Centralized Optimization of the Association in IEEE 802.11 Networks

Mohammed Amer

N° d'ordre NNT : 2018LYSE1238

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED 512

Spécialité de doctorat : Informatique

Soutenue publiquement le 12/11/2018, par :
Mohammed AMER

Centralized Optimization of the Association in IEEE 802.11 Networks

Devant le jury composé de :

Monsieur Thomas Noel, Professeur des universités, Université de Strasbourg	Rapporteur
Monsieur Andrzej Duda, Professeur des universités, Grenoble INP-Ensimag	Rapporteur
Monsieur Philippe Martins, Professeur des universités, Télécom ParisTech	Examinateur
Madame Katia Jaffres-Runser, Maître de Conférences, INPT-ENSEEIHT	Examinatrice
Madame Laure Gonnord, Maître de Conférences, Université Lyon 1	Examinatrice

Monsieur Anthony BUSSON, Professeur des universités, Université Lyon 1 Directeur de thèse
Madame Isabelle GUÉRIN LASSOUS, Professeur des universités, Université Lyon 1 Co-directrice de thèse

Résumé

Dans cette thèse nous étudions la problématique de l'association dans les réseaux Wi-Fi. Nous proposons des solutions qui permettent à un contrôleur d'optimiser et de gérer d'une manière centralisée les opérations d'association et de réassociation. L'association est exprimée sous forme de problèmes d'optimisations combinatoires. Les modèles proposés tiennent compte des interférences entre les APs et sont conformes avec la méthode d'accès DCF du standard IEEE 802.11.

Dans le premier modèle proposé nous avons considéré le cas d'un réseau saturé dans lequel on suppose que chaque AP dispose en permanence de trames à transmettre. Dans ce modèle, nous avons supposé que toutes les stations d'un même AP ont des chances de transmission équivalentes autrement dit le même nombre d'accès au medium. La fonction objectif proposée offre un bon compromis entre l'amélioration du débit des stations et l'équité. Les résultats numériques obtenus sur des simulations réalistes ont montré l'efficacité de cette solution et présentent une amélioration significative des performances du WLAN par rapport à une association basée sur la valeur du RSSI ou par rapport aux approches existantes.

Par la suite, étant donné que l'hypothèse d'un réseau tout le temps saturé n'est pas très réaliste, nous avons proposé une solution qui s'appuie sur des mesures réelles telles que les demandes de débit des stations et les taux d'erreur. Notre solution cherche à équilibrer la charge entre les APs. Plus précisément, nous cherchons à diminuer la charge de l'AP le plus chargé dans le WLAN. Pour évaluer cette charge, nous avons proposé un modèle mathématique qui permet d'estimer le BTF « Busy Time Fraction » d'un AP dans n'importe quelle configuration (schéma d'association). Ce modèle est basé sur un réseau de Markov. Le modèle associé au problème d'optimisation permet de proposer la meilleure association. L'évaluation de cette solution par simulation a montré à quel point notre estimation du BTF est précise, et a aussi montré sa capacité à équilibrer la charge entre les APs et à satisfaire la demande en débit des stations.

Pour généraliser cette solution aux nouvelles versions du standard IEEE 802.11 comme 802.11n/ac, nous avons adapté le modèle d'estimation du BTF pour qu'il tienne compte des nouvelles améliorations apportées par les couches physiques et MAC du Wi-Fi telles que l'agrégation des canaux, l'agrégation des trames et le bloc d'acquittement. Ainsi, nous avons proposé une nouvelle métrique qui permet d'exprimer à la fois le BTF d'un AP et les taux d'agrégation de trames de chacune de ces stations. L'évaluation numérique de cette solution a montré l'avantage de la nouvelle métrique par rapport au BTF pour améliorer le débit des stations et l'équilibrage de charge dans le WLAN.

Il est à noter que, pour la résolution des problèmes d'optimisation combinatoire formulés dans cette thèse, nous avons utilisé des heuristiques de recherche locale itérative. Ces heuristiques sont basées sur une même structure de voisinage, mais les procédures de recherches sont différentes selon la fonction objectif de chaque modèle. Ce choix est justifié par l'efficacité de la recherche locale à fournir des solutions acceptables dans un temps raisonnable pour des problèmes d'optimisation combinatoire complexes.

Abstract

In this thesis we study the problem of association in Wi-Fi networks. We propose solutions that allow a controller to optimize and manage in a centralized way the operations of association and reassociation. Association is expressed as combinatorial optimization problems. The proposed models consider interference between APs and are compliant with the DCF access method of the IEEE 802.11 standard.

In the first model proposed we considered the case of a saturated network in which it is assumed that each AP permanently has frames to transmit. In this model, we have assumed that all the stations of the same AP have equivalent chances of transmission, ie the same number of accesses to the medium. The proposed objective function offers a good compromise between improving station throughput and equity. The numerical results obtained on realistic simulations have shown the effectiveness of this solution and show a significant improvement in WLAN performance compared to an association based on the value of the RSSI or compared to existing approaches.

Subsequently, since the hypothesis of a saturated network all the time is not very realistic, we have proposed a solution that relies on real measurements such as station throughput requests and the error rates. Our solution seeks to balance the load between APs. Specifically, we seek to reduce the load of the most heavily loaded AP in the WLAN. To evaluate this load, we have proposed a mathematical model that allows to estimate the BTF "Busy Time Fraction" of an AP in any configuration (association scheme). This model is based on a Markov network. The model combined with the optimization problem allows to propose the best association. The evaluation of this solution by simulation has shown how accurate our BTF estimation, and has also shown its ability to balance the load between APs and satisfy the station throughput demands.

To generalize this solution to the new versions of the IEEE 802.11 standard such as 802.11n/ac, we adapted the BTF estimation model to take into account the new improvements

made by Wi-Fi on physical and MAC layers such as channel aggregation, frame aggregation and block acknowledgment. Thus, we have proposed a new metric that allows to express both the BTF of an AP and the frame aggregation rates of each of its stations. The numerical evaluation of this solution showed the advantage of the new metric compared to the BTF to improve the station throughputs and load balancing in the WLAN.

It should be noted that, for the resolution of the combinatorial optimization problems formulated in this thesis, we used iterative local search heuristics. These heuristics are based on the same neighborhood structure, but the search procedures are different depending on the objective function of each model. This choice is justified by the effectiveness of local research in providing acceptable solutions in a reasonable time for complex combinatorial optimization problems.

Contents

List of Figures

List of Tables

Introduction

General context

Wireless communications have become the most used means of communication to access the Internet. We distinguish two types of wireless technologies. The first one is the cellular wireless network that have evolved through different generations from 2G (GSM) to 4G (LTE) and 5G. These networks are deployed by telecom operators and ensure a very broad coverage almost offering a permanent and continuous connection to mobile users. The recent popularity of smartphones and the increasing demand for multimedia contents have boosted users' traffic and their requirements in terms of quality of service, especially throughput and latency. But, most of the time, the users are charged based on the volume of exchanged data and an intensive use of the cellular network may become very expensive to the customers.

The second type of wireless technologies is the wireless local area networks (WLANs). It relies on the deployment of Wi-Fi access points (APs) that are based on the IEEE 802.11 standard. Wi-Fi has been defined to operate on free frequency bands (ISM and U-NII). These technologies have experienced a rapid and spectacular development. They proposed transmission rates ranging from 1 $Mbps$ for the first standard defined in 1997 to 7 $Gbps$ for the latest standard defined in 2016 (including the IEEE 802.11ac amendment). These transmission ranges coupled to a relative simplicity of installation have quickly attracted users and operators for its use not only in homes, but also in business and public areas around the world.

In order to use transmission rates in concordance with the increase of traffic demand, advanced transmission techniques and channels merge have been recently standardized (IEEE 802.11n [1] and 802.11ac [2]). The high transmission rates offered by these technologies rely also, in practice, on a densification of APs belonging to the same Wi-Fi network, i.e. with the same Extended Service Set ID (ESSID). Such a densification aims to ensure an efficient coverage of a Wi-Fi network and allows stations (STAs) to have at least an AP in their close vicinity, which guarantees high transmission rates. This densification is also amplified by the increase of the number of different Wi-Fi networks (different ESSID or SSID), and a STA may observe, in its radio range, a large number of Wi-Fi networks [3, 4]. But the increase of transmission rates and densification are not the only factor that can provide a capacity growth of the wireless access. Channel spatial reuse and transmission rates cannot increase unlimitedly and their usage must be consequently optimized.

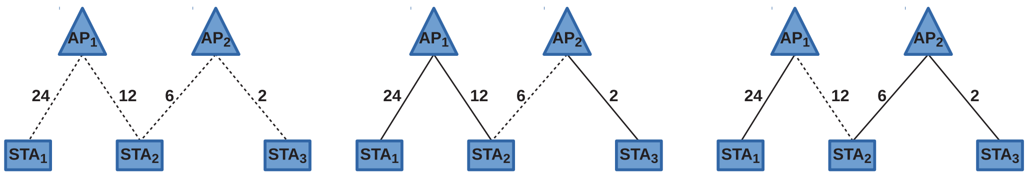
The recent trend of the Wi-Fi technology market offers a technological framework that allows such optimizations. Wi-Fi architectures have, for most of the products, a mode where a controller is in charge of a Wi-Fi network (consisting of a set of APs) in terms of configuration, management and optimization. This breaks the traditional approach where APs take all their decisions (association, channel selection, etc.) in an autonomous way. Even if most of the existing solutions are proprietary, different standards have been proposed to offer common protocols to support these centralized services, like for instance, CAPWAP [5] standardized by IETF, and IEEE 802.11v [6]. Furthermore, some propositions suggest to manage Wi-Fi networks through the SDN paradigm [7].

Thesis statement

Many people think that Wi-Fi is a plug-and-play technology, but it is actually quite complex. To work optimally, it requires a deep understanding of the underlying technology. Indeed, the deployment of a large number of APs without a rational design and a dynamic management of their configuration may lead to very poor performance. The design and management of Wi-Fi networks are related to the APs placement, channel assignment, power control, the association of STAs, etc.

Performance of WLANs has been much studied these two last decades. Many works study the impact, on the performance, of different operations, like for instance, the channel assignment, the transmission power assignment or the association. The literature mainly addressed the two firsts problem, even if the association problem begins to be more and more considered. Indeed, it appears that association has a non negligible impact on the performance of the WLAN. The association problem can be considered for a given deployment (e.g. for a given channel allocation) or combined with other design parameters.

The association with an AP is the first step that allows a STA to connect to the WLAN. Generally, in current Wi-Fi networks, the STA selects the AP with the best signal quality. The distribution of users among APs of the same Wi-Fi network is then dependent on the geographical locations of users and radio environment (path-loss, fading/shadowing) and not on the number of STAs already associated to each AP nor their throughput demands. It may lead to poor performance for users as they may be attached to the same AP whereas some other APs are idle [8]. Instead, the controller that manages the network dynamically can distribute STAs among APs in a centralized manner to optimize a given objective function, like, for instance, maximizing the overall throughput of the network. To illustrate the impact that association can have on the STAs and overall network performance, we present a motivating example illustrated in Figure 1.



(a) Data rate. (b) Default association. (c) Optimized association.

Figure 1: WLAN network example. (a) Data rates between APs and STAs, (b) Default association based on RSSI value, overall throughput = 18. (c) Optimized association to maximize overall throughput, overall throughput = 27.

We consider two APs (AP_1 and AP_2) and three STAs (STA_1, STA_2 and STA_3). The two APs use two different/orthogonal channels. The data rates of the wireless links between each AP and STA are indicated through the dotted lines in Figure 1a.

The default association is illustrated in Figure 1b where each STA connects to its AP with the highest value of RSSI. It is represented through the hard lines. If the AP has always packets to send to its STAs, then, in practice, the AP serves their STAs with equal throughput (as discussed later in this thesis). With this assumption, the throughput of an AP is equal to the harmonic mean of the data rates of its STAs. For instance, for AP_1 the throughput equals $2/(1/24 + 1/12) = 16$ and for AP_2 the throughput is 2. The overall throughput for this association is then equal to $16 + 2 = 18$.

The optimized association is illustrated in Figure 1c. The association is optimized to maximize the overall network throughput. The change with regard to the default association is the STA_2 which is associated with AP_2 rather than AP_1. The throughput for AP_1 is then

24 and $2/(1/6+1/2) = 3$ for AP_2. The overall throughput becomes $24+3 = 27$ leading to an improvement of 50% in terms of global throughput.

