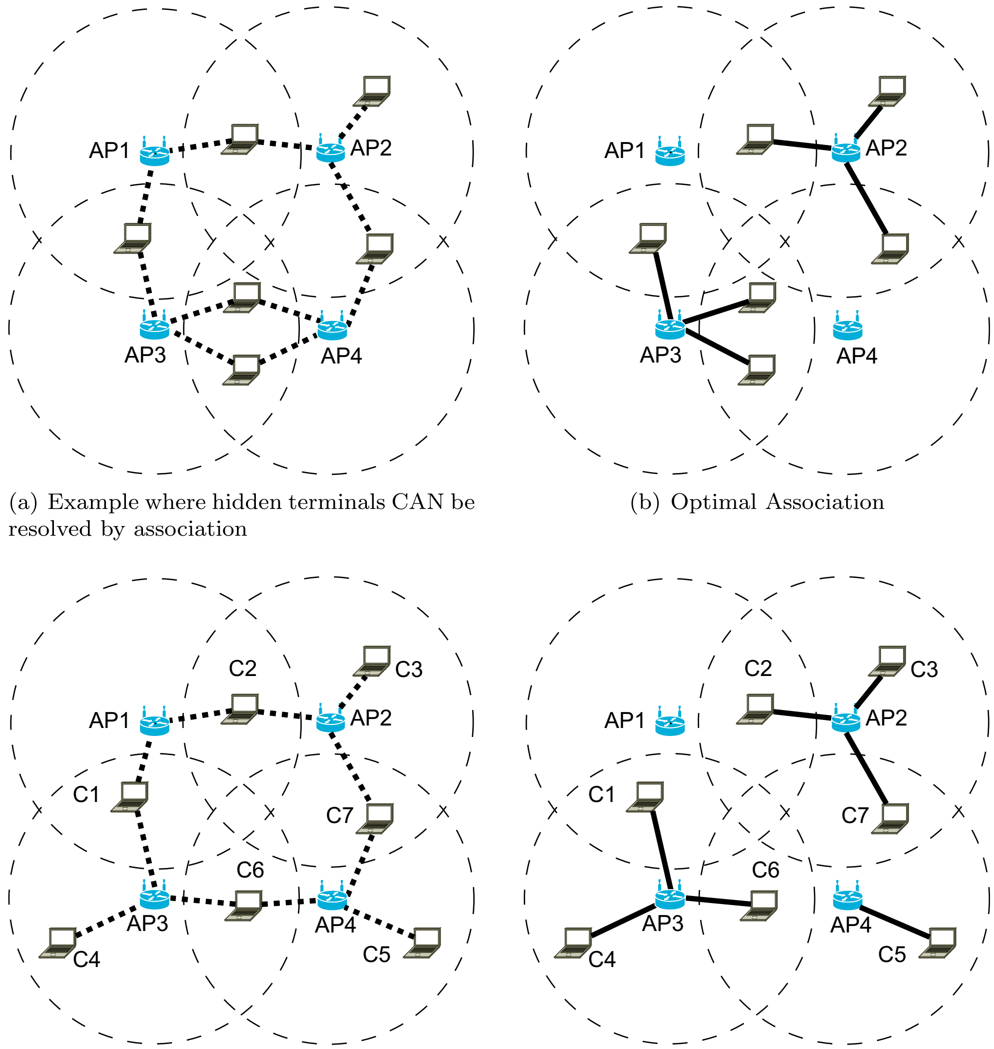
2.1. Removing hidden nodes by association

In the first example shown in Fig. 1a, there are five clients among which four have two possible associations. Clients should only associate with APs 2 and 3 as shown in Fig. 1b to avoid hidden node collisions and balance the load. APs 1 and 4 are not used and the solution shown is optimal in terms of throughput and fairness.

2.2. Cross-layer control

In the previous example, association control alone is capable of resolving the hidden terminal problem; however it may fail to do so in some other cases. Considering the network of Fig. 1c, clients 3–5 must associate with APs 2–4 respectively while clients 1 and 2 should not associate with AP 1 to prevent hidden node collisions. As a result, clients 6 and 7 are exposed to hidden node collisions for sure, and to minimize the severity of the problem, clients 6 and 7 should associate with APs 3 and 2 instead of AP 4 such that they only need to overcome collisions induced by the traffic flow from AP 4 to client 5. The optimal association for this case is shown in Fig. 1d.

Although association control reduced the amount of inter-AP interference, additional rate and contention



(a) Example where hidden terminals CAN be resolved by association

(b) Optimal Association

(c) Example where hidden terminals CANNOT be resolved by association

(d) Optimal Association does not achieve optimal rate: additional rate and contention control are needed

Fig. 1. Motivation of joint association, rate and contention control.

resolution are still needed to achieve an optimal fair throughput allocation. The transmission rate from AP 4 to client 5 should be limited to preserve a certain amount of airtime for clients 6 and 7. Also, the contention window size of APs 2 to 3 should be adjusted properly to cope with hidden node collisions. Under the IEEE 802.11 binary-exponential-backoff procedure, the average contention window size of APs 2 and 3 will tend to be much larger than that of AP 4 due to the hidden node collision. Without proper rate and contention control, elastic traffic like TCP which is very sensitive to packet drops and round-trip delay cannot be established between AP 2 (respect. 3) and client 7 (respect. 6) as we have shown previously in [14]. Clients 1–5 capture all the bandwidth. Whereas non-elastic traffic like UDP in which the source rate is fixed, not only clients 6 and 7 suffer but also clients 1–4 as the MAC layer transmission queue of APs 2 and 3 are filled up by the huge volume of retransmission traffic for clients 6 and 7. Packets of clients 1–4 are dropped due to the buffer lockout problem.

2.3. Periodic distributed optimization

In this paper, we tackle the problem by an offline distributed optimization algorithm. In practice, an association decision has to be made when a client arrives. However, without the user arrival patterns and complete profiles, an online algorithm may be harmful to the network as discussed in [25]. In Fig. 1c, if client 6 arrives before the others, without extra information, it is hard to decide that client 6 must not associate with AP 4. Therefore, a better approach to address the problem is to periodically optimize the network based on the current set of clients and

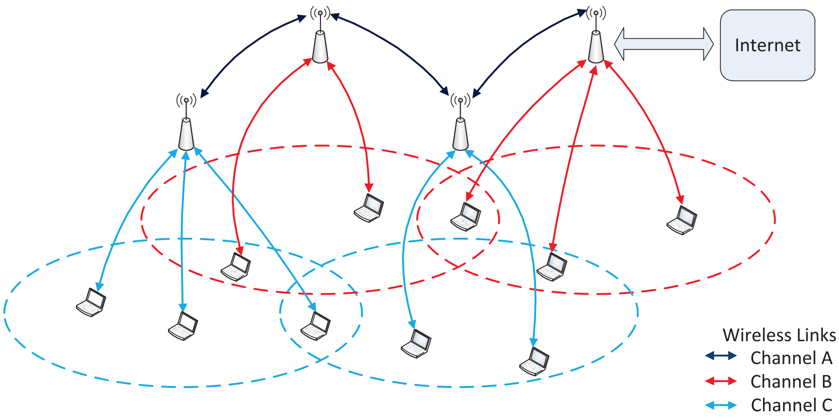


Fig. 2. The hidden-terminal scenario in a wireless mesh network.

their possible associations, and redeploy them to alternatives APs to achieve optimal performance periodically.

2.4. Relationship with the channel assignment problem

In the toy topology of Fig. 1, instead of performing a cross-layer optimization, one may solve the hidden terminal problem by assigning the three orthogonal channels available in the IEEE 802.11 standard to the APs intelligently. Such a channel assignment problem in wireless networks has been studied extensively [5,8,27,28,31]. Although channel assignment can improve network performance, beside being an NP-hard problem, recent studies [35] revealed that co-channel interference cannot be entirely eliminated by channel assignment in well managed production WLANs even with RF site surveys, indicating that other measures are needed to relieve the hidden terminal problem.

Moreover, the impact of the hidden terminal problem becomes even more severe in wireless mesh networks (WMNs). To reduce the infrastructure cost, WMNs often adopt a two tier hierarchy [10], the upper tier being the wireless back-haul network and the lower tier the local access network. The wireless back-haul is a multi-hop network consisting of mesh access points only, forwarding traffic back and forth between their clients and the Internet. The local access network is a set of BSSs similar to those of WLANs, established by the mesh access points and the clients. Clients join such a BSS to gain Internet access. To improve network throughput, multiple interfaces are often used, for instance, dual-interface mesh access points dedicate one interface to handle the back-haul traffic and the other to maintain the local access network. A simple linear wireless mesh networks formed by four such dual-interface mesh access points is depicted in Fig. 2. To preserve the connectivity, the same channel frequency is used for all back-haul interfaces, while it is desirable for the BSSs and the back-haul network to operate on different orthogonal channels to minimize co-channel interference between the two tiers. Furthermore, to enhance the spatial reuse, adjacent BSSs should also operate on different orthogonal channels. Given the three orthogonal channels (Channel A, B and C) of IEEE 802.11, the optimal channel assignment for this WMN is shown in Fig. 2.

Even with a careful channel assignment, we can clearly see that the hidden terminal problem between APs still exists in Fig. 2 due to the limited number of orthogonal channels. Increasing the number of radio interfaces may alleviate the problem [3,32,33]. However, as shown in [13], antennas on orthogonal channels have to be separated by at least 18 in. to prevent the transmission signal from interfering with the ongoing signal reception. It is a non-trivial task to fully exploit the three orthogonal channels with mesh access points equipped with three interfaces. In our mesh example, along with channel assignment, the cross-layer control aforementioned in Section 2.2 is essential to solve the hidden terminal problem.

Due to the problem complexity, we do not include channel assignment in our joint optimization problem in this paper. By intelligently assigning orthogonal channels to the APs, the number of hidden terminals can be reduced and thus, cushions the impact of hidden node collision. However, the channel assignment problem is known to be NP-hard even for finding an approximate solution. In our problem, the association control and channel assignment are two optimization dimensions which are interdependent. Different channel assignments lead to different problem formulations as the packet collision probability of a client depends on the co-channel interference coming from its surrounding APs. One can obtain the optimal solution by enumerating each possible channel assignment, solving the corresponding optimization problem and picking the configuration that results in the best overall performance. Such a brute-force approach is however indisputably infeasible in practice for production networks such as campus or enterprise networks because of the network size and the time required to locate the best configuration. Designing an effective heuristic that combines channel assignment with the cross-layer control that is the subject of the present paper is an interesting yet hard research problem left for a future paper.

3. Related work

Studies on association control fall into two categories: optimization-based and measurement-based.

In the optimization-based category, the association control problem is formulated as an optimization problem. Bejerano et al. propose centralized algorithms to determine AP association with the objective of allocating bandwidth according to the max–min principle [4]. Li et al. consider the proportional fair bandwidth allocation in multi-rate and multi-cell WLANs [25]. Two approximation algorithms are proposed to solve a linear and nonlinear association problem. Both [4,25] make use of the rounding procedure developed for job scheduling in [34] to approximate an integral association. Xie et al. [36] extend the problem formulation in [25] to the association problem of a vehicular network. Both offline and online association problems have been studied and corresponding centralized approximation algorithm are proposed. Koutsopoulos et al. study a joint channel assignment and AP selection problem in [19]. An algorithm that minimizes the number of required orthogonal channels for a specific loading is proposed.

As for the measurement-based approach, the association control decisions are based on the real-time measurement over some network parameters such as channel idle time or the size of the backlog at different APs. Lee et al. [21] propose a new association metric which estimates the available residual bandwidth of a WLAN. Clients make their own association decision solely based on this metric without external information from APs or other clients. The estimation is done by observing the changes of channel status from idle to busy and from busy to idle. Jardosh et al. present a queue-based AP association management system in [15]. When a new client joins the network, it issues an association request to an AP. If the network is heavily loaded, only the client at the head of the waiting queue of an AP is allowed to join the network and others have to wait for admission. The admitted user is served for a limited period of time. When such period expires, this user will be placed at the back of the queue and another client gains access. The maximum number of permissible users is determined by channel utilization measurement and the size of the queue length at the AP.

A considerable amount of work [7,9,20,11,16,22,26] has been done in the past to study the performance of IEEE 802.11 DCF. However, the performance impact upon the MAC layer throughput induced by inefficient association has seldom been studied. Optimization based association control algorithms proposed in [4,25,19] ignored the specific characteristics of the MAC layer in IEEE802.11 networks like the random backoff procedure and hidden node collisions. In this paper, we incorporate association control into the cross-layer control problem in which the influence on the MAC layer performance due to the client-AP association settings is included.

4. System model

Consider an IEEE 802.11 multi-cell network with several APs interconnected via a back-haul wired (or wireless) distribution network. Since we consider the effects of co-channel interference, we assume without loss of generality that these APs operate on the same channel. An AP provides a limited coverage over a certain area, and to strike a good balance between quality of connectivity and spatial reuse, the APs are placed such that their coverage areas overlap partially yet the APs themselves are not synchronized by the DCF similar to the example networks of Fig. 1. Previous studies [4,25,36] focus on the case where adjacent APs are assigned orthogonal channels, resulting in a total absence of inter-AP interference. In this paper, we drop this simplifying assumption and study the network performance with the consideration of such interference. Since APs are not synchronized by the DCF, hidden node collisions may occur.

Denote the set of wireless clients $\mathcal{C}$ and the set of APs $\mathcal{A}$ and assume that the clients are quasi-stationary as supported by the user mobility study in [18]. Each client receives saturated downlink traffic via its associated AP from the Internet. Denote by $\mathcal{A}_c$ the set of APs that are available to client c for association. Similarly, define $\mathcal{C}_a$ as the set of clients allowed to associate with AP a. We assume that each client should be able to communicate with at least one AP, i.e. $\mathcal{A}_c \neq \emptyset$. We focus on downlink performance in this paper as it is the primary target for clients.

When client c is receiving a data frame from AP a, we define $\mathcal{I}_{ac}$ as the set of APs that may interfere with the on going transmission. For example, considering the network of Fig. 1c, we have $\mathcal{I}_{3,1} = \{1\}$ as AP 1 may interfere with the traffic from AP 3 to client 1 due to hidden node collisions. We emphasize here that the definition of set $\mathcal{I}$ only depends on spatial relationships between clients and APs not on association. For example, in the optimal association of Fig. 1d, AP 1 does not cause any hidden node collision on client 1 due to the absence of association but we still have $\mathcal{I}_{3,1} = \{1\}$ based on the network topology.

4.1. AP activity model

We model the channel view of an AP as a renewal process and focus on the time spent by the AP transmitting or performing backoff. Let bo_a and T_a be the average backoff time and transmission time for a data frame at the MAC layer. The renewal period is equal to $bo_a + T_a$ where $T_a = T_H + L_a + T_{SIFS} + T_{ACK} + T_{DIFS}$ is defined as the channel occupancy time per MAC data frame. T_{DIFS} and T_{SIFS} are the durations of the protocol defined DCF inter-frame space and short inter-frame space respectively; T_H represents the transmission time for the physical and MAC layer header; T_{ACK} is the transmission time of an ACK frame; L_a is the transmission time of the payload. All these quantities are measured in units of the physical layer time slot σ. The normalized active time of AP a is $T_a/\{bo_a + T_a\}$ and the normalized idle time is $bo_a/\{bo_a + T_a\}$.

4.2. Client association

For each possible association between AP a and client c, we define a binary variable y_{ca} to indicate the association status. y_{ca} is equal to 1 if client c is associated with AP a otherwise y_{ca} is equal to 0. A client can only be associate with one AP at any time. Therefore, we have the constraints $\sum_{a\in\mathcal{A}_c} y_{ca}=1$ and $y_{ca}\in\{0,1\}$ for each client $c\in\mathcal{C}$.

4.3. Link throughput

Given that client c is associated with AP $a\in\mathcal{A}_c$, it receives a data packet from AP a successfully if the transmission starts when all APs in $\mathcal{I}_{ac}$ (Let's assume all APs in $\mathcal{I}_{ac}$ are servicing at least one client for the time being) are performing backoff and they remain in this status for the whole transmission period for $L_a + T_H$ slots. If AP n is in backoff stage at the current slot, the probability that it remains in backoff for the next slot is $1-1/bo_n$ on average [20]. Given AP a is going to transmit to client c, the conditional successful transmission probability s_{ac} is defined as

$$s_{ac}=\prod_{n\in\mathcal{I}_{ac}}\frac{bo_n}{bo_n+T_n}\times\left(1-\frac{1}{bo_n}\right)^{L_a+T_H}. \tag{1}$$

If client c is associated with AP a, we denote the fraction of time of AP a transmitting to client c as t_{ac}. If client c is not associated with AP a, we have $t_{ac}\times y_{ca}=0$. The throughput x_{ac} of link (a,c) can be expressed in terms of bo_a, s_{ac} and $t_{ac}\times y_{ca}$ as

$$x_{ac}=s_{ac}\times\frac{T_a}{bo_a+T_a}\times t_{ac}\times y_{ca}\times\frac{L_a}{T_a}\times R_a, \tag{2}$$

where R_a is the physical layer date rate of AP a, t_{ac} is a fraction between 0 and 1. The total downlink throughput allocated to client c is equal to $\sum_{a\in\mathcal{A}_c} x_{ac}$. We assume the average transmission time T_a is given.

5. Problem formulation

In this paper, we tackle the problem by an offline distributed optimization algorithm. In practice, an association decision has to be made when a client arrives. However, without the user arrival patterns and complete profiles, an online algorithm may be harmful to the network as discussed in [25]. In Fig. 1c, if client 6 arrives before the others, without extra information, it is hard to decide that client 6 should not associate with AP 4. Therefore, a better approach to address the problem is to periodically optimize the network based on the current set of clients and their possible associations, and redeploy them to alternatives APs to achieve optimal performance periodically.

We formulate the cross-layer throughput maximization problem as a mathematical optimization problem. The objective of the problem is to maximize the clients' downlink throughput while maintaining proportional fairness to ensure an acceptable spatial coverage. For this purpose we invoke the network utility maximization (NUM) framework [17]. We associate the total downlink throughput between an AP and a client with a strictly concave, increasing and twice differentiable utility function $U(\cdot)$ and maximize the aggregate utility. To achieve proportional fairness, the utility function is chosen as the log function [29]. Defining $\rho_a = T_a/bo_a$, we formulate the throughput optimization problem as

$$\begin{aligned}
\max\quad & \sum_{c\in\mathcal{C}}\log\left(\sum_{a\in\mathcal{A}_c} x_{ac}\right)\\
\text{s.t.}\quad & x_{ac}\leqslant s_{ac}b_a t_{ac}y_{ca}\times\frac{L_aR_a}{T_a},\quad \forall a\in\mathcal{A}_c,\ \forall c\in\mathcal{C},\\
& b_a\leqslant\frac{\rho_a}{1+\rho_a},\quad \forall a\in\mathcal{A},\\
& s_{ac}\leqslant\prod_{n\in\mathcal{I}_{ac}}\frac{1}{1+\rho_n}\left(1-\frac{\rho_n}{T_n}\right)^{T_H+L_a},\quad \forall a\in\mathcal{A}_c,\ \forall c\in\mathcal{C},\\
& \sum_{c\in\mathcal{C}_a} t_{ac}\leqslant 1,\quad \forall a\in\mathcal{A},\\
& 0\leqslant t_{ac},\quad \forall c\in\mathcal{C}_a,\ \forall a\in\mathcal{A},\\
& \sum_{a\in\mathcal{A}_c} y_{ca}=1,\quad \forall c\in\mathcal{C},\\
& y_{ca}\in\{0,1\},\quad \forall c\in\mathcal{C}_a,\ \forall a\in\mathcal{A},\\
& \rho_a^{min}\leqslant\rho_a\leqslant\rho_a^{max},\quad \forall a\in\mathcal{A}.
\end{aligned} \tag{3}$$