Contributions

In this thesis we study the problem of association in Wi-Fi networks. We propose solutions that allow a controller to optimize and manage, in a centralized way, the operations of association and reassociation. Association is expressed as combinatorial optimization problems. The proposed models consider interference between APs and are compliant with the DCF access method of the IEEE 802.11 standard.

First contribution: the first model considers the case of a saturated network where APs have always a frame to transmit. In this model, we assumed that all the STAs associated to the same AP have the same chance/probability of transmission, and thus the same number of medium accesses in average. The proposed objective function offers a good trade-off between the overall throughput and fairness. The numerical results obtained with the network simulator ns-3 have shown the effectiveness of this solution and a significant improvement of the WLAN performance compared to the default association based on the RSSI but also with regard to the existing solutions proposed in the literature.

Second contribution: we release the assumption of a saturated network. We have proposed a solution that relies on real traffic measurements. It is based on traffic demand or the observed traffic for each STA and the frame error rates. Our solution seeks to balance the load between APs while verifying the traffic demands. More precisely, the optimization problem aims to reduce the load of the most heavily loaded AP in the WLAN. To evaluate this load, we have proposed a mathematical model that estimates the BTF "Busy Time Fraction" of an AP for any configuration (association scheme). This model is based on a Markov network. It is combined to an heuristic that explores a subset of solutions and propose a better (often the best) association. The evaluation of this approach by simulation has shown the accuracy of our BTF model and thus its capacity to infer the network load for any association. Also, results of the optimization problem showed its ability to balance the load between APs and satisfy the STAs' demand.

Third contribution: to generalize the second contribution to the new versions of the

IEEE 802.11 standard such as 802.11n / ac, we adapted the BTF estimation model to take into account the new improvements brought by these standards as channel aggregation, frame aggregation and block acknowledgment. We have proposed a new metric that allows to express the load through the BTF of an AP and the frame aggregation rate of each STA. The numerical evaluation of this solution showed the benefit of the new metric compared to the BTF to improve the STAs throughput and the load balancing in the WLAN.

Fourth contribution: to solve the combinatorial optimization problems formulated in this thesis, we used iterative local search heuristics. These heuristics are based on the same neighborhood structure, but the search procedures are different depending on the objective function of each model. This choice is justified by the effectiveness of local search in providing acceptable solutions in a reasonable time for complex combinatorial optimization problems.

Manuscript organization

In this manuscript, we present in Chapter 1 the WLANs based on the IEEE 802.11 standard (Wi-Fi), their mode of operation, their evolution and their main characteristics. We present a state of the art dealing with association in WLANs in Chapter 2. In this chapter we establish a classification of existing works according to several criteria. It is followed by a benchmark of our solutions with regard to these works. In the same chapter, we give the general description of the network model, the generic assumptions and the notations used in the different mathematical models.

In Chapter 3, we present our first association model. This model is proposed for a scenario with a saturated WLAN and where we assume that the numbers of accesses for each STA associated with the same AP are fair. In Chapter 4, we present our second solution to the association problem. It considers an unsaturated WLAN. The proposed model is based on the demands of STAs in terms of throughput. In Chapter 5, we present our model that generalizes the solution presented in Chapter 4 for the recent versions of Wi-Fi such as IEEE 802.11n/ac.

We conclude in the last chapter: we summarize the main results of this thesis and give possible extensions of this work.

Chapter 1

IEEE 802.11 Networks

Sommaire

The 802.11 standard defines two operating modes: infrastructure and ad-hoc modes. With the infrastructure mode, a STA must associate with a central node called access point (AP) to connect to the network. The STA is associated to a single AP and its transmissions are sent to this AP or received from this AP. Instead, in ad-hoc mode, a STA may transmit frames to any other node in the same mode and within its transmission range.

In this thesis, we consider only the infrastructure mode. Therefore, in this chapter, we describe the main features of this mode and its operations.

1.1 DCF Access Mode

The IEEE 802.11 standard proposes the Distributed Coordination Function (DCF) [9]. It is implemented by each STA and AP to access and share the medium.

This mode is based on the Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA) method. CSMA/CA is a Media Access Control (MAC) protocol for wireless networks. Basically, it consists in verifying the presence of other transmission before transmitting.Thus, DCF allows frames transmission only when the medium is detected as idle. This reduces the risk of collision, i.e. the fact that two frame transmissions overlap, probably leading to the bad reception of one or both frames.

In wireless networks, collision detection is not possible by a transmitter node during its transmission. To detect a collision, a source node should be able to listen and transmit at the same time, but the transmission prevents the STA from listening at the same time. Therefore, with DCF, after the correct reception of a frame, the receiver transmits an acknowledgment (Ack) back to the sender. If the sender does not receive the Ack after a certain period, called acknowledgment time out (AckTimeOut), it considers that the frame has not been correctly received and must retransmit it.

1.1.1 Clear Channel Assessment

For determining whether the medium is idle or not, DCF implements the Clear Channel Assessment (CCA) mechanism. It includes two modes: carrier sensing and energy detection. In carrier sensing, CCA measures the received signal strength for valid 802.11 symbols, whereas for energy detection the energy level on the medium is measured and compared to a given threshold. The thresholds for both carrier sense and energy detections are predefined in the standard. As an alternative to carrier sensing, the network allocation vector (NAV) is updated with the duration field given in the frame header and used to inform the other nodes about the time of the current transmission. A NAV with a non null value indicates that there is no need to sense the channel during this period.

Furthermore, DCF is a time-based system that uses three key sets of time durations, the slot time, inter-frame space, and the contention window. In the next paragraph, we detail these different timers.

1.1.2 Slot Time

The slot time is the basic unit of time measure for DCF. This value is different in the different versions of IEEE 802.11. IEEE 802.11b specifies a slot time of $20\mu s$, while more recent specifications including IEEE 802.11a/g/n/ac define a shorter slot time of $9\mu s$.

1.1.3 Inter Frame Space

To manage the access to the medium for the STAs, DCF sets up a system of priorities between the different types of frames. This system relies on an Inter-Frame Spacing (IFS) mechanism as shown in Figure 1.1. An IFS is a period of inactivity between two successive frames. There are four types of IFS:

- Short Inter-Frame Spacing (SIFS) is used to separate the different frames transmitted within the same transmission opportunity, i.e. the consecutive frames sent after a allowed medium access, for example, between data frames and their Ack or between different fragments of the same frame.

- DCF Inter-Frame Spacing (DIFS) is the time that a STA must wait, during the medium access process, before sending a frame in DCF mode. If the medium is continuously detected idle for DIFS duration, the STA is allowed to transmit. Otherwise, it must defer its transmission. The value of the DIFS is equal to that of a SIFS plus two slot times.

- PCF Inter-Frame Spacing (PIFS) is the time that a STA must wait before sending a frame in PCF mode. This mode is explained below. The value is lower than the DIFS to favor this mode against DCF.

- Extended Inter-Frame Spacing (EIFS) is the longest IFS. When a STA receives an erroneous frame, it must wait during an EIFS when the medium becomes idle. It gives the opportunity to the intended receiver (of the erroneous frame) to return an Ack. Indeed, the erroneous frame may be intended to another STA for which the following Ack is not necessarily received/detected.

To reduce overhead and thereby increase network efficiency, IEEE 802.11n introduces a new IFS called Reduced Inter-Frame Spacing (RIFS). RIFS may be used in place of SIFS to separate multiple High Throughput (HT) transmissions from a single transmitter when no SIFS separated response transmission (like an Ack) is expected from the receiver.

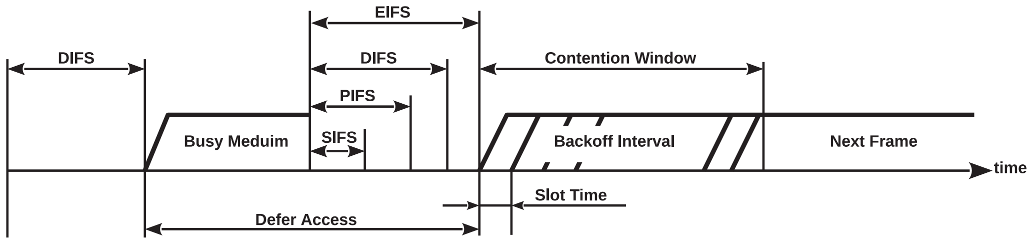


Figure 1.1: Access in DCF mode.

1.1.4 Contention Window

If the medium is sensed busy, the STA has to wait for a free DIFS, then the STA must additionally wait for a backoff interval (see Figure 1.2). The backoff interval is a number, chosen randomly in a predefined range called Contention Window (CW). The number represents the amount of slot times during which the STA must detect the medium idle before it may attempt to start its transmission. The backoff number decrements each time the medium has been sensed idle during a time slot. When the backoff expires, i.e. the backoff value is set to 0, the STA can begin its transmission. It is possible for two concurrent STAs to choose the same backoff interval, which may result in a collision. In case of unicast transmission, CW is doubled with each retransmission attempt of a frame until a maximum is reached. After a fixed number of unsuccessful retransmissions, the frame is rejected. In the case where a frame is transmitted successfully, the CW is set to its minimum initial value.

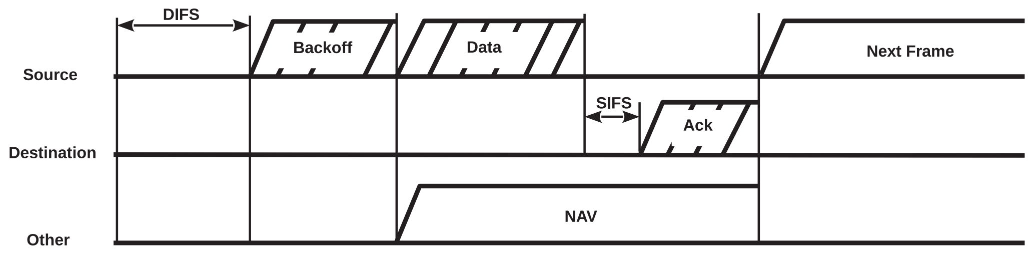


Figure 1.2: Frame transmission with DCF mode.

1.2 Other access modes

The 802.11 standard defines other media access methods. The Point Coordination Function (PCF) mode is a contention-free mode that works by probing STAs and giving them the opportunity to send information without conflicting with other equipments. PCF mode is optional and almost never implemented by the manufacturers.

The other access method "Hybrid Coordination Function" (HCF) has been introduced with 802.11e for QoS requirements. The HCF consists of two channel access methods for the support of differentiated QoS. One of them is a contention-based channel access method named Enhanced Distributed Channel Access (EDCA), and the other one is HCF Controlled Channel Access (HCCA). EDCA is the fundamental and mandatory method of IEEE 802.11e and delivers traffic based on differentiated Access Categories (ACs). HCCA is optional and requires centralized polling and scheduling algorithms to allocate the resources.[10].

The wireless mesh networks (IEEE 802.11s) includes an additional access mode called Mesh Coordination Function (MCF). Mesh STAs shall implement the MCF only. MCF has both a contention-based channel access and contention free channel access mechanism. The contention based mechanism is EDCA that is mandatory and the contention free mechanism is called the MCF Controlled Channel Access (MCCA). MCCA is an optional access method that allows mesh STAs to access the WM at selected times with lower contention than would otherwise be possible.

While PCF is optional, HCF has been designed only to support differentiated QoS requirements over IEEE 802.11 WLANs and MCF that is usable only in an MBSS, this thesis covers only the mandatory DCF access method.

1.3 IEEE 802.11 evolution

In 1999, Apple launched the first commercial product using this new standard by adding a Wi-Fi interface in its new computers, called "AirPort". The use of Wi-Fi was quickly widespread in businesses and homes. In 2005, Wi-Fi shipments reached 100 millions per year and in 2009, 1 billion Wi-Fi chipsets were sold. It is estimated that more than 27 billion units will be equipped with a Wi-Fi interface worldwide in 2021 [11]. Wi-Fi is based on the IEEE 802.11 standards.

The working Group 11 of IEEE (IEEE 802) LAN / MAN Standards Committee have published the first standard in 1997. This standard (IEEE 802.11) uses the unlicensed 2.4 GHz radio frequency spectrum which is taken from the industrial, scientific and medical (ISM) band. This first version of Wi-Fi, which uses frequency hopping, allows data transfer of up to 2 Mbps.