As we have seen in our first example network in Fig. 1a, some of the APs do not serve any client in the optimal association. As a consequence, when evaluating the successful transmission probability s_{ac}, this must be taken into account as it is possible that not all APs in $\mathcal{I}_{ac}$ will generate interference. Two approaches are possible, the first not considered here because it renders the problem intractable, consists into redefining (1) by incorporating a binary variable to indicate whether the AP generates interference or not. The second which is an accurate approximation, consists in setting the average backoff time bo_n to a very large constant when an AP is not associated, and then AP n's active time $T_n/(bo_n + T_n)$ trends to zero. As a consequence, we set ρ_n^{min} which is the lower bound of T_n/bo_n as a very small positive constant. As for ρ_n^{max}, we set it as $T_n/8$ as CW_{min} = 16 in IEEE 802.11g standard and the random backoff time is uniform. In our experiments, we find that an AP's average backoff time never exceeds the system limit, $CW_{max}/2$, if some clients are associated to it because of the proportional fairness constraint. ρ_a is equal to ρ_a^{min} only for APs without any client associated.

We represent the normalized activation time of AP a as b_a. The optimization variables in problem (3) are ρ_a, t_{ac} and y_{ca} as other variables can be expressed in terms of these. The fourth inequality ensures that the sum of the fractions of time that AP a spends on different clients does not exceed 1. Since x_{ac} is an increasing function of s_{ac}, t_{ac} and b_a, the equalities in the first four inequalities will always be satisfied at optimality.

Solving problem (3) yields the optimal association, the average backoff time as well as the throughput allocation. Given the average channel occupation time per data packet T_a, the optimal average backoff can be obtained by T_a/ρ_a. However, since the first three constraints are non-convex, problem (3) is a non-convex mixed integer programming problem which is NP-hard in general and a local optimum may not be the global optimum [24]. Therefore, instead of solving the optimization problem directly, we propose an approximation which overcomes the computational burden due to the integral constraint and the difficulty due to the non-convexity. We will first solve a problem with fractional association that assumes that a client can be associated with more than one AP at a time, then propose a rounding algorithm to recover the binary nature of the association variables.

5.1. Fractional association

Let us first relax the integral constraints then transform the relaxed problem into a convex programming problem where fractional association is allowed. In fractional association, a client may associate with more than one AP and receives traffic from different APs at different times. In Sections 6 and 7, we show how we solve the convex problem and approximate a binary association by rounding the fractional association obtained from the solution of the convex problem.

We relax the integral constraints by substituting the product $t_{ac} \times y_{ca}$ with a new variable f_{ac} which indicates the fractional association of client c to AP a: $f_{ac} = 0.5$, means AP a spends half of its active time on forwarding traffic to client c. Therefore, we have the following constraints to ensure that the total downlink traffic forwarded by an AP a is within the capacity limit

$$\sum_{c \in \mathcal{C}_a} f_{ac} \leqslant 1, \quad \forall a \in \mathcal{A}. \tag{4}$$

By doing so, we expand the feasible space of the problem and a client can associate with more than one AP. We transformed the discrete integral association problem into a continuous problem.

5.2. Convex programming relaxation

After relaxing the integer constraints, we still have a non-convex optimization problem with continuous variables. We then apply a logarithmic transformation to this problem and reformulate it as an equivalent convex problem. To begin with, we approximate the term $(1-\rho_n/T_n)^{T_H+L_a}$ in the constraint on collision probability s_{ac} with its continuous analogue $\exp(-\rho_n(T_H + L_a)/T_n)$ without any loss of accuracy [23], i.e.,

$$s_{ac} \leqslant \prod_{n \in \mathcal{I}_{ac}} \frac{1}{1+\rho_n} \times \exp\left(-\frac{\rho_n(T_H+L_a)}{T_n}\right), \quad \forall a \in \mathcal{A}_c, \ \forall c \in \mathcal{C}. \tag{5}$$

Then, we introduce a set of logarithmic changes of variables: $x'_{ac} = \log x_{ac}$, $s'_{ac} = \log s_{ac}$, $b'_a = \log b_a$, $\rho'_a = \log \rho_a$. By logarithmic changes of variables and logarithmic transformation on the first two constraints in problem (3) and the modified constraint (5), we can transform the non-convex constraints in problem (3) into convex constraints as

$$\begin{aligned}
&0 \geqslant x'_{ac} - s'_{ac} - b'_a - \log f_{ac} - \log\frac{L_a R_a}{T_a}, \quad \forall a \in \mathcal{A}_c, \ \forall c \in \mathcal{C},\\
&0 \geqslant b'_a - \rho'_a + \log(e^0 + e^{\rho'_a}), \quad \forall a \in \mathcal{A},\\
&0 \geqslant \sum_{n \in \mathcal{I}_{ac}} \left(\log(e^0 + e^{\rho'_n}) + e^{\rho'_n} \times \frac{T_H + L_a}{T_n}\right) + s'_{ac},\\
&\quad \forall a \in \mathcal{A}_c, \quad \forall c \in \mathcal{C}.
\end{aligned} \tag{6}$$

We remind the reader that $t_{ac} \times y_{ca}$ has been replaced by f_{ac}. Since both $-\log$, exp and log–sum–exp functions are convex [6], the transformed constraints are convex.

As for the objective function, if we apply the logarithmic changes of variables, it becomes

$$\sum_{c \in \mathcal{C}} \log\left(\sum_{a \in A_c} e^{x'_{ac}}\right), \tag{7}$$

which is a convex function. Even if we further apply a logarithmic transformation on this convex objective, the resulting function is still not concave in general. Maximizing a non-concave objective over a convex set is not a convex problem and a local maximum may not be a global maximum. To address this issues, we will instead maximize a concave objective function which is a lower bound on the original objective function (7). We then reduce the gap between the lower bound and the objective (7) iteratively.

We derive a lower bound by invoking the inequality for the weighted arithmetic mean and weighted geometric mean. Considering n positive variables $y_1, \ldots, y_n$ and weight $w_1, \ldots, w_n$, the weighted geometric mean is always a lower bound of the weighted arithmetic mean, i.e.,

$$\sum_{i=1}^{n} \frac{w_i y_i}{w} \geqslant \prod_{i=1}^{n} y_i^{w_i/w}, \tag{8}$$

where $w = w_1 + \cdots + w_n$. Substituting $x_i = \theta_i y_i$ where $\theta_i = w_i/w$, we can rewrite the above inequality as

$$\sum_{i=1}^{n} x_i \geqslant \prod_{i=1}^{n} \left(\frac{x_i}{\theta_i}\right)^{\theta_i}. \tag{9}$$

Using this relation, we approximate the lower bound of client c's downlink throughput $\sum_{a \in \mathcal{A}_c} x_{ac}$ as $\prod_{a \in \mathcal{A}_c} \left(\frac{x_{ac}}{\theta_{ac}}\right)^{\theta_{ac}}$ where $\sum_{a \in \mathcal{A}_c} \theta_{ac} = 1$ and $\theta_{ac} \geqslant 0$. θ_{ac} is a constant here, we will discuss how it is determined shortly. Applying the logarithmic change of variable, the concave lower bound is then

$$\log \sum_{a \in \mathcal{A}_c} x_{ac} \geqslant \log \prod_{a \in \mathcal{A}_c} \left(\frac{x_{ac}}{\theta_{ac}}\right)^{\theta_{ac}} = \sum_{a \in \mathcal{A}_c} \theta_{ac} x'_{ac} - \theta_{ac} \log \theta_{ac}. \tag{10}$$

As the term $-\sum_{a \in \mathcal{A}_c} \theta_{ac} \log \theta_{ac}$ is a constant, we will omit it in the problem formulation. We are now ready to rewrite problem (3) as a convex problem:

$$\begin{aligned}
\max \quad & \sum_{c \in \mathcal{C}} \sum_{a \in \mathcal{A}_c} \theta_{ac} x'_{ac}\\
\text{s.t.} \quad & 0 \geqslant x'_{ac} - s'_{ac} - b'_a - \log f_{ac} - \log \frac{L_a R_a}{T_a},\\
& \quad \forall a \in \mathcal{A}_c, \quad \forall c \in \mathcal{C},\\
& 0 \geqslant b'_a - \rho'_a + \log(1 + e^{\rho'_a}), \quad \forall a \in \mathcal{A},\\
& 0 \geqslant \sum_{n \in \mathcal{I}_{ac}} \left(\log(1 + e^{\rho'_n}) + e^{\rho'_n} \times \frac{T_H + L_a}{T_n}\right) + s'_{ac},\\
& \quad \forall a \in \mathcal{A}_c, \quad \forall c \in \mathcal{C}.\\
& \sum_{c \in \mathcal{C}_a} f_{ac} \leqslant 1, \quad \forall a \in \mathcal{A},\\
& 0 \leqslant f_{ac}, \quad \forall c \in \mathcal{C}_a, \forall a \in \mathcal{A},\\
& {\rho'_a}^{min} \leqslant \rho'_a \leqslant {\rho'_a}^{max}, \quad \forall a \in \mathcal{A},
\end{aligned} \tag{11}$$

where ${\rho'_a}^{min} = \log \rho_a^{min}$ and ${\rho'_a}^{max} = \log \rho_a^{max}$.

5.3. Determining the coefficient θ_{ac}

Inequality (10) becomes an equality if $\theta_{ac} = x_{ac}/\sum_{a\in\mathcal{A}_c} x_{ac}$. That is, at any fixed point $\mathbf{x}^0$, the value of the lower bound is exactly equal to the original objective if $\theta_{ac} = x^0_{ac}/\sum_{a\in\mathcal{A}_c} x^0_{ac}$. To gradually reduce the gap between the lower bound and the original objective, we adopt an iterative approach proposed in [12]. At iteration $k+1$, the coefficient matrix $\boldsymbol{\theta}(k+1)$ is evaluated based on the optimal throughput allocation of problem (11) obtained in iteration k, and we denote such an allocation as $\mathbf{x}^*(k)$ which is the matrix, $\{x^*_{ac}(k)\}_{\forall a\in\mathcal{A}_c,\ \forall c\in\mathcal{C}}$. $\theta_{ac}(k+1)$ is equal to

$$\theta_{ac}(k+1) = \frac{x^*_{ac}(k)}{\sum_{a\in\mathcal{A}_c} x^*_{ac}(k)}. \tag{12}$$

We then solve problem (11) again with this new $\boldsymbol{\theta}(k+1)$. We iterate this procedure until $\boldsymbol{\theta}$ converges. The procedure works as follow:

(1) Set $k = 0$ and initialize $\boldsymbol{\theta}(k)$ with any feasible setting
(2) Increment k by 1
(3) Apply concave approximation (10) to the objective and transform the optimization problem into a convex problem
(4) Solve the resulting problem and obtain optimal $\mathbf{x}^*(k)$
(5) Update $\theta_{ac}(k+1)$ by (12)
(6) Jump back to step 2 until $\boldsymbol{\theta}$ converges

As shown in [12] through extensive numerical experiments, a global optimal solution is always achieved in various power control optimization problems. We also observed that this procedure always converges when we apply it to our problem in different random scenarios in Section 8.

6. Fractional association algorithm

To solve problem (11), we consider its Lagrange relaxation and dual problem. It can be verified easily that Slater's condition is satisfied and the strong duality holds [6]. There is no duality gap between the primal problem and the dual problem.

We relax the first sets of constraints in problem (11) with a set of Lagrangian multipliers $\boldsymbol{\lambda} = \{\lambda_{ac}\}_{\forall a\in\mathcal{A}_c, \forall c\in\mathcal{C}}$ The Lagrangian is then

$$\sum_{c\in\mathcal{C}}\sum_{a\in\mathcal{A}_c}\left(\theta_{ac}x'_{ac} + \lambda_{ac}\left(s'_{ac} + b'_a + \log f_{ac} + \log\frac{L_aR_a}{T_a} - x'_{ac}\right)\right). \tag{13}$$

If we differentiate the Lagrangian with respect to x'_{ac}, the variable vanishes and we cannot recover it from the dual solution. Several optimization algorithms like the proximal optimization algorithm [6] can be used to recover the primal solution. However, such algorithms require synchronized global updates of some variables. It creates additional complexity on the solution algorithm. To keep our algorithm simple, we tackle this problem by adding a set of concave quadratic terms to the objective function, as

$$\max \sum_{c\in\mathcal{C}}\sum_{a\in\mathcal{A}_c}\theta_{ac}x'_{ac} - \frac{1}{2\beta}\sum_{c\in\mathcal{C}}\sum_{a\in\mathcal{A}_c}\left(x'_{ac}\right)^2. \tag{14}$$

By choosing constant β sufficiently large, the primal variables can be recovered very accurately. In Section 8, we find that the additional concave terms are less than 0.01% of the optimal objective value by setting β = 10,000. Such small perturbation indicates that this approach is simple yet very accurate.

6.1. Dual decomposition

With the new objective (14), the corresponding dual function $D(\boldsymbol{\lambda})$ is

$$\begin{aligned}
D(\boldsymbol{\lambda}) = &\max_{\mathbf{x},\mathbf{s},\mathbf{b},\mathbf{f}} L(\boldsymbol{\lambda},\mathbf{x},\mathbf{s},\mathbf{b},\mathbf{f})\\
\text{s.t.}\quad & 0 \geqslant b'_a - \rho'_a + \log\left(1+e^{\rho'_a}\right), \quad \forall a\in\mathcal{A},\\
& 0 \geqslant \sum_{n\in\mathcal{I}_{ac}}\left(\log\left(1+e^{\rho'_n}\right) + e^{\rho'_n}\times\frac{T_H+L_a}{T_n}\right) + s'_{ac},\\
&\quad \forall a\in\mathcal{A}_c, \quad \forall c\in\mathcal{C}.\\
& \sum_{c\in\mathcal{C}_a} f_{ac} \leqslant 1, \quad \forall a\in\mathcal{A},\\
& 0 \leqslant f_{ac}, \quad \forall c\in\mathcal{C}_a, \forall a\in\mathcal{A},\\
& {\rho'_a}^{min} \leqslant \rho'_a \leqslant {\rho'_a}^{max}, \quad \forall a\in\mathcal{A},
\end{aligned} \tag{15}$$

where the Lagrangian $L(\boldsymbol{\lambda},\mathbf{x},\mathbf{s},\mathbf{b},\mathbf{f})$ is defined as

$$\begin{aligned}
L(\boldsymbol{\lambda},\mathbf{x},\mathbf{s},\mathbf{b},\mathbf{f}) = &\sum_{c\in\mathcal{C}}\sum_{a\in\mathcal{A}_c}\theta_{ac}x'_{ac} - \frac{1}{2\beta}\sum_{c\in\mathcal{C}}\sum_{a\in\mathcal{A}_c}\left(x'_{ac}\right)^2\\
&+\sum_{c\in\mathcal{C}}\sum_{a\in\mathcal{A}_c}\lambda_{ac}\\
&\times\left(s'_{ac} + b'_a + \log f_{ac} + \log\frac{L_aR_a}{T_a} - x'_{ac}\right),
\end{aligned} \tag{16}$$

and the dual problem is defined as

$$\begin{aligned}
\min\quad & D(\boldsymbol{\lambda}),\\
\text{s.t.}\quad & \boldsymbol{\lambda} \succeq 0.
\end{aligned} \tag{17}$$

We use a gradient descent method to obtain the optimal solution $\boldsymbol{\lambda}^*$ of problem (17) [6]. In iteration t, $\lambda_{ac}(t)$ are updated as follow

$$\lambda_{ac}(t+1) = \max(\lambda_{ac}(t) - \gamma g(\lambda_{ac}(t)), 0), \tag{18}$$

where γ is a positive step size and $g(\lambda_{ac}(t))$ is the gradient with respect to $\lambda_{ac}(t)$.

Given the multipliers $\boldsymbol{\lambda}(t)$, we denote the corresponding maximizers as $s'^{\lambda(t)}_{ac}$, $b'^{\lambda(t)}_a$, $f^{\lambda(t)}_{ac}$, $x'^{\lambda(t)}_{ac}$ which maximizes Lagrangian $L(\boldsymbol{\lambda},\mathbf{x},\mathbf{s},\mathbf{b},\mathbf{f})$. The gradient $g(\lambda_{ac}(t))$ can be obtained by

$$g(\lambda_{ac}(t)) = s'^{\lambda(t)}_{ac} + b'^{\lambda(t)}_a + \log f^{\lambda(t)}_{ac} + \log\frac{L_aR_a}{T_a} - x'^{\lambda(t)}_{ac}. \tag{19}$$

In the later part of this section, we show that the maximizers $s'^{\lambda(t)}_{ac}$, $b'^{\lambda(t)}_a$, $f^{\lambda(t)}_{ac}$, $x'^{\lambda(t)}_{ac}$ are always unique as the objective function is strictly concave and the feasible set is convex [6]. The dual function is differentiable everywhere and the gradient g always exists. The gradient projection method can be used to obtain the optimal dual solution.

Step sizes γ can be a constant or can be determined by various convergence rules [6] beyond the scope of this paper.

To find the maximizers of the dual function $D(\lambda)$, we decompose problem (15) into three subproblems and formulate them as, a rate control problem in (20),

$$\max \sum_{a\in\mathcal{A}}\sum_{c\in\mathcal{C}_a}\left((\theta_{ac}-\lambda_{ac})x'_{ac}-\frac{1}{2\beta}(x'_{ac})^2\right); \tag{20}$$

a fractional association problem in (21),

$$\begin{aligned} \max\quad & \sum_{a\in\mathcal{A}}\sum_{c\in\mathcal{C}_a}(\lambda_{ac}\log f_{ac}),\\ \text{s.t.}\quad & \sum_{c\in\mathcal{C}_a} f_{ac}\leqslant 1,\quad \forall a\in\mathcal{A},\\ & 0\leqslant f_{ac},\quad \forall c\in\mathcal{C}_a,\ \forall a\in\mathcal{A}; \end{aligned} \tag{21}$$

$$\begin{aligned} \max\quad & \sum_{a\in\mathcal{A}}\sum_{c\in\mathcal{C}_a}\lambda_{ac}\left\{\rho'_a-\log\left(1+e^{\rho'_a}\right)-\sum_{n\in\mathcal{I}_{ac}}\left(\log\left(1+e^{\rho'_n}\right)+e^{\rho'_n}\times\frac{T_H+L_a}{T_n}\right)\right\},\\ \text{s.t.}\quad & {\rho'_a}^{min}\leqslant\rho'_a\leqslant{\rho'_a}^{max},\quad \forall a\in\mathcal{A}, \end{aligned} \tag{27}$$

and a contention resolution problem in (22),

$$\begin{aligned} \max\quad & \sum_{a\in\mathcal{A}}\sum_{c\in\mathcal{C}_a}\lambda_{ac}(s'_{ac}+b'_a),\\ \text{s.t.}\quad & 0\geqslant b'_a-\rho'_a+\log(1+e^{\rho'_a}),\quad \forall a\in\mathcal{A},\\ & 0\geqslant \sum_{n\in\mathcal{I}_{ac}}\left(\log(1+e^{\rho'_n})+e^{\rho'_n}\times\frac{T_H+L_a}{T_n}\right)+s'_{ac},\\ & \quad \forall a\in\mathcal{A}_c,\quad \forall c\in\mathcal{C}. \end{aligned} \tag{22}$$

The maximizers of the rate control problem (20), can be obtained by maximizing a set of quadratic functions. The maximizer x'^{λ}_{ac} is equal to

$$x'^{\lambda}_{ac}=\beta(\theta_{ac}-\lambda_{ac}). \tag{23}$$

We solve the second problem (21), which is a fractional association problem, by considering its dual problem. We relax the first set of constraints with Lagrange multiplier $\boldsymbol{v}_a$. The problem is separable and each AP solves its own maximization problem. The local dual problem D_a of AP a is defined as $\min D_a(\boldsymbol{v})$, where

$$D_a(\boldsymbol{v})=\sum_{a\in\mathcal{A}}\max\sum_{c\in\mathcal{C}_a}(\lambda_{ac}\log f_{ac}-v_a f_{ac}). \tag{24}$$

The price λ_{ac} from the master dual problem serves as a weight in the maximization problem of the dual function. We again invoke the gradient projection method to solve the problem. We emphasize that λ_{ac} is a constant in the dual function D_a.