The IEEE working group ratified 802.11a and 802.11b standards in 1999. The 802.11a standard used the same data link layer protocol and frame format as the original standard. However, 802.11a added to the physical layer an air interface based on Orthogonal Frequency Division Multiplexing (OFDM). The 802.11a standard exploits the 5 GHz band that is taken from the Unlicensed National Information Infrastructure (U-NII) band with a maximum throughput of 54 Mbps.

The IEEE 802.11b standard that uses at the physical layer the Direct Sequence Spread Spectrum (DSSS) technique, offers a maximum data rate of 11 Mbps. Due to the cost and complexity of implementation, most early commercial deployments used the 802.11b standard which implemented a simpler modulation technique with 22 MHz bandwidth.

The third generation, IEEE 802.11g, was released in 2003. It operated in the 2.4 GHz frequency band such as IEEE 802.11b, but could also reach a speed of 54 Mbps using the same OFDM technology as the IEEE 802.11a standard. The 802.11g standard was backwards compatible with 802.11b hardware. Thus, the transition to this new technology was easy because the user could keep some of his equipment during the transition to new standards.

In 2009, amendment 802.11n was published. The IEEE 802.11n standard has introduced several new approaches to improve WLAN performance. First, at the physical layer, the 802.11n standard has advanced both in terms of data rate (reaching 450 Mbps in the 2.4 GHz band) and reliability by using more antennas and multiple data streams. Multiple

Input Multiple Output (MIMO) technology uses multipath propagation to send and receive multiple data signals simultaneously on the same frequency. At the MAC layer, the frame aggregation mechanism is used so that multiple frames or datagrams are aggregated into a single frame before transmission. The 802.11n standard supports the 2.4 GHz and 5 GHz frequency bands and is backward compatible with older 802.11a/b/g technologies.

The IEEE working group has released two amendment of the 802.11ac standard. The first one in 2014 where a device was capable of transmitting at 1.3 Gbps, while the second amendment in 2015 was announcing an incredible rate of 3.4 Gbps. 802.11ac devices achieve nearly eight times the speed of 802.11n by leveraging beamforming technology to increase the number of MIMO streams and increase the channel bandwidth through optimized OFDM modulation (up to QAM 256). The 802.11ac standard only supports 5 GHz frequency bands, so, is backward compatible with previous 802.11a/n versions.

802.11n/ac standards are revolutionary departures for WiFi. The IEEE 802.11n/ac have introduced several new approaches to improve WLAN performance. Conceptually, 802.11ac is an evolution from 802.11n. To improve efficiency and increase speed, many of the techniques, either for the physical layer or the MAC layer, used in 802.11ac are taken from 802.11n and improve them to a new level, except with one exception. Rather than using MIMO only to increase the number of data streams sent to a single client, 802.11ac is pioneering a multi-user form of MIMO that enables an AP to send to multiple STAs at the same time.

In the next section, we will detail more specifically the new features of IEEE 802.11n standard. It is important to describe these features and discuss them, as our models, which will be presented in the following sections, are based on its behavior.

1.4 Features of IEEE 802.11n

At the physical layer, a 802.11n STA can be equipped with multiple antennas and MIMO is used to increase data rates and reliability by transmitting multiple spatial streams simultaneously or by exploiting the spatial diversity. In addition, the maximum coding rate is increased from 3/4 to 5/6 and a short guard interval (400 ns) between orthogonal frequency division multiplexing (OFDM) symbols is introduced to improve spectral efficiency, and therefore the maximum physical flows. In addition, a wider channel link is applied to further enhance data rate by combining contiguous and non-overlapping 20 MHz channels (40 MHz channel for 802.11n and 80 - 160 MHz form 802.11ac). The combination of the spatial streams, the modulation type and the coding rate are represented by a Modulation and Coding Scheme index (MCS). Coupled to the channel width and the guard interval length, different data rates are then possible for transmitting the frames.

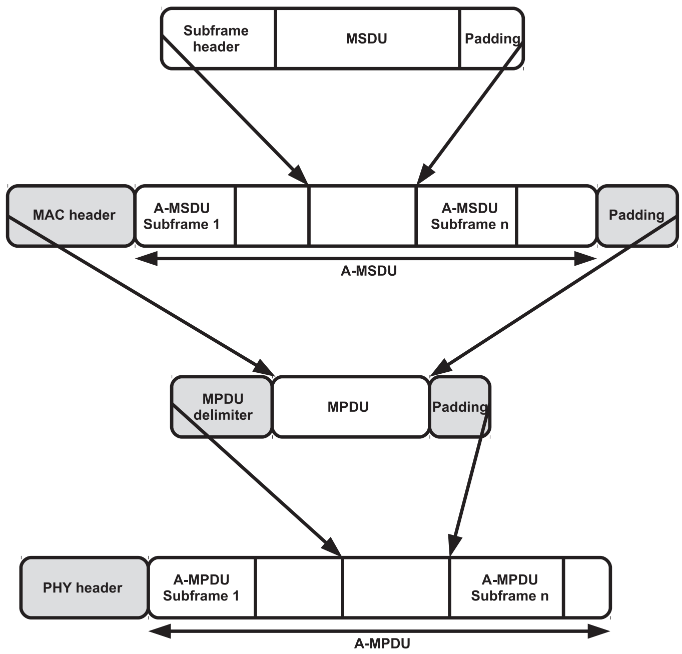


Figure 1.3: Two level aggregation.

At the MAC layer, two aggregation mechanisms have been proposed (see Figure 1.3). Aggregate MAC Service Data Unit (A-MSDU) is an aggregation of several SDUs (Service Data Units, typically IP packets) with one common MAC header. Aggregate MAC Protocol Data Unit (A-MPDU) is a scheme where several IEEE 802.11 frames or A-MSDU are aggregated into a single packet.

We focus on A-MPDU because it is mandatory in the standard and implemented in the products, whereas A-MSDU is optional [12]. Each A-MPDU aggregated frame is acknowledged by a block ACK frame, in which a bitmap is used to acknowledge the subframes. In this way, both the MAC overhead (due to the numerous acknowledgments) and the duration to access the medium (DIFS, SIFS, Back-off, etc.) are reduced with regard to the number of transmitted bits. Ns-3 simulations with MCS 7 (corresponding to a maximal physical transmission rate of 150 Mbps) show that a throughput of 42.5 Mbps is attainable without aggregation and 136 Mbps with aggregation. Obviously, aggregation is a prerequisite to benefit from the physical transmission rate increase offered by the new standards.

1.5 Centralized management

The control plan of WLAN is responsible for ensuring mobility between APs, coordinating the selection of channels and associating users, among other tasks. Much of the development of WLANs has involved refining the control plan. The first WLANs were built from completely independent access points. The control plan was practically nonexistent. Networks based on autonomous APs did not automatically select channels and did not always support the smooth handoff between APs without proprietary solution.

The development of WLAN controllers in recent years has led to a redesign of the way networks have been built, and the control plans being centralized in new WLANs. In a typical controller-based deployment, APs have limited functionality without connection to the controller. Access control and authenticating are handled by the controller, as are the channel selection and STA association. Centralized control has made WLANs much more important, and nearly every large-scale WLANs are now built using a controller-based architecture.

1.6 Wireless Network Design

The design of wireless LANs applies to a large number of parameters as the number of APs and their location, the frequencies/channels to assign, the associations between wireless STAs and APs, etc. Optimization models and algorithms aim to reduce costs and to improve performance through a better management of the available resources. Unfortunately, even with the development of computational technologies and parallel processing, many of these problems cannot be solved optimally in a reasonable computational time because of their internal nature or size. Consequently, the challenge is to design algorithms or heuristics offering a good trade-off between optimality and complexity. A more complete description of these problems, models, as well as the main optimization tools are provided in [13, 14, 15, 16].

1.7 IEEE 802.11 WLAN

In infrastructure mode each STA associates to an AP. The set formed by the AP and its associated STAs is called a basic service set (BSS) and constitutes a cell. Each BSS is identified by a BSSID. In infrastructure mode, the BSSID is the MAC address of the AP.

Several APs may be connected through the distribution system (DS) and form a unique

Wi-Fi network. It constitutes an extended service set (ESS). The DS can be a wired network with cables between APs or even a wireless network as illustrated in Figure 1.4.

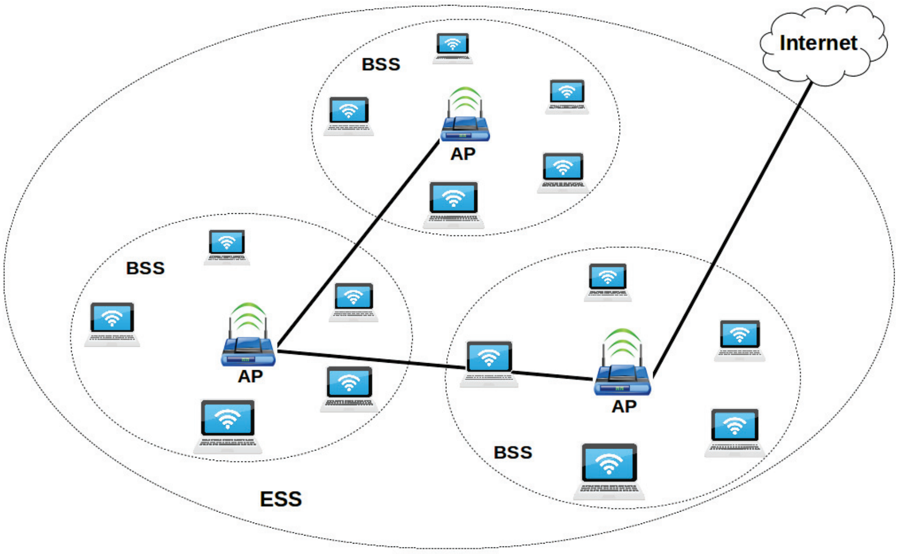


Figure 1.4: Extended Service Set.

An ESS is identified by an Extended Service Set Identifier (ESSID), which is a 32-character identifier (in ASCII format) used as the name for the network. The ESSID, often abbreviated as SSID, is the name of the network ("eduroam" for instance). In some way, it is a first level of security as the knowledge of the SSID is required for a STA to connect to the ESS.

When a STA moves within the ESS, it is able to associate to a new AP belonging to the same ESS in order to keep a good link quality. The APs communicate with each other through the DS in order to exchange information on the STAs and, if necessary, to transmit the data from the mobile STAs. This feature allowing STAs to "seamlessly pass" from one AP to another is called handover.

1.8 Association process

The APs are bridges that carry the traffic between mobile STAs and Internet. Before an STA can send traffic through an ESS, it must be associated with an AP. The association process is achieved by exchanging a series of 802.11 management frames between the STA and the selected AP as presented in Figure 1.5.

When a STA enters an ESS, at first, it looks for an AP to associate (point 1 in the figure). This step is done according to two modes. In the active mode, an STA sends a probe

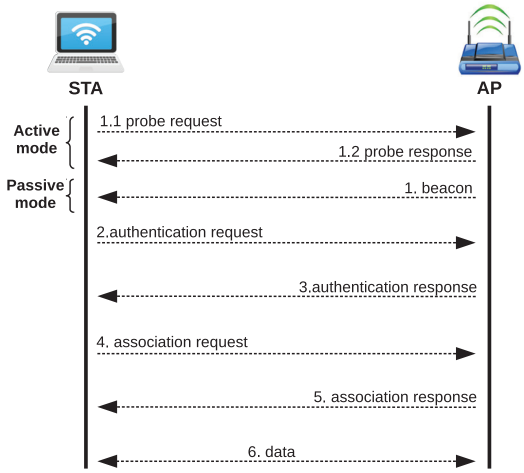


Figure 1.5: Association process.

request on each channel to discover 802.11 networks within its radio range. Probe requests advertise the SSID for which it is configured and supported data rates. The AP receiving the probe request checks the SSID and supported data rates. If their SSIDs match and they have compatible data rates, the AP sends a probe response advertising the SSID, supported data rates, encryption types if required, and other IEEE 802.11 capabilities of the AP. If no SSID is configured, the STA uses the passive mode. Each AP broadcasts regularly (approximately every 0.1 seconds) a beacon frame giving information on its BSSID, its characteristics and possibly its SSID (this latter it may be hidden and not given in the beacon frames). The STA listening to the different channels and receiving beacon and/or probe response is then able to list the different BSS, ESS, and their respective characteristics. The STA is thus also able to evaluate the quality of the signal emitted by the APs on the beacon and/or probe response frames. Usually, in a company, campus, or in a public area, a STA may reach several APs belonging to the same ESS and choose the one offering the best signal quality. Indeed, choosing the better in terms of signal quality leads to a better data rate.

Once the STA determines to which AP it wants to associate to, it sends a low-level 802.11 authentication frame to an AP setting in order to be authenticated (point 2 in the figure). The AP that receives the authentication request frame will respond to the STA with an authentication frame configured to open (point 3 in the figure).

If the STA succeeds to authenticate, it sends an association request to that AP (point 4). The association request contains the encryption types if required and other compatible 802.11 capabilities. If the parameters in the association request match the capabilities of the AP (point 5), the AP creates an Association ID for the STA and responds with an association response with a success message granting network access to the STA. Now the STA is successfully associated to the AP and data transfer can begin (point 6).

1.9 Which AP to choose?

Therefore, since in a IEEE 802.11-based infrastructure network, a STA must be associated with one AP to be allowed to use the network, when several APs are available within its reception range, one AP must be selected. By default, most of IEEE 802.11 devices use the received signal strength indicator (RSSI), from the different APs they detect, to choose the AP to associate with. This approach does not consider the number of already attached STAs per AP. It does not consider either the impact of the STA data rates on the user and global throughput. Indeed, STAs using a low data rate occupy the channel longer than the STAs with a high data rate [17]. High rate STAs may then be significantly penalized as low and high rate STAs attached to the same AP tend to have the same throughput. Therefore, the association between STAs and APs is a key step that has an impact on the user performance as well as on the overall wireless network performance. A survey on the association optimization in IEEE 802.11 networks is presented in the next chapter.

Chapter 2

Survey on Association Optimization

Sommaire

2.1 Introduction

In a IEEE 802.11-based infrastructure network, the association with an AP is the first step that allows a STA to connect to a WLAN. Generally, the STA selects the AP based only on the highest value of the received signal strength indicator (RSSI) from the different APs detected. Several works, discussed in this section, claim that the use of the RSSI metric is not an efficient approach and have proposed different solutions for the association.

In this chapter, we propose a set of criteria to classify the different existing approaches in Section 2.2. It is followed by a description of the most recent solutions and their classification in Section 2.3. We describe, in Section 2.4, the network model, common assumptions and the notations used for the mathematical formulation of the various association solutions proposed in this thesis.

2.2 Classification of the solutions

In Table 2.1, we present a taxonomy of these solutions according to different criteria:

- **saturated / unsaturated (Sat/Uns)**: saturation is a key parameter to take into account in the performance study in wireless networks and the association in particular. Saturated network scenario corresponds to a case where devices (STAs and/or APs) have always a frame to send. Such an assumption did not take into account real traffic demands but allows to express the minimum amount of throughput a device can obtain. An unsaturated scenario corresponds to a case where at least one AP is not saturated. In this context, we rather consider the request of downlink/uplink throughput for each STA. Indeed, at given point, there may be some devices that do not have any frame for transmission. This assumption, allows to consider a more realistic scenario.

- **centralized or distributed (Cnt/Dst)**: the association decision can be done in a central or distributed way. If the association is distributed in many IEEE 802.11 products, since each STA locally makes its association choice based on the RSSI metric, the solutions proposed to improve the association are mainly centralized, as shown in Table 2.1. However, there also exist proposals that aim to improve the association based on the RSSI value in a distributed way. Some papers also propose both approaches [18, 19].

- **On-line / Off-line (On/Off)**: in an on-line approach, the association is triggered upon a new event. With this approach, every time there is a change in the network (i.e. a STA moves or leaves the network, the network conditions sensed or measured by the STAs change), the association is reconsidered. On the other hand, in the off-line approach, the solution is periodically executed from each current association in the network.

- **Access-based / Time-based Fairness (AbF/TbF)**: most of these papers consider a specific bandwidth sharing between users. This share has an impact on the throughput that can be obtained by each STA and is part of the optimization model when such a model exists. Some solutions assume that the medium share is fair in time (denoted TbF hereafter), meaning that each STA has the same proportion of times in average. Solutions based on this assumption aim to improve network performance while ensuring fairness in terms of service time between STAs on the same access point. It ensures that each STA obtains a throughput proportional to its physical transmission rate used for transmissions between the AP and the STA. Other solutions consider that the share is fair in the number of accesses to the channel (AbF). This guarantees that all STAs on the same access point receive the same throughput regardless of the used physical transmission rate. Some solutions consider that the share of the radio medium between users is driven by QoS requirements (denoted by 'QoS' in Table 2.1). When considering downlink traffic in a cell, the fairness (in time or in access) is provided by APs that apply an appropriate scheduling to send packets to the STAs according to the targeted fairness model.

- **Objective function / Use of a metric (Obj F/Metric)**: When the solution is based on an optimization model, the solution seeks to optimize an objective function. 'log' in Table 2.1 refers to proportional fairness objective function. In these solutions, the authors look for a proportional fair association by optimizing the sum of logs of the users' throughput. 'Max-AP' refers to minimizing the maximum AP utilization. 'Max-flow' corresponds to the maximum flow problem. With 'Max-min', the goal is to maximize the minimal throughput among all the STAs. When the proposed solution is distributed, it does not rely on an explicit optimization model. It is rather based on one (or more) metrics that is used by each STA to select its AP, like, for instance, the achieved throughput by each STA, the least loaded AP, the signal quality on the packet loss rate.

- **Downlink / Uplink traffic (Down/Up)**: in most of the proposed solutions, only downlink traffic (*i.e.* traffic sent from APs to STAs) is considered in the association process because it represents the majority of traffic flows. Other solutions make no assumption on traffic direction and consider downlink and uplink traffic.

- **Simulation / Experimentation (Sim/Exp)**: most of the proposed solutions are evaluated by simulation, while a limited number of solutions are evaluated experimentally. In many approaches based on an optimization model, solutions are evaluated numerically by using a tool that solves optimization problems, such as CVX in [20], or CPLEX in [21]. In some papers, the designed heuristics are coded and evaluated in Python or in C. In all these papers, only the model/algorithm is implemented and evaluated. The performance evaluation does not consider the behavior of the algorithms in more realistic networking scenarios taking into account the 802.11 DCF principles: realistic radio environment, hidden terminals/APs, IP/TCP layers, etc. On the other hand, only few papers perform realistic simulations, as in [18] where the simulations are realized with the OMNetpp simulator, with the ns-2 simulator in [22] and the ns-3 simulator in [23]. In the very short list of papers describing real experimentations, we can mention the work of [24] that uses the NITOS wireless testbed [24] and the ones of [25, 23] with a homemade testbed.

Paper	Sat	Uns	Cnt	Dst	On	Off	AbF	TbF	Obj F	Metric	Down	Up	Sim	Exp
[8]	X		X		X	X	X			Max-min	X		C	
[20]	X		X			X		X	log		X		X	
[21]	X		X			X		X	log				X	
[24]		X		X	X		X			Throughput	X	X		X
[22]		X		X	X		X			Min load	X		ns-2	
[18]	X		X	X	X	X		X	log	log	X		OMNet++	
[26]		X		X	X		X		Max-AP		X		X	
[27]		X		X	X				Max-flow		X		C	
[25]		X	X		X		QoS			signal quality loss rate	X	X		X
[23]		X	X			X	X		Max-min		X	X	ns-3	X
[19]		X	X	X		X		X	log		X		Python	
[28]		X	X			X		X	log		X		X	
[29], [30]		X	X			X			Power		X	X	Matlab	
[31]	X			X		X			Throughput				X	
[32]	X		X		X			X		Utility function	X		X	
[33]	X		X		X		QoS			Delay, Throughput	X		ns-2	
[34]		X		X	X		FIFO			Utility function	X		ns-3	
[35]	X		X		X		X		Throughput, Handover		X		X	
[36]	X		X			X	FIFO		Throughput			X	Matlab	
[37]	X		X			X	X			SINR	X		OPNET	
[38]	X		X			X	X			Throughput	X		X	
[39]	X			X			X	log		X	X		X	X

Table 2.1: Related work taxonomy with 6 criteria taken into account and described in Section 2.2. The "X" in the Simulation column means that it is either a house-made simulator or there is no detail in the paper about its implementation.

2.3 Description of the solutions

In this section, we provide a description of some recent solutions for the association optimization problem.

In [8], the authors propose a solution to manage user-AP association by ensuring a max-min fair bandwidth allocation. This optimal max-min allocation is obtained through load-balancing techniques. The authors also extend their off-line approach to an on-line solution that computes the off-line solution each time the time elapsed between two calls to the optimization algorithm is longer than a given time threshold or when the maximal load among the access points is greater than an allowed load.

The authors of [20] consider the association problem between users/subscribers of a same provider and their set-top boxes. Most APs in urban areas share the same upstream Internet provider. The idea is then to offer a collaborative solution where a user may associate to any box. The association is solved at the common upstream provider to maximize the provider network throughput as well as user experience. The solution is centralized and formulated as a proportional fairness optimization problem.

The authors of [21] address the association problem in multi-rate wireless LANs. They consider an objective function that achieves overall proportional fairness. This function tries to maximize the total logarithmic utility function expressed in terms of bandwidth, and to provide equal channel occupancy time to each user.

The authors of [24] propose a distributed solution. In this solution, a metric, based on estimation of end user throughput in 802.11 infrastructure networks, is used by each STA. This metric allows the STA to choose an AP that maximizes throughput on both uplinks and downlinks. This solution takes into account the effects of contention and interference in the neighborhood and possible hidden STAs.

To achieve load balancing that guarantees fairness among the STAs, the authors of [22] have proposed a distributed and self-stabilized association scheme in multi-rate WLANs. The proposed scheme gradually balances the AP load in a distributed manner. In this approach, STAs associate to APs according to their load.

The authors of [18] propose an algorithm for the user-AP association to achieve time-based fairness in multi-rate wireless LANs. The problem is formulated as a non-linear programming problem with an objective function that maximizes the total users bandwidth utility in the whole network. The authors also propose a distributed version of their algorithm, initially designed in a central way.

In [26], the authors design a distributed algorithm that consider the problem of optimizing associations in 60-GHz wireless access networks. The objective, in their problem formulation, is to minimize the maximum AP utilization while ensuring balanced and fair resource allocation.

The authors in [27] have formulated the AP association as a max-flow problem to improve the overall throughput, fairness, load balancing and also resilience to client mobility. The proposed distributed model relies on shared local information from multiple APs.

In [25], the author proposes a centralized approach based on fuzzy logic for load balancing. In this solution, the STAs change of access point based on the load of the access points, the signal strength received by each STA and QoS parameters such as loss rate and required deadlines.

Wong et al. [23] propose an AP association that maximizes the minimum user throughput. The problem is subject to constraints on the user migration cost implying overhead in handshaking, authentication and data flow management.

In [19] an association control algorithm is proposed to optimize the throughput in wireless LANs. The authors consider an objective function that maximizes a logarithmic utility function on the overall throughput. The solution achieves load balancing between APs and time-based fairness while considering users traffic demand.

In [28], the association optimization is formulated as a proportional fairness problem. It is solved periodically by a central controller. The solution considers the cost of handovers and the minimum throughput requirements (e.g. video or best-effort traffic) of each user when assigning users to APs.

The authors of [29, 30] propose a model to optimally design green wireless LANs. It consists in minimizing the power consumption of a WLAN, when the load is scarce, by powering-on a subset of APs and associating STAs to them. The model takes into account data rates between STAs and APs, users' mobility, and channel conditions.

To understand and improve the performance of several association control schemes, a theoretical framework that analyzes association problems in vehicular Wi-Fi networks is described in [31]. They formulate the association problem as a non-linear integer program taking into account influence of vehicles' mobility, available effective bit rate from APs, and handover cost. Their offline algorithm is compared to existing online algorithms.

In [32], the problem of AP association in WLAN is formulated through a mixed strategic game with a utility function that maximizes the throughput. They propose to users to move from their positions to improve their throughput. Distances traveled to a new AP are incorporated as a cost in the strategy game.

The authors in [33] propose a solution for differentiated access service selection based on network applications, which are classified into four types according to their QoS requirements. Their approach can be used in an on-line or off-line strategy.

In [34] a utility-based strategy is proposed to select the best AP according to the distance, data rate and delay. These three metrics are normalized between zero and one. Then, an equal weight is given to each metric within the utility function. The AP with the highest utility value is selected.