At iteration n, with the price $v_a(n)$, the maximizer $f_{ac}^{v_a(n)}$ is equal to

$$f_{ac}^{v_a(n)}=[\lambda_{ac}/v_a(n)]_0^1, \tag{25}$$

where $[\cdot]_0^1$ is the projection between 0 and 1. At iteration $n+1$, the price $v_a(n+1)$ is updated by

$$v_a(n+1)=\max\left(v_a(n)-\alpha\times\left(1-\sum_{c\in\mathcal{C}_a}f_{ac}^{v_a}\right),0\right), \tag{26}$$

and α is a positive constant. Such iterative updates continue until the gradient projection method converges and we reach the optimal price v_a^*. At the optimum point, we have $f_{ac}^{\lambda}=f_{ac}^{v_a^*}=\lambda_{ac}/v_a^*$.

To solve the third problem (22), which is a contention resolution problem, we rewrite the problem by substituting the constraints on b'_a and s'_{ac} into the objective:

We can further reformulate the above problem as a separable problem. We introduce a new notation, $\mathcal{H}_a$, which represents a set of clients that may suffer from hidden node collision due to AP a's transmission. $\mathcal{H}_a$ is defined as

$$\mathcal{H}_a=\bigcup k: a\in\mathcal{I}_{nk},\quad \forall k\in\mathcal{C}_n,\ \forall n\in\mathcal{A}. \tag{28}$$

AP a only needs to solve a local maximization problem:

$$\begin{aligned} \max\quad & \sum_{c\in\mathcal{C}_a}\left(\lambda_{ac}\left(\rho'_a-\log\left(1+e^{\rho'_a}\right)\right)\right)\\ & -\sum_{k\in\mathcal{H}_a}\sum_{n\in\mathcal{A}_k\setminus\{a\}}\lambda_{nk}\left(\log\left(1+e^{\rho'_a}\right)+e^{\rho'_a}\times\frac{T_H+L_n}{T_a}\right),\\ \text{s.t.}\quad & {\rho'_a}^{min}\leqslant\rho'_a\leqslant{\rho'_a}^{max}. \end{aligned} \tag{29}$$

By defining $C_1=\sum_{c\in\mathcal{C}_a}\lambda_{ac}$, $C_2=\sum_{k\in\mathcal{H}_a}\sum_{n\in\mathcal{A}_k\setminus\{a\}}\lambda_{nk}$ and $C_3=\sum_{k\in\mathcal{H}_a}\sum_{n\in\mathcal{A}_k\setminus\{a\}}\lambda_{nk}\frac{T_H+L_n}{T_a}$, the local contention resolution problem is equivalent to

$$\begin{aligned} \max\quad & CR_a(\rho'_a)\\ \text{s.t.}\quad & {\rho'_a}^{min}\leqslant\rho'_a\leqslant{\rho'_a}^{max}, \end{aligned} \tag{30}$$

where $CR_a(\rho'_a)$ is defined as

$$CR_a(\rho'_a)=C_1\rho'_a-C_3e^{\rho'_a}-(C_1+C_2)\log\left(1+e^{\rho'_a}\right). \tag{31}$$

We can obtain the maximizer ρ'^{λ}_a by solving the equation $\nabla CR_a(\rho'_a)=0$ which is

$$C_1-C_3e^{\rho'_a}-(C_1+C_2)\frac{e^{\rho'_a}}{(1+e^{\rho'_a})}=0. \tag{32}$$

If we substitute $\rho_a = e^{\rho'_a}$ back to the above equation, we have a quadratic equation. Distinct real roots of this equation always exist and are given by

$$\rho_a^{\lambda} = \frac{-(C_2+C_3) \pm \sqrt{(C_2+C_3)^2 + 4C_1C_3}}{2C_3}. \tag{33}$$

We can see that one of them is always positive while the other is always negative. We discard the negative one and we obtain $\rho_a'^{\lambda}$ by

$$\rho_a'^{\lambda} = \left[\log\left(\frac{-(C_2+C_3) + \sqrt{(C_2+C_3)^2 + 4C_1C_3}}{2C_3}\right)\right]_{\rho_a'^{\,min}}^{\rho_a'^{\,max}}, \tag{34}$$

where $[\cdot]_{\rho_a'^{\,min}}^{\rho_a'^{\,max}}$ is the projection which ensures that $\rho_a'^{\lambda}$ is within the feasible range. If $\rho_a'^{\lambda} > \rho_a'^{\,max}$, then we set $\rho_a'^{\lambda} = \rho_a'^{\,max}$. If $\rho_a'^{\lambda} < \rho_a'^{\,min}$, then we set $\rho_a'^{\lambda} = \rho_a'^{\,min}$. After obtaining $\rho_a'^{\lambda}$, we can evaluate $s_{ac}'^{\lambda}$ and $b_a'^{\lambda}$ by

$$\begin{aligned} b_a'^{\lambda} &= \rho_a'^{\lambda} - \log(1 + e^{\rho_a'^{\lambda}}), \\ s_{ac}'^{\lambda} &= -\sum_{n \in \mathcal{I}_{ac}} \left(\log\left(1 + e^{\rho_n'^{\lambda}}\right) + e^{\rho_n'^{\lambda}} \times \frac{T_H + L_a}{T_n}\right). \end{aligned} \tag{35}$$

6.2. Distributed algorithm

We propose a distributed algorithm in which the APs carry out the computations separately and cooperate via information exchanges on the wired distribution network. Clients only assist in information gathering during the initialization, therefore no firmware modification is needed. We assume the algorithm is executed periodically for an offline optimization. The clients are associated with the APs before the execution of the algorithm, however the result of the algorithm may lead to a reassociation that optimizes the performance of the network.

At initialization, the clients detect the set of accessible APs from the beacon frames received. We denote such a set for client c, B_c. Since we assume APs are not synchronized by the DCF, we can estimate $I_{ac} = B_c \backslash \{a\}$ for all $a \in B_c$. Such an estimation is an online approximation and it excludes the interference that clients cannot decode. With proper firmware modification, more accurate interference estimation like micro-probing [2] can be used to detect sources of interference that cannot be decoded. After setting I_{ac}, we initialize $\mathcal{A}_c = B_c$. We can limit the choices for client c by setting $\mathcal{A}_c \subseteq B_c$. We apply such a limit when we round the fractional association (more detail will be discussed in Section 7). Client c notifies APs in $\mathcal{A}_c$ about its existence. All other clients perform a similar notification. AP a can construct its possible clients set $\mathcal{C}_a$ for all $a \in \mathcal{A}$. The contents of $\mathcal{A}_c$ and $\mathcal{I}_{ac}$ are also transmitted to AP a for information exchange during the execution of the algorithm. In addition, client c sends the information $\mathcal{A}_c$ to AP n for all $n \in \mathcal{I}_{ac}$. Based on such information, AP n constructs the set $\mathcal{H}_n$ and obtains the list of $\mathcal{A}_k$ for all $k \in \mathcal{H}_n$.

The detail of the algorithm is shown in Algorithm 1. The algorithm, is iterative, and comprises two tiers. The outer tier is the concave objective approximation and the inner tier is the gradient projection method.

Algorithm 1. Fractional association algorithm $FA(\{\mathcal{A}_\rfloor\}, \{\mathcal{C}_\dashv\}, \{\mathcal{I}_{\dashv\rfloor}\}, \mathcal{A}, \mathcal{C})$

1: Set $k = 1$
2: Initialize λ, $\mathbf{x}'$, β, $\mathbf{f}$ and ρ' at each AP
3: **repeat**
4: **for** each AP $a \in \mathcal{A}$ **do**
5: Apply concave approximation (10) to the objective
6: **end for**
7: Set $t = 1$
8: **repeat**
9: **for** each AP $a \in \mathcal{A}$ **do**
10: Evaluate $x_{ac}'^{\lambda(t)}$ by (23)
11: Obtain $f_{ac}^{\lambda(t)}$ by the gradient projection method (25) and (26)
12: Obtain $\rho_a'^{\lambda(t)}$ by (34)
13: Disseminate $\rho_a'^{\lambda(t)}$ to AP $n,\ \forall n \in \mathcal{A}_k,\ \forall k \in \mathcal{H}_a$
14: **end for**
15: **for** each AP $a \in \mathcal{A}$ **do**
16: Obtain $b_a'^{\lambda(t)}$ by (35)
17: **for** each client $c \in \mathcal{C}_a$ **do**
18: Obtain $s_{ac}'^{\lambda(t)}$ by (35)
19: Obtain price $\lambda_{ac}(t+1)$ by (18)
20: Disseminate $\lambda_{ac}(t+1)$ to AP $n,\ \forall n \in \mathcal{I}_{ac}$
21: **end for**
22: **end for**
23: $t = t + 1$
24: **until** Convergence of $\mathbf{x}'$, and we denote the optimal throughput allocation as $\mathbf{x}'^{*}$
25: $x_{ac}^{*}(k) = \log x_{ac}'^{*}$
26: $\theta_{ac}(k+1) = \frac{x_{ac}^{*}(k)}{\sum_{a \in \mathcal{A}_c} x_{ac}^{*}(k)},\ \forall a \in \mathcal{A}_c,\ \forall c \in \mathcal{C}$
27: $k = k + 1$
28: **until** Convergence of the coefficient θ

In the inner tier (from lines 6.2 to 6.2 in Algorithm 1) we solve the throughput optimization via a gradient projection method. To achieve a distributed computation, the APs exchange their own respective price information λ_{ac} and the variables ρ'_a and f_{ac} with other APs on the wired distribution network. In the outer tier (from lines 6.2 to 6.2 in Algorithm 1) we carry out a series of concave approximations on the objective function. The algorithm converges when no further change in the coefficient θ can be made.

Our algorithm does not require any global information exchange or centralized scheduling. APs cooperate with each other via the wired infrastructure network. It is very reasonable as all these APs are part of the same extended network and are outside the radio communication range of each other. All arithmetic operations are very simple and are well supported by common low power embedded platforms like commercial WiFi routers.

7. Rounding algorithm

We now discuss how to round off the fractional association to an integral association. Classical iterative approaches like branch-and-bound, generalized benders decomposition and outer approximation [24] are not suitable for a distributed solution as a centralized node is needed to coordinate the calculation between iterations. Also, unlike previous studies with no inter-AP interference, our problem cannot be transformed into the classic generalized assignment problem [34] in which jobs are assigned to independent machines such that the total reward is maximized. In our problem, the service time of APs are correlated due to the hidden node collisions which leads to extra complexity. To the best of our knowledge, no approximation rounding procedure has been proposed to solve such a problem.

To tackle the rounding problem, we performed extensive numerical experiments and observed that the sum of the total downlink throughput of a client is dominated by only one of its associations nearly in all the cases. Also, we can obtain the optimal integral association by simply considering the association that contributes the largest portion of the downlink throughput. Resolving the throughput optimization problem again with such a fixed association setting, we can get the optimal average backoff time and transmission rate. Such an approach is distributed and can be implemented without centralized server. A formal description of the algorithm is summarized in Algorithm 2.

Algorithm 2. Rounding algorithm

1: Obtain $\mathbf{x}^*$ by $FA(\{\mathcal{A}_\rfloor\},\{\mathcal{C}_\dashv\},\{\mathcal{I}_{\dashv\rfloor}\},\mathcal{A},\mathcal{C})$
2: **for** each client $c \in \mathcal{C}$ **do**
3: Fix the association $\mathcal{A}_c^* = \mathrm{argmax}_a x_{ac}^*$
4:
end for
5: **for** each client $a \in \mathcal{A}$ **do**
6: Setting $\mathcal{C}_a^*$ based on $\mathcal{A}_c^*$, $\forall c \in \mathcal{C}$
7: **end for**
8: Obtain the optimal throughput and windows setting by $FA\left(\left\{\mathcal{A}_\rfloor^*\right\},\{\mathcal{C}_\dashv^*\},\{\mathcal{I}_{\dashv\rfloor}\},\mathcal{A},\mathcal{C}\right)$

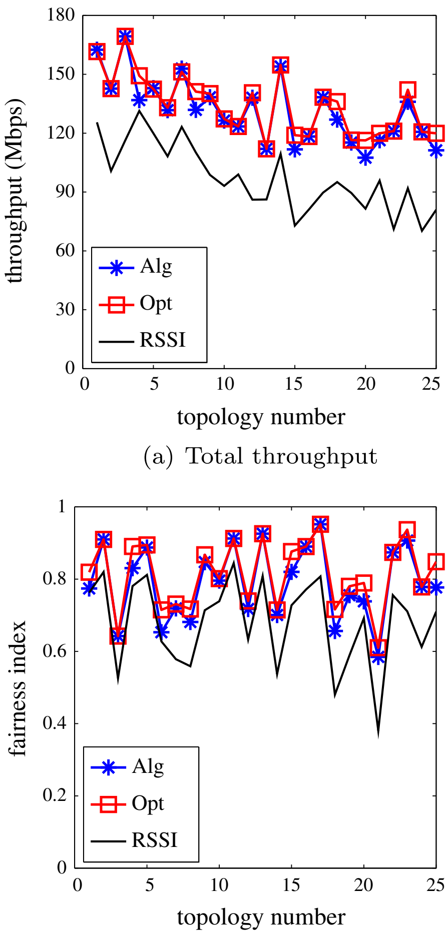
In the second optimization problem of Algorithm 2 at line 7, we fixed the association for each client. The objective function is $\max \sum_{c\in\mathcal{C}}\sum_{\mathcal{A}_c^*} x_{ac}$ and no concave approximation on the objective is needed.

7.1. Applying the optimal settings

Solving the optimization problem yields the optimal average backoff time for each AP, the transmission rate for each downlink flow and the optimal association. As suggested in [4,25], the optimization can be executed periodically in an offline manner. Clients can be reassociated when their current transmission session is over in order not to disrupt the ongoing transmission. To achieve the optimal backoff time, AP a can set its CW_{min} and CW_{max} as $2 \times bo_a^*$ since the average backoff time is a uniform random variable in the IEEE 802.11 DCF. Also, AP a can schedule the transmission among different downlink flows by weighted fair queueing where the weight of each flow is equal to f_{ac} and the resulting transmission rate should be x_{ac}^* for all client $c \in \mathcal{C}_a$. To achieve a reliable transmission,

Table 1
Number of association combinations for different topologies.

T	# of Comb.	T	# of Comb.	T	# of Comb.
1	512	10	55,296	18	1,048,576
2	3072	11	147,456	19	1,179,648
3	16,384	12	393,216	20	1,327,104
3	24,576	13	393,216	21	2,359,296
5	24,576	14	393,216	22	2,654,208
6	24,576	15	786,432	23	3,538,944
7	36,864	16	786,432	24	4,478,976
8	36,864	17	995,328	25	4,718,592
9	36,864				



(a) Total throughput
(b) Fairness index

Fig. 3. Performance evaluation via numerical experiments.

the maximum transmission limit per data frame for AP a, m_a, should satisfy $(1 - s_{ac})^{m_a} \approx 0,\ \forall c \in \mathcal{C}_a$.

8. Performance evaluation

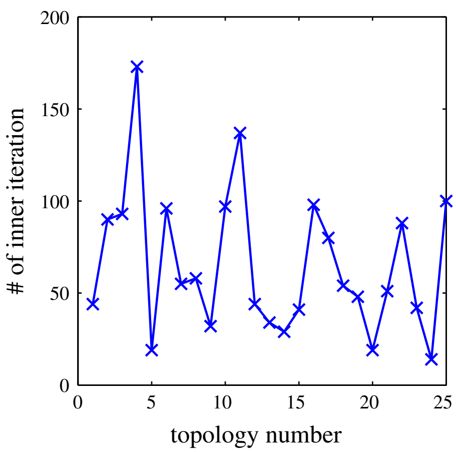
In this section, we evaluate the performance of the distributed algorithm via numerical experiments and simulation. We consider a 3×3 grid network with 9 APs. Its structure is similar to the 2×2 example in Fig. 1. There are 30 clients distributed uniformly within the coverage area. A client can at most detect the beacon message from four different APs. We generated 25 random topologies and the respective number of possible association combinations for these topologies are listed in Table 1 (sorted from the smallest to the largest).

8.1. Numerical experiments

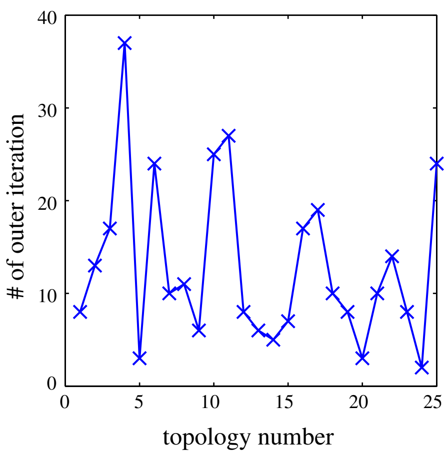
We have implemented the algorithm in MATLAB and conducted several numerical experiments to study its performance. We examine the accuracy of our algorithm by comparing its performance to the optimal solution. The optimal solution for each topology is obtained by brute-force enumeration: when examining a combination, we fix the association for each client and solve the corresponding optimization problem. We studied 2.55×10^7 combinations in total. We emphasize that, when the association is fixed, the concave lower bound approximation (10) is not needed when solving the optimization problem as the objective function is already a linear function in this case.

Table 2
Simulation parameters.