A multi-objective optimization function that maximizes the download user throughput and minimizes the number of handovers in saturated mode is also proposed in [35].

The authors of [36] propose a centralized approach to improve users' throughput in dense WLAN. They use signal-interference-noise-ratio (SINR) between APs and STAs to control the association. In order to further coordinate interference and increase spatial reuse, an algorithm is proposed to adjust the clear channel assessment (CCA) threshold of the 802.11 MAC protocol in each AP cell to address the problem of overlapped basic service set that degrades overall network throughput.

Taking into account the propagation environment, the authors of [37] investigate the impact of the AP deployments and STA association in dense WLAN on the aggregate throughput.

To maximize the aggregate downlink throughput, Babul et al. propose in [38] a greedy algorithm that jointly deals with the channel assignment and association control while avoiding interference.

Based on the Markov model to estimate the uplink and downlink throughput of clients, the authors of [39] propose an on-line AP association algorithm for 802.11n WLANs with heterogeneous clients. In this approach, authors seek to improve the overall network throughput and fairness by using a logarithmic utility function.

In this thesis, we are interested in centralized association solutions based on explicit optimization models. We have proposed three different models. The first one considers

WLAN in saturated mode. In the second model, we assume WLAN in unsaturated mode. We extend the second model to high throughput WLANs (IEEE 802.11n/ac...) in a third model.

Considering WLANs in saturated mode, many optimization models assume a time-based fairness between the APs and the STAs [20, 21, 18, 19]. This approach does not correspond to most of the AP implementation which uses FIFO queues rather than complex scheduling (necessary for time share). Moreover, to obtain a medium share in time in the whole network, the IEEE 802.11 DCF mode must be modified. Since the DCF mode of IEEE 802.11 provides an access-based share of the medium in cell, an access-based fairness model for the medium share seems more appropriate. Other solutions based on explicit optimization models use such an access-based fairness scheme. In the first model, we maximize the logarithmic utility function. Furthermore, in contrast to most of the proportional fairness solutions based on optimization models (and all considering a time-based fairness medium share), we evaluate our solution with a network simulator.

In the case of WLANs in unsaturated mode, most of the cited approaches that consider an unsaturated network do not taken into account the traffic demand in the associations except in [19]. In this latter, where the fraction of time a STA needs to satisfy its traffic demand is simply calculated as that demand divided by the link capacity which is a strong approximation. The motivation of the second model is to design association algorithms able to adapt to traffic demands. It allows the controller to balance the load according to the real traffic, alleviate congested AP, and offer bandwidth to STAs that need it. It is based on measurements available on most of the Wi-Fi products (e.g., busy time, data rates, error rates, etc.).

2.4 System Model

In this section, we describe the common assumptions used in our different contributions and the notations used for the mathematical formulation of our centralized association solutions.

We consider an IEEE 802.11 infrastructure based wireless network configured to use the DCF mode. We can distinguish three different entities within our centralized management network (see Figure 2.1): a fixed number of APs, the STAs that are assumed to be covered by the APs, and one or several controller(s). We assume that the channel assignment, power

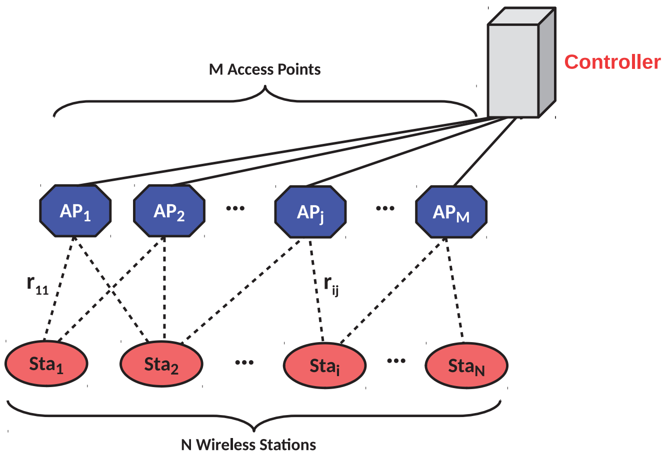


Figure 2.1: Access points and wireless STAs in the network managed by a controller. Dotted lines represent the possible associations between APs and STAs.

setting, and the placement of the APs have already been set and probably optimized during the deployment of the WLAN. The deployment parameters (e.g. channel) can be changed. In this case, these changes can trigger new association decisions at the controller.

The set of APs is assumed to belong to the same extended service set (ESS) and is managed by a WLAN controller. A controller manages the APs, and, in particular, is in charge of determining the associations between STAs and APs. In the formulation of the models, we only take into account downlink traffic, from the APs to the STAs, as downlink traffic is preponderant compared to uplink traffic. The amount of uplink traffic is considered negligible, or at least not significant, with regard to the downlink traffic [40, 41].

When the STAs initially join the network, they do not know the association decisions that are going to be made or even the presence of the controller. Therefore, when the association process starts, STAs connect to the APs with the strongest signal. This is the initial association step, and the association between the STA and AP is not considered completely established. Therefore, there is no data exchange at this step. Association decisions are then made by the controller and announced to the STA through the initial association. Once the decision is received, the STA connects to the AP given by the controller. The controller decision can be sent to the STA via A control frames.

The controller (see Figure 2.2) implements one of our proposed algorithms to compute the association. It may be performed at regular interval or at certain events (e.g. arrival/departure of STAs). The application of a new association induces a re-association cost. The condition for applying a new association may be function of the cost and gain of

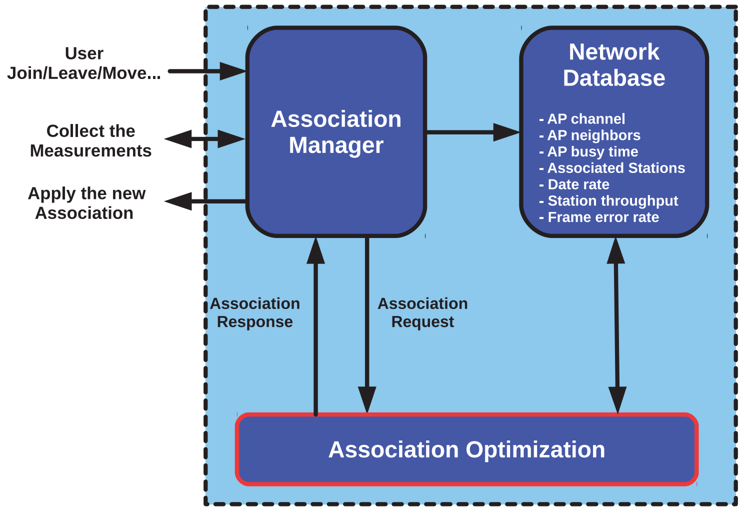


Figure 2.2: Centralized association optimization process in the controller.

the new configuration.

The association optimization is based on measurements made on each AP that allow the controller to know the performance level of the current association, and on our models that predict the performance for other association schemes. The controller can periodically collect from the APs measurements such as:

- the current association,
- for each AP, the APs that are within its sensing range (with which it is in conflict to access the medium),
- the busy time fraction for each AP,
- for each STA:
  - the data rates between APs and the STA,
  - the throughput and the average frame size received by the STA from its AP,
  - the error rate (or equivalently the probability of success) between the STA and its AP.

It is worth noting that most of these measurements are already available on most of the AP products (e.g., Cisco Aironet Series APs).

In our study, data rate refers to the transmission rate at the physical layer. Each IEEE 802.11 standard version has different possible data rates that depend on the used modulation.

In the same way, in our study, link capacity will refer to the maximum rate theoretically possible at the application layer. This link capacity can easily be computed from the data rate taking into account all overheads induced by the networking sublayers.

Table 2.2 summarizes the different notations used throughout this thesis.

Symbol	**Description**
A	Set of APs in the network
M	Number of APs in the network $M = \|A\|$
S	Sat of STAs in the network
N	Number of wireless STAs in the network $N = \|S\|$
S_j	Set of STAs associated with AP j
r_{ij}	Link capacity between AP j and STA i
t_{ij}	Mean transmission time of one frame from AP j to STA i
L_i	Mean frame size to be transmitted to STA i
d_{ij}	Mean throughput obtained by STA i when associated to AP j
D_j	Mean outgoing throughput of AP j
N_{ij}	Mean number of frames transmitted from AP j to STA i
x_{ij}	1 if STA i is associated to AP j, 0 otherwise
s_{ij}	1 if AP i is in sensing range of AP j, 0 otherwise
λ_i	Mean number of datagrams transmitted to STA i in one second
$\overline{T_{ij}}$	Mean time required to transmit one datagram from AP j to STA i
$T_{ij}(k)$	Time required to transmit one frame at backoff stage k from AP j to STA i
$T_{ij}^c(k)$	Time required to consider that the transmission failed at backoff stage k
T_{BA}	Duration of the block acknowledgment frame
$T_{BO}(k)$	Mean back-off after k unsuccessful successive transmission attempts
m	Maximum number of retransmissions
R_{ij}	Link data rate between AP j and STA i
p_{ij}	Probability of success to transmit one frame between AP j and STA i
q_{ij}	Probability of transmission failure of at least one frame in an A-MPDU
τ	Average MPDU size
τ_{max}	Maximum MPDU size

Table 2.2: Notations

Chapter 3

A new Association based Access Fairness

Sommaire

3.1 Introduction

In the previous chapter, we have presented the network model and assumptions that define the general context of the thesis. For this chapter, we first present the scenario for which we propose an association solution. We formulate this first model as well as other complementary assumptions related to this scenario.

In this scenario, we consider a WLAN in saturated mode, i.e., where all APs have always a frame waiting for transmission. This scenario allows us to evaluate the capacity of the WLAN in an extreme case and the performance gain obtained by our association optimization in this particular case.

Our model relies on an access based fairness. We consider that an AP access the medium, in average, the same number of times to serve each STA associated with it. It is thus different of the works of the literature where the AP serves each STA during the same amount of time. Our model is thus more realistic as it corresponds to the current implementation of APs and to the IEEE 802.11 DCF mode. To solve the formulated problem (being NP-hard), we propose a local search heuristic. This heuristic has the benefit to be tuned according to time and CPU constraints of the controller.

The rest of the chapter is organized as follows. In Section 3.2, we present the mathematical model and the optimization problem when orthogonal and non-orthogonal channels are used. Then, in Section 3.3, an heuristic is proposed to solve this problem. A performance evaluation of our solution based on ns-3 simulations is carried out in Section 3.4. Our solution is compared to the time-based fairness approach in Section 3.5.

3.2 Association Model

3.2.1 Assumptions and Scenario

As we have discussed in chapter 2, most existing association optimization models assume a time-based fairness between the APs and the STAs, *i.e.* assume that the time is equally shared between STAs and APs. This assumption requires to apply an appropriate scheduling on each AP that must take into account different parameters like the packet sizes and the physical transmission rates. In practice, APs use very simple scheduling policies like a FIFO scheduling and the DCF (Distributed Coordination Function) mode of IEEE 802.11 that provides an access-based share of the medium between APs/STAs. Therefore, considering an access-based fairness model for the medium share seems more appropriate. Very few solutions, based on explicit optimization models, use such an access-based fairness scheme. This is the case in [8], but the solution aims to minimize the maximal load on all the APs. In our work, we choose to maximize a logarithmic utility function. Its optimization offers a good tradeoff between the overall network throughput and the fairness of user throughputs.

We also assume that the amount of data intended to the STAs associated to the same AP is equal in average, or in a long term period. To this end, we assume that the mean number of frames transmitted to each STA and the mean frame size are the same for each STA. Obviously, it will not correspond to reality, but it allows to express the problem with an equal priority to each STA [24, 42]. This assumption is motivated by different reasons:

- The optimization problem is thus addressed without privileging a STA because it has more traffic than the others at a given time;
- Internet traffic is quite sporadic and the time scale in terms of dynamics is very likely smaller than the one of the association problem, which implies that, in average, STAs may receive the same amount of data;
- The association problem output consists in associating STAs with APs and the goal is not to directly set/reserve any resource for each STA; consequently, STAs that receive more traffic still benefit of the statistical multiplexing offered by the Wi-Fi technology.