Slot time σ	9 µs
SIFS time	10 µs
Preamble duration	20 µs
Signal extension time	6 µs
Default CW_{min}, CW_{max}	15, 1023
Data rate	54 Mbps
MTU	1500 bytes
Basic rate for ACK and MAC header	6 Mbps
Maximum queue length	100 packets

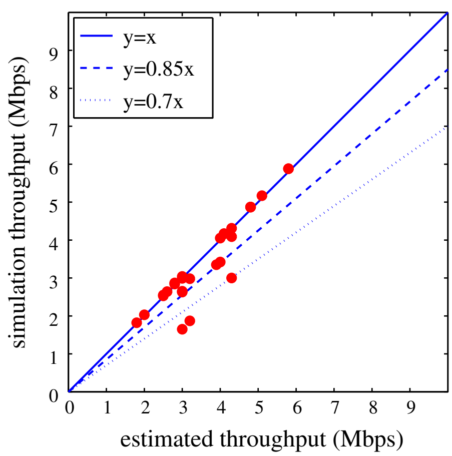


(a) Convergence: inner tier

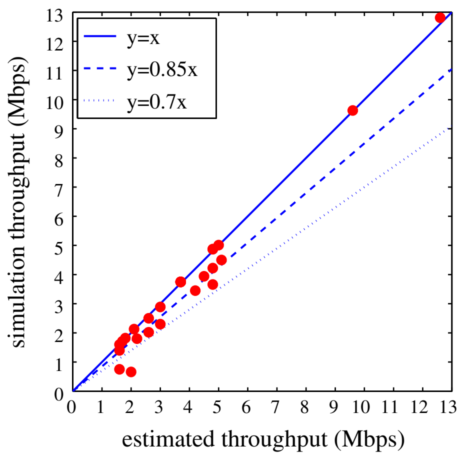


(b) Convergence: outer tier

Fig. 4. Convergence of the distributed algorithm.



(a) Topology 13



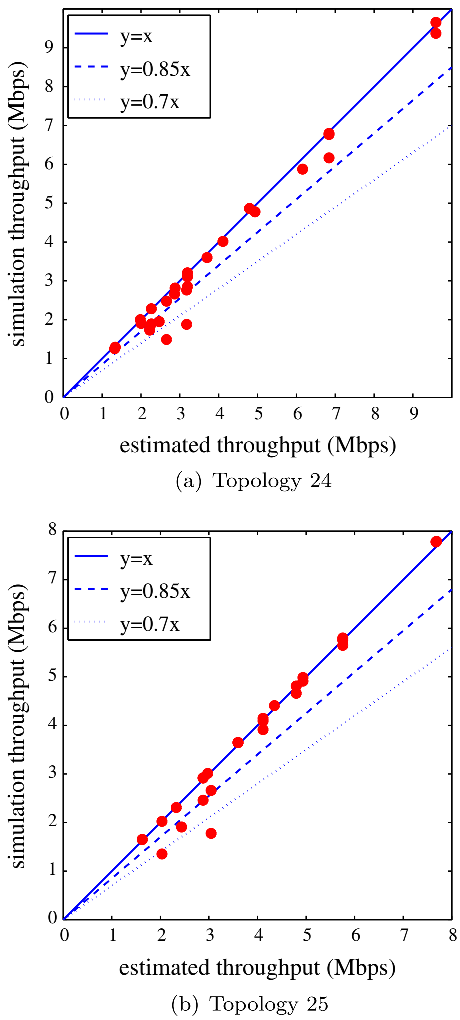
(b) Topology 19

Fig. 5. Performance evaluation through ns-2 simulation.

We also study the effectiveness of our algorithm by comparing its performance to that of the strongest-RSSI-based association. The signal strength is determined by the path loss models. To the best of our knowledge, no similar joint cross-layer optimization scheme has been proposed for this problem either. To obtain the performance of the signal strength based association, we fixed the association based on the strongest RSSI and apply our optimization framework.

We show in Fig. 3a the total throughput achieved by the three methods and in Fig. 3b the associated fairness index for the different topologies. We can see that our algorithm matches very well the optimal solution for both total throughput of the network and the fairness index. This indicates that the throughput assignment to each client by our algorithm is very close to the optimal one. The calculated average error is 2.3% and 3.07% for the total throughput and fairness index respectively. Also, we can observe that our algorithm outperforms the RSSI-based association in all topologies. Our algorithm gains on average 34Mbps more than the RSSI-based association in terms of total throughput. The distribution is also fairer as shown in Fig. 3b. Setting $\beta = 10{,}000$ in our experiments, we find that the value of the extra concave terms in the objective is less than 0.01% of the total objective. In addition, no active AP's average backoff time is larger than the $CW_{min}/2$ slots.

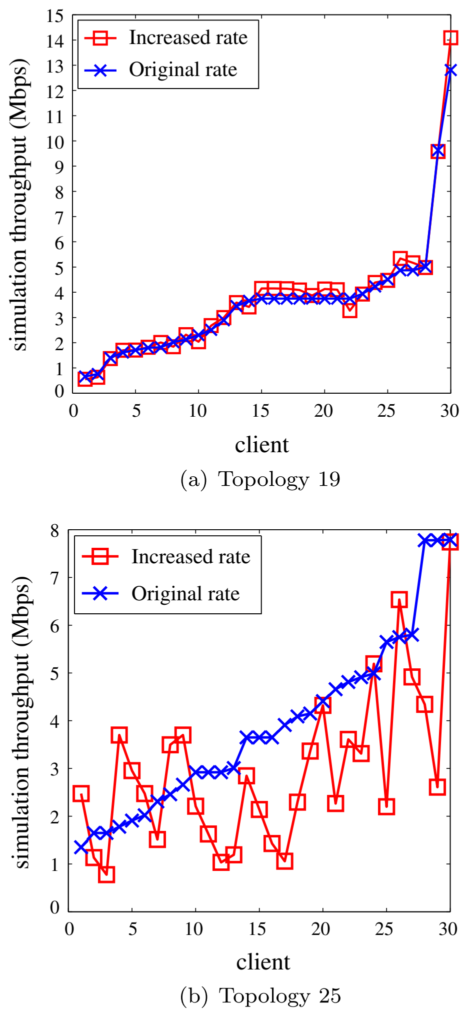
In our distributed algorithm, we invoke the fractional association algorithm *FA* twice to solve the problem. We define the convergence of the inner tier as: the consecutive changes of x'_{ac} is less than 1%, $\forall a \in \mathcal{A}_c,\ c \in \mathcal{C}$. A similar convergence rule is applied to θ_{ac} for the outer tier. In some topologies, when the algorithm is close to the optimum, there are some small oscillations for some x'_{ac} and they only account for less than 1% of a client's throughput. We exclude them in the convergence test. We report the total number of executions of the inner tier in Fig. 4a and outer tier in Fig. 4b. For example, around fifty inner tier iterations have been executed for topology 1. Whereas eight outer iterations are executed for the same case. We can observe that the algorithm converges very quickly and most of the topologies only require less than a hundred inner tier iter-



(a) Topology 24

(b) Topology 25

Fig. 6. Performance evaluation through ns-2 simulation.



(a) Topology 19

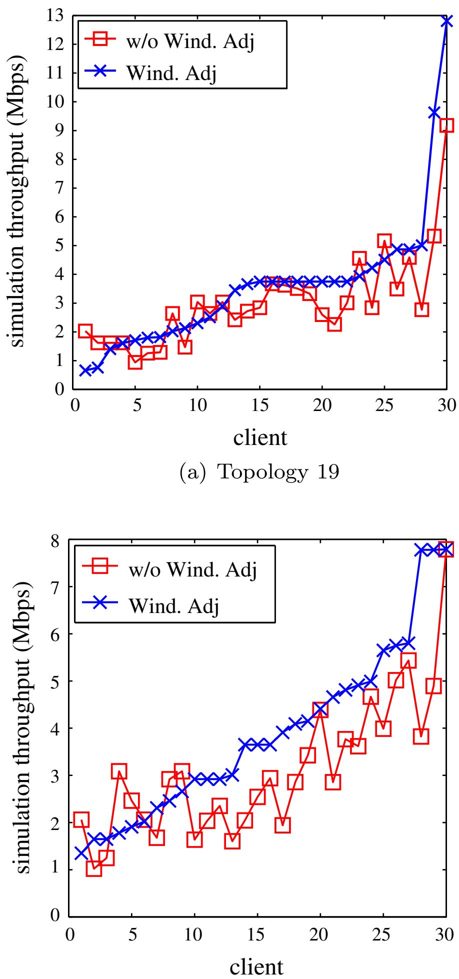
(b) Topology 25

Fig. 7. Under-prediction test.

ations and twenty outer iter iterations. As the arithmetic operations in our algorithm are very simple, we foresee that the algorithm converges within few seconds in real-world deployment.

8.2. Performance via ns-2 simulation

We study the performance of the algorithm in IEEE 802.11g networks by using ns-2 [30] simulation. We present the results for topologies 13, 19, 24 and 25 where topologies 24 and 25 are the most severely under the influence of hidden node collisions as they have the highest number of association combinations, and topologies 13 and 19 are under moderate influence. The IEEE 802.11g standard parameters are the same as those in Table 2. We apply the optimal settings from the algorithm's output to the APs. Fig. 5a and b compares the estimated throughput of a client by the algorithm and the actual allocation in simulation for topologies 13 and 19 respectively; Fig. 6a and b compares the estimated throughput of a client by the algorithm and the actual allocation in the simulation



(a) Topology 19

(b) Topology 25

Fig. 8. Window adjustment test.

for topologies 24 and 25 respectively. The three lines in the figure are $y = x$, $y = 0.85x$ and $y = 0.7x$. The predictions are very accurate and the average estimation error is less than 10% in all topologies under consideration.

We also want to determine if our algorithm under-predicts the achievable throughput or not. It is hard to detect this as the solution space spans three dimensions: association, rate control and contention resolution. To have a basic idea, we use a simple approach by only setting the source transmission rate 10% more than the optimal solution for each flow and repeat the simulation. We take topologies 19 and 25 as examples and the resulting throughput is reported in Fig. 7a and b respectively where the simulation results of the original source rates and increased source rates are reported. The simulation results are sorted according to the throughput obtained from the original source rate scenario. After increasing the source transmission rate, in topology 19, we observe that most of the clients' throughput remain the same; in topology 25, some of the clients have obtained even lesser throughput than before. This implies that the algorithm has already achieved the optimal rate allocation and there is no room to improve the throughput by individually increasing the sending rate.

Finally, we demonstrate the importance of the contention resolution control by only applying the rate control to the APs without window adjustment and repeat the same simulation. Taking topologies 19 and 25 as examples, Fig. 8a and b compares between the throughput with window adjustment and that obtained without window adjustment. When the window adjustment is absent, a throughput reduction, 19% for topology 19 and 25% for topology 25, occurs, indicating the importance of the window adjustment.

9. Conclusion

In this paper, we studied the problem of optimizing the throughput and fairness in multi-cell WLANs with inter-AP interference. We identified the need for a cross-layer approach and presented a practical joint association control, rate control and contention resolution algorithm. We formulated a cross-layer throughput optimization problem whose objective is to allocate downlink throughput according to the proportional fairness principle. As the problem turned out to be a non-convex mixed integer programming problem, known to be NP-hard, we relaxed and transformed it into a continuous convex problem and proposed a distributed algorithm to solve it. Such a distributed algorithm can be executed periodically to optimize network performance. The effectiveness, convergence speed and validation of the algorithm were demonstrated via ns-2 simulation as well as numerical experiments. A very interesting yet not trivial possible extension of this study would be to incorporate channel assignment as part of the optimization problem.

Acknowledgement

This work is supported in part under Grant GRF 610409.

References

[1] 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, 2007.

[2] N. Ahmed, U. Ismail, K. Papagiannaki, Online estimation of RF interference, in: ACM CoNEXT, 2008.

[3] M. Alicherry, B. Randeep, L. Li, Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks, in ACM MobiCom, 2005.

[4] Y. Bejerano, S.-J. Han, L. Li, Fairness and load balancing in wireless lans using association control, IEEE/ACM Transactions on Networking 15 (3) (2007) 560–573.

[5] Y. Bejerano, S.-J. Han, M. Smith, A novel frequency planning algorithm for mitigating unfairness in wireless lans, Computer Networks 54 (15) (2010) 2575–2590.

[6] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific, 1999.

[7] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE Journal on Selected Areas in Communications 18 (3) (2000) 535–547.

[8] S. Bosio, A. Capone, M. Cesana, Radio planning of wireless local area networks, IEEE/ACM Transactions on Networking 15 (6) (2007) 1414–1427.

[9] F. Cali', M. Conti, E. Gregori, IEEE 802.11 wireless LAN: capacity analysis and protocol enhancement, in: IEEE INFOCOM, 1998.

[10] J. Camp, E. Knightly, The IEEE 802.11s extended service set mesh networking standard, IEEE Communications Magazine 46 (8) (2008) 120–126.

[11] M. Chen, S.C. Liew, Z. Shao, C. Kai, Markov approximation for combinatorial network optimization, in: IEEE INFOCOM, 2010.

[12] M. Chiang, C.W. Tan, D.P. Palomar, D. O'Nell, D. Julian, Power control by geometric programming, IEEE Transactions on Wireless Communications 6 (7) (2007) 2640–2651.

[13] A. Dhananjay, H. Zhang, J. Li, L. Subramanian, Practical, distributed channel assignment and routing in dual-radio mesh networks, in: ACM SIGCOMM, 2009.

[14] K.-L. Hung, B. Bensaou, TCP performance optimization in multi-cell WLANs, in: ACM MSWiM, 2009.

[15] A.P. Jardosh, K. Mittal, K.N. Ramach, E.M. Belding, K.C. Almeroth, ABSTRACT IQU: practical queue-based user association management for WLANs, in: ACM MobiCom, 2006.

[16] L. Jiang, J. Walrand, A distributed CSMA algorithm for throughput and utility maximization in wireless networks, in: Allerton Conference on Communication, Control, and Computing, 2008.

[17] F.P. Kelly, A.K. Maulloo, D.K.H. Tan, Rate control for communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society 49 (3) (1998) 237–252.

[18] D. Kotz, K. Essien, Analysis of a campus-wide wireless network, in: ACM MobiCom, 2002.

[19] I. Koutsopoulos, L. Tassiulas, Joint optimal access point selection and channel assignment in wireless networks, IEEE/ACM Transactions on Networking 15 (3) (2007) 521–532.

[20] A. Kumar, E. Altman, D. Miorandi, M. Goyal, New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs, IEEE/ACM Transactions on Networking 15 (3) (2007) 588–601.

[21] H. Lee, S. Kim, O. Lee, S. Choi, S.-J. Lee, Available bandwidth-based association in IEEE 802.11 wireless LANs, in: ACM MSWiM, 2008.

[22] J. Lee, J. Lee, Y. Yi, S. Chong, A. Proutiere, M. Chiang, Implementing utility-optimal CSMA, in: Allerton Conference on Communication, Control, and Computing, 2009.

[23] A. Leon-Garcia, Probability and random processes for electrical engineering, second ed., Addison Wesley, 1994.

[24] D. Li, X. Sun, Nonlinear Integer Programming, Springer, 2006.

[25] L. Li, M. Pal, Y.R. Yang, Proportional fairness in multi-rate wireless lans, in: IEEE INFOCOM, 2008.

[26] J. Liu, Y. Yi, A. Proutiere, M. Chiang, H.V. Poor, Towards utility-optimal random access without message passing, Wireless Communications and Mobile Computing 10 (1) (2010) 115–128.

[27] A. Mishra, S. Banerjee, W. Arbaugh, Weighted coloring based channel assignment for WLANs, ACM SIGMOBILE Mobile Computing and Communications Review 9 (3) (2005) 19–31.

[28] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, W. Arbaugh, A client driven approach for channel management in wireless lans, in: IEEE INFOCOM, 2006.

[29] J. Mo, J. Walrand, Fair end-to-end window-based congestion control, IEEE/ACM Transactions on Networking 8 (8) (2000) 556–567.

[30] ns-2, <http://www.isi.edu/nsnam/ns/>.

[31] D.P. Palomar, M. Chiang, A tutorial on decomposition methods for network utility maximization, IEEE Journal on Selected Areas in Communications 24 (2006) 1439–1451.

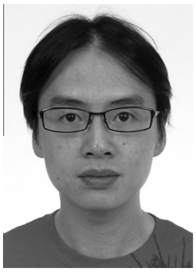
[32] K. Ramachandran, E. Belding, K. Almeroth, M. Buddhikot, Interference-aware channel assignment in multi-radio wireless mesh networks, in: IEEE INFOCOM, 2006.

[33] A. Raniwala, T. Chiueh, Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network, in: IEEE INFOCOM, 2005.

[34] D.B. Shmoys, E. Tardos, An approximation algorithm for the generalized assignment problem, Mathematical Programming 62 (3) (1993) 461–474.

[35] V. Shrivastava, N. Ahmed, S. Rayanchu, S. Banerjee, S. Keshav, K. Papagiannaki, A. Mishra. CENTAUR: realizing the full potential of centralized WLANs through a hybrid data path, in: ACM MobiCom, 2009.

[36] L. Xie, Q. Li, W. Mao, J. Wu, and D. Chen, Achieving efficiency and fairness for association control in vehicular networks, in: IEEE ICNP, 2009.



Ka-Lok Hung received his B.Eng., M.Phil. and Ph.D. degrees in Computer Science and Engineering in 2004, 2006 and 2011 respectively from the Hong Kong University of Science and Technology (HKUST). His research interests include cross-layer optimization, mathematical modelling and performance evaluation of multi-cell WLANs and multi-hop wireless networks. He served on the technical program committees of IEEE GLOBECOM 2007, IEEE WCNC 2010, IEEE INFOCOM 2011 Workshop on Cognitive and Cooperative Networks and IEEE ICCT 2011, and has reviewed extensively for IEEE Communications Letters, IEEE Wireless Communications, Elsevier Computer Networks, Elsevier Performance Evaluation and Annals of Telecommunications. In 2009, he received the Professor Samuel Chanson Teaching Assistant Award at the HKUST.

Brahim Bensaou received an Engineering Degree in Computer Science (with distinction) from the University of Science and Technology Houari Boumediéne of Algiers Algeria in 1982, and a Diplôme d́Etudes Approfondies degree from University Paris XI in Computer Science in 1988. He earned his Doctorate degree in Computer Science from the University Paris VI in 1993. From 1990 to 1994, he was a research assistant at France Telecom Research labs at Issy-les-Moulineaux near Paris, where he was involved in the early designs and performance evaluation of ATM technology. In mid 1995 he joined the Hong Kong University of Science and Technology as a Research Associate where he spent nearly 2 years working on various problems in congestion and traffic control. In 1997 he joined the Centre for Wireless Communications, a national R&D center in Singapore (now under the Institute of Infocomm Research, I2R, A-Star) as a Member of Technical Staff, where he worked as a System Architect on the design of Quality of Service (QoS) enabled MAC protocols and scheduling algorithms in a wireless ATM network prototype. In 1998 he was promoted to a Senior Member of Technical Staff, and was instrumental in forming a small R&D group in the area of wireless networking at the CWC. He led the group, for a year and half, until he moved to Academia at the HKUST in the fall of 2000 where is now a faculty member in the Department of Computer Science and Engineering. He has published more than 100 research papers, received numerous research grants, graduated more than 20 postgraduate students and co-invented 3 US patented inventions of which one was licensed in 2006. Prof. Bensaou served as, an Associate Editor of IEEE Communications Letters, the technical program chair of the IEEE International symposium on Wireless Pervasive Computing 2011, the technical program committee co-chair of three consecutive ACM conference on Modelling Analysis and Simulation of Wireless and Mobile Systems. He guest-edited several special issues for OSA Journal of Optical Networking, Elsevier Computer Networks, Performance Evaluation, and Ad Hoc Networks. Prof. Bensaou is a Senior member of the IEEE, and a member of the ACM.