The objective function that we optimize is based on the mean throughputs between APs and STAs, denoted d_{ij} ($i \in \{1, ..., N\}$ and $j \in \{1, ..., M\}$ where N is the number of STAs and M the number of APs). By convention, we set $d_{ij} = 0$ if Sta_i is not associated to AP_j. This throughput depends, among others, on the number of STAs associated with the AP, and the corresponding link capacity. The link capacity r_{ij} is defined here as the maximum amount of data that can be exchanged between AP_j and Sta_i in one second. The throughput d_{ij} is the throughput when considering the other STAs and, in one of the proposed models (see Section 3.2.3), the other interfering APs. In other words, d_{ij} takes into account the fact that the medium is shared whereas r_{ij} does not.

We present our optimization problem under two variants. The first approach assumes that the channels used by the APs are orthogonal, meaning that they can not detect each other and can transmit at the same time without interfering. It is equivalent to assume that there are as many orthogonal channels as APs. Then, in the second approach, we consider that the number of orthogonal channels is limited. Consequently, APs which use the same channel and which are in the sensing range of each other share the medium. The formula that characterizes the throughput between an AP and a STA is refined accordingly.

3.2.2 Orthogonal channels

We assume that all APs use different orthogonal channels, or equivalently the APs using the same channel are far enough to avoid any interference and signal detection. Therefore, each AP can be considered as an independent sub-network and the mean aggregate throughput for the whole Wi-Fi network is the sum of the mean AP throughputs. We begin by computing the mean overall throughput offered by an AP from which we derive the mean throughput between this AP and one of its associated STAs.

The mean throughput D_j of AP_j is defined as the downlink throughput sent by this AP to the set of its associated STAs:

$$D_j = \sum_{i=1}^{N} d_{ij}$$

It can also be expressed as the ratio between the mean quantity of data transmitted to all wireless STAs associated to it and the time required for these transmissions:

$$D_j = \frac{\sum\limits_{i=1}^{N} N_{ij} L_i x_{ij}}{\sum\limits_{i=1}^{N} N_{ij} t_{ij} x_{ij}} \tag{3.1}$$

where N_{ij} is the mean number of frames sent from AP_j to Sta_i, L_i is the mean size of these frames, x_{ij} indicates if Sta_i is associated to AP_j (it is equal to 1 if it is true, and 0 otherwise) and t_{ij} is the mean time to send a frame from AP_j to Sta_i. This time is given by the ratio between the mean frame size and the link capacity:

$$t_{ij} = \frac{L_i}{r_{ij}} \tag{3.2}$$

By substituting (3.2) in (3.1), we get:

$$D_j = \frac{\sum\limits_{i=1}^{N} N_{ij} L_i x_{ij}}{\sum\limits_{i=1}^{N} N_{ij} \frac{L_i}{r_{ij}} x_{ij}} \tag{3.3}$$

As we assume that the mean number of frames transmitted to each STA and the mean frame size are identical for each STA, the mean overall throughput of an AP is then given by:

$$D_j = \frac{\sum_{i=1}^{N} x_{ij}}{\sum_{i=1}^{N} \frac{x_{ij}}{r_{ij}}} \tag{3.4}$$

Also, as we assume that the STAs associated to the same AP receive the same amount of data in average, then the throughput of the AP is equally shared among its wireless STAs. Therefore, the mean throughput d_{ij} obtained by Sta_i from its AP_j becomes:

$$d_{ij} = \frac{D_j}{\sum_{k=1}^{N} x_{kj}} \tag{3.5}$$

Substituting D_j in (3.5), we get:

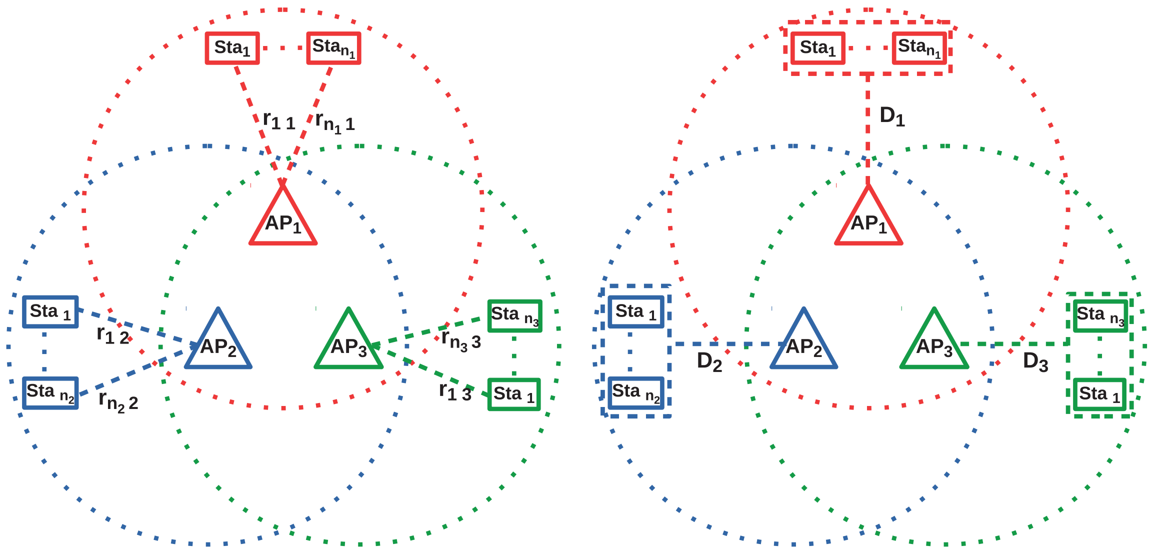
$$d_{ij} = \frac{1}{\sum_{k=1}^{N} \frac{x_{kj}}{r_{kj}}} \tag{3.6}$$

From Equation (3.6), we can easily see that the mean throughput d_{ij} of Sta_i associated to AP_j is the same for all STAs associated to this AP, whereas they may experience different link capacities with this AP.

Our optimization aims to maximize the total downlink throughput for the whole network while ensuring fairness between wireless STAs. In order to introduce fairness in the objective function, we use the logarithmic utility function proposed by Kelly in [43]. The association optimization problem with orthogonal channels can then be formulated as follows:

$$\begin{aligned}
&\max \quad \sum_{i=1}^{N} \log\left(\sum_{j=1}^{M} d_{ij} x_{ij}\right) && \\
&\text{with} \quad d_{ij} = \frac{1}{\sum_{k=1}^{N} \frac{x_{kj}}{r_{kj}}} && 1 \le i \le N, 1 \le j \le M \\
&\text{subject to} \quad \sum_{j=1}^{M} x_{ij} = 1 && 1 \le i \le N, \\
&\qquad x_{ij} \in \{0, 1\} && 1 \le i \le N, 1 \le j \le M, \\
&\qquad \text{if } r_{ij} = 0 \text{ then } x_{ij} = 0 && 1 \le i \le N, 1 \le j \le M.
\end{aligned} \tag{3.7}$$

The objective is thus to find the set of association variables x_{ij} that maximizes the total network throughput while ensuring a certain fairness. The two first constraints are related to the association variables x_{ij} and ensure that a STA is connected to a single AP. The third constraint aims to guarantee that a wireless STA cannot associate with an AP that is not within its receiving range.



(a) Simple representation (b) Aggregated representation

Figure 3.1: Wireless network with non-orthogonal channels

3.2.3 Non-orthogonal channels

We propose to refine the model by considering non-orthogonal channels. A certain number of orthogonal channels are available but their number is limited, so several APs may use the same channel.

In practice, it is difficult to know the various interference that can undergo a radio signal in a wireless network. A source of interference can belong to the same Wi-Fi network, *e.g.* a nearby AP with the same SSID, part of the same Extended Service Set (ESS), or can be external such as another wireless network, or any radio source in the same frequency band. As information on interference and traffic are not easily available and unpredictable for external sources, we consider only interference that exist between APs of the same ESS.

We assume that the assignment of the channels to the APs has been set. Two APs will detect transmissions of each other if they use the same channel and are in the carrier sense range of each other. It leads to a share of the medium, as transmissions can not take place at the same time or collisions may happen if they transmit at the same time. The two APs are then in conflict.

According to the IEEE 802.11 DCF mode, APs in mutual conflict have equivalent opportunities to access the medium [44]. Therefore, we assume that the number of accesses to the medium is equal, in average, between conflicting APs. In Figure 3.1a, we represent three APs in mutual conflict.

To compute the mean throughput D_j^* of AP_j in presence of conflicts, we use Equation (3.6) and adapt it to this context. The throughput of STAs associated to the same AP is seen as an aggregation, as shown in Figure 3.1b. d_{ij} is then replaced by D_j^*. D_j corresponds to the mean throughput of AP_j without conflict, therefore r_{kj} (which is the bandwidth obtained by one STA without conflict) is replaced by D_j. Finally, in this context, the share comes from the APs in conflict: x_{kj} is replaced by s_{kj} that represents the number of APs in conflict with AP_j. Note that this adaptation of Equation (3.6) is possible because the opportunity to access the channel is the same for all APs in mutual conflicts (as the throughput of an AP is equally shared among its associated STAs in the previous model):

$$D_j^* = \frac{1}{\sum\limits_{k=1}^{M} \frac{s_{kj}}{D_k}} \tag{3.8}$$

Substituting (3.4) in (3.8), we get :

$$D_j^* = \frac{1}{\sum\limits_{k=1}^{M} \left(\frac{s_{kj}}{\sum\limits_{i=1}^{N} x_{ik}} \sum\limits_{i=1}^{N} \frac{x_{ik}}{r_{ik}} \right)} \tag{3.9}$$

As for the case with orthogonal channels, we assume that the AP throughput is equally shared among the STAs associated with it. Therefore, the mean throughput for a particular STA is:

$$d_{ij}^* = \frac{1}{\sum\limits_{i'=1}^{N} x_{i'j}} \cdot \frac{1}{\sum\limits_{k=1}^{M} \left(\frac{s_{kj}}{\sum\limits_{i'=1}^{N} x_{i'k}} \cdot \sum\limits_{i'=1}^{N} \frac{x_{i'k}}{r_{i'k}} \right)} \tag{3.10}$$

The formulation of the association optimization problem in a wireless network with non-orthogonal channels is given as follows:

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{N} \log\left(\sum_{j=1}^{M} d_{ij} x_{ij}\right) \\
\text{with} \quad & d_{ij} = \frac{1}{\sum\limits_{i'=1}^{N} x_{i'j}} \cdot \frac{1}{\sum\limits_{k=1}^{M}\left(\frac{s_{kj}}{\sum\limits_{i'=1}^{N} x_{i'k}} \cdot \sum\limits_{i'=1}^{N} \frac{x_{i'k}}{r_{i'k}}\right)} \\
\text{subject to} \quad & \sum_{j=1}^{M} x_{ij} = 1 \qquad 1 \le i \le N, \\
& x_{ij} \in \{0,1\} \qquad 1 \le i \le N, 1 \le j \le M, \\
& \text{if } r_{ij} = 0 \text{ then } x_{ij} = 0 \quad 1 \le i \le N, 1 \le j \le M.
\end{aligned}
\tag{3.11}
$$

The objective here is to maximize the overall network throughput while ensuring a certain fairness between the wireless STAs when the APs use non-orthogonal channels. The expression of the mean throughput between an AP and an associated STA has changed, compared to the orthogonal channel case, to take into account conflicts between APs. The constraints are the same as in the orthogonal channel case.

3.3 Optimization Problem Solving

The optimization association problem is formulated, in the previous section, as a centralized optimization approach based on the use of a logarithmic utility function. The problem is modeled as a non-linear programming problem with binary decision variables representing the association of wireless STAs to APs, which is known to be NP-Hard [45].

Most of the studies that deal with optimization of Wi-Fi associations, and that have been presented in Section 2 [20, 18, 21, 19], use approximation algorithms based on relaxation to a non-linear convex program. It allows them to apply the rounding process proposed by Shmoys and Tardos for the generalized assignment problem [46], to provide binary values of the association variable x_{ij}. This often does not allow an exact solution of the problem in a reasonable computational time.

Instead, to solve our optimization problem, we propose an iterative heuristic based on the principle of local search, also called descent or iterative improvement. Local search is an important class of heuristics used to solve combinatorial optimization problems. The key idea of a local search algorithm is to start from an initial feasible solution and iteratively find, at each iteration, a solution called a best neighbor that improves the objective function [47]. Among the existing procedures to choose a better neighbor, two approaches are widely used. The first one is an exhaustive exploration of the neighborhood to choose the best neighbor. It is this procedure that we consider in this work. The second approach consists in exploring partially the neighborhood and choosing the first neighbor that obtains a better objective function. The search procedure stops when the current solution is better than all its neighbors, this solution is then a local optimum.

The main benefits of local search lie in its effectiveness (with some type of problems), its simplicity for actual implementation and its iterative process which can stop the optimization process at any time to comply with a constraint like the computation time for instance. This is supported by the fact that the local search algorithms consider only complete feasible solutions during the search. The proposed algorithm has then the advantage to improve Wi-Fi associations at each iteration, and can be stopped at any time with a feasible solution. The time that the system spends in computing a solution can thus be bounded and tuned.

Our iterative local search method is based on two essential elements: a neighborhood structure and a procedure exploiting this neighborhood. The method can be summarized as follows:

1. It starts with an initial feasible solution.
2. At each iteration, it chooses, among all the neighbors of the current solution, one of the solutions that maximizes the objective function. This neighbor becomes the current solution on which to apply the next iteration.

We present in the two next paragraphs, the notion of neighborhood, and a more detailed version of the local search algorithm.

3.3.1 Neighborhood structure

Since the search space associated with a combinatorial optimization problem is often non-enumerable in a reasonable time, the local search defines a neighborhood relationship that connects certain solutions to each other. At each step, the algorithm searches for a better solution in the neighborhood of the current one.

Therefore, for our optimization problem we define a neighborhood structure that consists of changing the association of a single STA in the current solution.

Let $X(X = (x_{ij})_{1\leq i\leq N, 1\leq j\leq M})$ be a feasible solution of the optimization problem. X is a $N \times M$ matrix (we remind that N is the number of STAs and M the number of APs). Its terms $x_{i,j}$ correspond to the association variables of the optimization problem.

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,j'} & \cdots & x_{1,M} \\ x_{2,1} & \cdots & x_{2,j} & \cdots & x_{2,j'} & \cdots & x_{2,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,j'} & \cdots & x_{N,M} \end{pmatrix}$$

X being a feasible solution, constraints (3.7) and (3.11) of the optimization problem hold: all variables x_{ij} in a line are equal to 0 except one equal to 1, and $x_{ij} = 0$ if $r_{ij} = 0$.

The neighborhood $\mathcal{N}(X)$ of a feasible solution X is the set $\mathcal{N}(X) = \{X^1, X^2, ..., X^e, ..., X^{L(X)}\}$. It is composed of all the feasible solutions where only a single STA has changed of AP compared to the solution X. Basically, a neighbor X^e is then equal to X except that two elements have been permuted in a same line. These two elements are chosen in such a way that they respect the constraints. Formally, let $X = (x_{ij})_{i,j}$ be a feasible solution. A $n \times m$ matrix $X^e = (x^e_{ij})_{i,j}$ belongs to $\mathcal{N}(X)$ if and only if $\exists (l,p) \in \{1,..,M\}^2$, $l \neq p$, and $k \in \{1,..,N\}$ such that:

- $x_{kl} = x^e_{kp} = 1$,
- $x_{kp} = x^e_{kl} = 0$,
- $r_{kp} > 0$,
- $x_{ij} = x^e_{ij}$ $\forall (i,j) \neq (k,p)$, $\forall (i,j) \neq (k,l)$.

For instance, a neighbor of the matrix X (given above) can be:

$$X^e = \begin{pmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,j'} & \cdots & x_{1,M} \\ x_{2,1} & \cdots & x_{2,j} & \cdots & x_{2,j'} & \cdots & x_{2,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,j'} & \cdots & x_{N,M} \end{pmatrix}$$

The neighborhood of a solution X contains at most $(N \times M) - 1$ elements $(L(X) \leq N \times M - 1$ for all $X)$. The cardinality $L(.)$ is less than $(N \times M) - 1$ as only feasible solutions are considered: we can move a STA k from AP_l to AP_p only if $r_{kp} > 0$.

3.3.2 Local search algorithm

The iterative local search algorithm is given in Algorithm 1. It starts from an initial feasible solution X_0. At each iteration, it computes the objective function $\mathcal{F}()$ (defined in 3.7 and 3.11) for all the neighbors of the current solution. After an iteration, there is at most one STA that changes of AP. It corresponds to the best neighbor of the current solution, i.e. the one that maximizes the objective function.

Algorithm 1 Local Search

//Initialization
Collect measurements for current solution X_0
Infer the APs conflict graph for each channel
//The optimization loop
while (Convergence() = false) **do**
    $\mathcal{V} = \mathcal{N}(X_0)$;
    $X = \arg\max_{\mathcal{U} \in \mathcal{V}} \mathcal{F}(\mathcal{U})$;
    if $(\mathcal{F}(X) > \mathcal{F}(X_0))$ **then**
        $X_0 = X$;
    end if
end while
end procedure

The condition to exit the optimization loop is implemented by the function $Convergence()$ which returns a boolean. This function is not fixed and depends on the context. For instance, $Convergence()$ may return $TRUE$ when it has:

1. found a local maximum (as the solution space is finite, the local search reaches a local optima in a finite number of iterations when the current solution has no neighbor with a greater objective function),
2. reached a maximum fixed threshold for the number of iterations,
3. or exceeded a maximum fixed threshold for the runtime of the optimization program.

The two last conditions ensure that a feasible and better solution (compared to the initial one) may be found while respecting the time constraint of the system.

The result of the local search algorithm greatly depends on the initial solution (i.e. the starting point of the search algorithm X^0). Some of them give a globally optimal solution while others give only locally optimal solutions. Also, the number of iterations to reach the output solution depends on the initial configuration. A way to improve the performance of the algorithm consists in running several instances with different starting points. In this case, Algorithm 1 is repeated several times, with, each time, a different starting solution. This aspect is further discussed in Section 3.3.3.

3.3.3 Heuristic evaluation

In order to assess the effectiveness of the proposed heuristic for solving the optimization problem, we perform tests over 100 different network configurations. We consider 4 APs deploy as a grid. We assign to each AP an orthogonal channel. 20 STAs are then deployed around these APs. An example of such a topology is shown in Figure 3.2 (but with 25 APs and 100 STAs in this example). The configurations are obtained by randomly changing STAs and APs location. In this part, we use the network simulator ns-3 [48] to generate topologies and extract the initial configurations which are based on RSSI values, whereas the main features of our heuristic are evaluated with the C++ code implementing this heuristic. The capacities r_{ij} are assumed to be known (we explain in the next section how to get them).

Conf	**RSSI**	**Brute Force**		**Local Search (RSSI)**			**Local Search (multi-start)**		
	Obj. F	**Obj. F**	**Time(s)**	**Obj. F**	**Iter**	**Time(ms)**	**Obj. F**	**Iter**	**Time(ms)**
1	17.661	21.474	225.16	21.474	6	0.233	21.474	359.00	9.943
2	18.604	21.038	57.24	21.038	6	0.185	21.038	355.00	9.328
3	18.320	19.487	1800.56	19.487	4	0.149	19.487	452.00	14.394
4	18.451	20.896	12806.50	20.896	5	0.202	20.896	406.00	13.286
5	19.018	21.226	399.98	21.226	5	0.154	21.226	386.00	10.850
6	19.275	21.364	712.93	21.364	6	0.196	21.364	408.00	12.164
7	20.034	23.442	2381.80	23.442	6	0.207	23.442	400.00	12.429
8	17.807	21.417	2436.58	21.417	11	0.338	21.417	426.00	13.447
9	20.367	22.403	15090.80	22.392	4	0.165	22.403	418.00	14.118
10	18.896	20.832	46.86	20.832	5	0.152	20.832	368.00	9.556
...	...	...	...	...	...	...	...	...	...
100	19.154	22.016	3028.12	22.016	4	0.142	22.016	435.00	13.509
Mean	**19.008**	**21.402**	**2729.568**	**21.396**	**5.57**	**0.183**	**21.402**	**391.06**	**11.724**

Table 3.1: Comparison results between the Local Search Algorithm and Brute Force Algorithm

For each network configuration, we consider different initial associations (*i.e.* initial feasible solutions). The first one is based on the RSSI (each STA is associated with the AP that gives the best signal strength). We also test the local search algorithm in multi-start mode with 30 different initial associations. For each network configuration, we keep, among the 30 initial instances randomly chosen among the feasible solutions, the one that gives, after optimization, the best result with regard to the objective function. We give, in Table 3.1, the values of the objective function before optimization ("RSSI" column) and after optimization with the initial association based on RSSI ("Local search (RSSI)" column) and based on 30 initial associations ("Local search (multi-start mode)" column). Beside, we also consider the optimum solution: for a given network configuration, we evaluate the objective function for all possible associations and keep the best one ("Brute force" column).

The first observation is that the default association based on RSSI is not optimal. The objective function at the optimum is increased by 10-20% compared to the RSSI based association. With our heuristic in multi-start mode, the local search algorithm is able to find the optimum in all cases, i.e. for the 100 network configurations. When our heuristic starts from the RSSI association, the optimum is found for 87 configurations over 100. For the other 13 configurations, the maximum difference is less than 1%.

Also, we have evaluated, for each algorithm, the number of iterations and computation times to converge towards its solution. Computation times have been measured on a laptop (RAM $8GB$, CPU *Core* $i7$ "$4 \times 1.8GHz$", OS $Linux - Ubuntu$). The number of iterations and computation times of the brute force algorithm vary according to the network configurations. It is due to the constraints of our optimization problems, more precisely to the number of APs in the communication range of each STA, which leads to a different number of feasible solutions. The optimum is obtained after 8.49 *seconds* in the best case and after approximately 6 hours (21368 *seconds*) in the worst case. It is approximately 15×10^6 times greater than for the local search algorithm. It empirically proves that an exhaustive search is not an accurate approach in terms of complexity even for such simple configurations. The multi-start mode of our algorithm requires between 300 and 456 iterations to find its solution (we sum the number of iterations over the 30 instances), corresponding to 7.535 *ms* and 15.294 *ms* respectively. The local search algorithm based on RSSI needs only 11 iterations in the worst case, with 5.57 iterations in average. The time to obtain the optimum or local optimum takes 0.184 *ms* in average, with a maximum of 0.338 *ms*. This variant of our algorithm is clearly more efficient than the other ones. The number of iterations is at least two times less than with the multi-start mode (13.03 iterations for an instance of the multi-start mode in average versus 5.57 for the one based on RSSI).

It clearly appears that the heuristic starting from the default RSSI association offers an interesting tradeoff between performance and complexity. It leads to an efficient solution, close to the optimum, and requires only one initial configuration. This initial configuration seems relevant as it converges faster to the solution compared to a random initial configuration. Consequently, it is this variant of our algorithm that will be used in the performance evaluation part presented in the next section.

3.4 Evaluation

We now use the network simulator ns-3 to evaluate the performance of our heuristic with a more realistic and richer environment, as all network aspects from the physical to the application layers are simulated.

Simulations are performed as follows. The first step consists in using ns-3 to create the network topologies, to compute the link capacities between the APs and the STAs (r_{ij}) and to extract the initial association based on the RSSI values. A link capacity between one AP and one STA corresponds to the throughput received by the STA when a saturated constant bit rate (CBR) flow is generated (using an UDP application) between the two considered nodes and when all other STAs and interferences from the other AP/STAs are neglected. Note that these capacities are computed at the application layer of the TCP/IP stack. This has the advantage of:

- taking into account the headers generated by the sub-layers and the overhead induced by the IEEE 802.11 DCF mode (e.g. the MAC header and the Acknowledgment frame),
- directly obtaining the useful throughput,
- designing the proposed model independently of the standard (802.11 a/b/g/n, ...).

In a second step, ns-3 generates CBR traffic between APs and STAs for the RSSI association. This traffic is homogeneous between STAs and saturates the medium. The generated payloads have a size of 1500 bytes. We then measure the obtained throughput (d_{ij}) for each STA.

The last step consists in running our heuristic to find an optimized association. This step has been integrated to ns-3. Once our heuristic has found the solution, we force the STAs to associate to the corresponding APs. We then generate again the same CBR traffic between APs and STAs and measure the obtained throughput (d_{ij}) for each STA. Other types of traffic are also tested. They are described when they are specifically tested.

As the objective function aims to offer a trade-off between throughput and fairness, the performance metrics we consider are the overall throughput (sum of the throughput for all STAs) and the Jain's index. These two metrics are computed from the results obtained with ns-3 simulations.

The Jain's Index [49] evaluates the fairness achieved in the network. It is defined as follows:

$$Jain = \frac{\left(\sum_{i=1}^{n} d_{iAP(i)}\right)^2}{n \sum_{i=1}^{n} d_{iAP(i)}^2}$$

where STA i is associated with AP(i).

Wireless interfaces are configured to use the IEEE 802.11n standard. We shall simulate it on the two frequency bands: 2.4 GHz and 5 GHz. The transmission power is 40 mW (16.00206 dBm). We use the rate adaptation algorithm $IdealHtWifiManager$ of ns-3 to set the physical rate between STAs and APs. We had to develop it as it was not available for 802.11n. The code may be found in [50]. This manager determines the best physical transmission rate to use between a STA and its AP according to the SNR measured on packets sent from the source to the destination.

The Wi-Fi network consists of 25 (5x5) APs, deployed on a square grid such that the distance between two adjacent APs is 100 meters. APs are then randomly moved within a circle with a diameter of 25 meters (the center being the grid points) to obtain more realistic topologies. This distance leads to overlapping zones. A STA may then have several choices for its association. STAs are randomly distributed in the coverage area of the APs. The distribution is Gaussian, centered in the middle of the grid. A topology sample is shown in Figure 3.2.

For each scenario, the number of APs is fixed (25), and we increase the number of wireless STAs from 25 to 250. For each scenario, we perform 30 different configurations for a given number of STAs. These configurations are obtained by randomly changing the STA and AP positions. In the different figures, each point is the mean of these 30 simulations with a confidence interval at 95%.

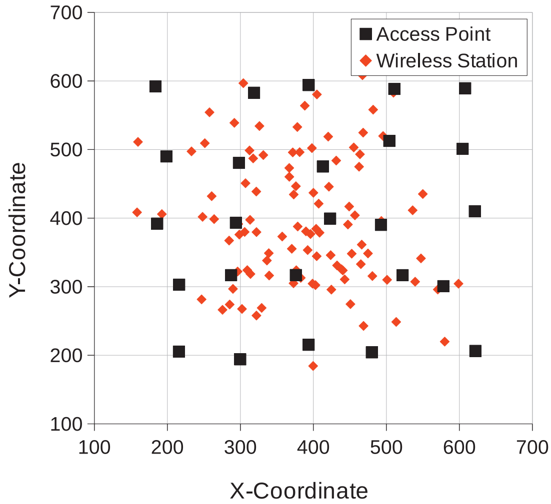
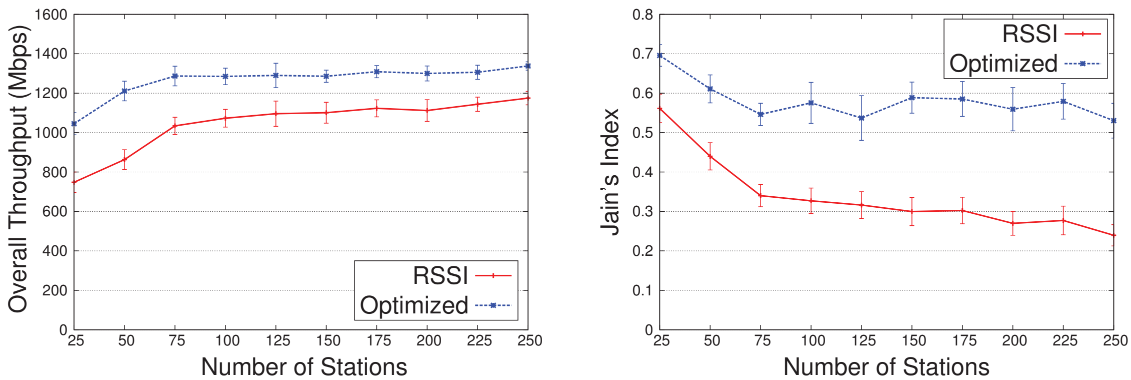


Figure 3.2: Placement of APs and wireless STAs for one simulation.

3.4.1 Orthogonal channels

Figure 3.3 illustrates the performance results when all APs have orthogonal channels in the 2.4 GHz band. This scenario enables to show the solution performance when there is no radio conflict. Figure 3.3a represents the overall network throughput when associations are based on RSSI values and our heuristic (based on the first model (Eq. 3.7)).

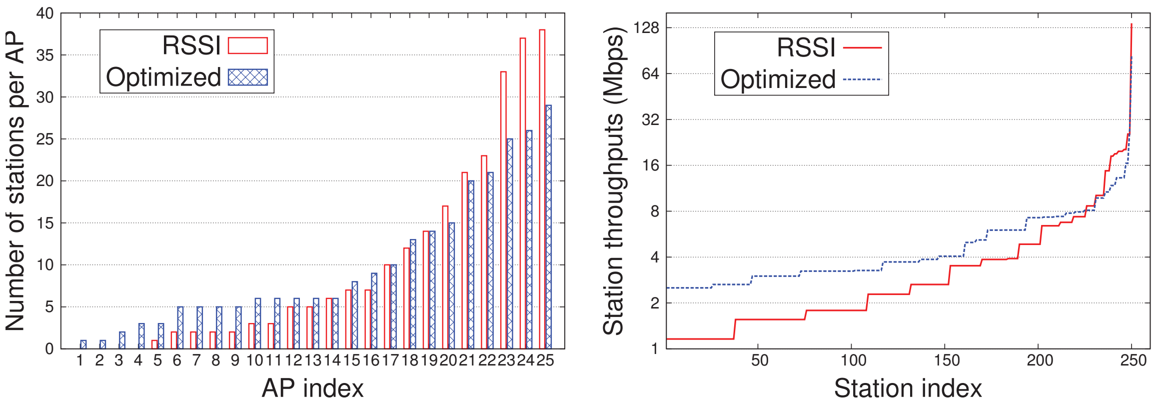


(a) Overall Network Throughput.

(b) Jain's Fairness Index.

Figure 3.3: All Orthogonal channels.

We observe that our algorithm improves the overall throughput by about 40% for a low number of STAs, and by 20% when the number of STAs reaches 250. Also, we can see that the overall throughput of the network increases until 75 STAs (3 STAs per AP in average) and remains stable for a greater number of STAs. Figure 3.3b shows the evolution of the Jain's Fairness index before and after optimization. The optimization significantly improves the fairness, up to 120% for 250 STAs. Moreover, we observe that, with the RSSI association,



(a) Number of STAs per AP.

(b) Throughput per STA.

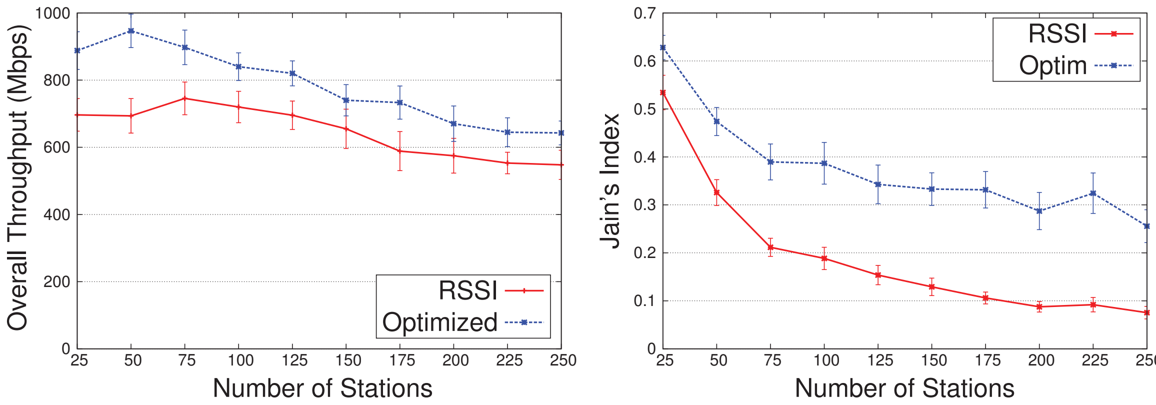
Figure 3.4: Comparison of the number of STAs per AP and STA throughput for one simulation sample before (RSSI) and after optimization (Optimized).

the fairness decreases with the number of STAs whereas it seems to remain stable with our algorithm (at least with 75 STAs and more).

Fairness is also illustrated in Figure 3.4, where for one simulation (250 STAs), we plot the distribution of the number of STAs associated to each AP, and the STA throughput (d_{ij}) before and after optimization. This simulation is representative: the observed trends are similar for all simulations. In Figure 3.4a, we can observe that 4 APs do not have any STAs associated with them with the RSSI association, whereas there is only one AP without STA after optimization. With our solution, it appears that STAs are more homogeneously distributed between APs compared to the RSSI case. A more homogeneous distribution of STAs among the APs leads to more balanced throughput among STAs (Figure 3.4b). In this figure, the x-axis represents the indexes of the 250 STAs in an increasing order of the STA throughput. The y-axis represents the STA throughput (with a log scale). It varies from 1.15 $Mbps$ to 136 $Mbps$ for the RSSI association, and from 2.51 $Mbps$ to 83 $Mbps$ after optimization. It clearly shows a better usage of Wi-Fi resources: STAs use more APs and they are more homogeneously shared between APs leading to a better fairness and a throughput increase.

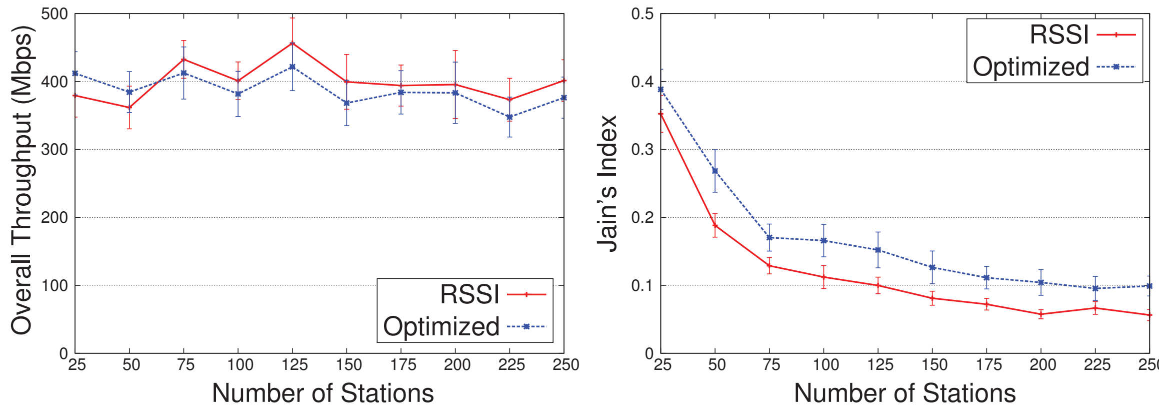
3.4.2 Non-orthogonal channels

We simulate two cases: one in the 5 GHz band with 8 orthogonal channels and one in the 2.4 GHz band with 3 orthogonal channels. We distribute channels on APs in a way that minimizes the number of conflicts and interference. It corresponds to a scenario where the AP deployment has been planned. Figure 3.5a (8 orthogonal channels) shows that our



(a) Overall Network Throughput. (b) Jain's Fairness Index.

Figure 3.5: 8 Orthogonal channels.



(a) Overall Network Throughput. (b) Jain's Fairness Index.

Figure 3.6: 3 Orthogonal channels.

optimization (based on the second model (Eq. 3.11)) improves the overall throughput up to 20% regardless of the number of STAs. The Jain's Fairness index, shown in Figure 3.5b, is improved by our optimization by a factor varying from 1.18 for 25 STAs to 3.39 for 250 STAs.

Figure 3.6 illustrates the simulation results obtained in the 2.4 GHz band with 3 orthogonal channels. Improvements are clearly less significant than in the 8 channels case. Figure 3.6a shows that the optimization does not increase the overall throughput and the obtained results are almost equivalent with these two solutions, with a small advantage to the RSSI-based association. Nevertheless, we can observe an improvement of the Jain's index, with our optimization, varying from 10% to 100% (35% in average).

In our simulations, the sensing range is approximately 221 meters. With our channel allocation and our topologies, an AP detects transmissions from at most 3 APs. As we have seen during the formulation of the problem, APs/STAs that share the medium tends to obtain the same throughput. Consequently, in this very constrained scenario, performance cannot be significantly improved. Throughput can hardly be increased since the high number of STAs on each channel does not allow to separate STAs with high and low link capacities, and fairness is already imposed by STAs with low link capacities.

Variable packet sizes To assess the impact of the packet size, we simulate the same scenario with packets size that varies according to a distribution designed from a real trace [51]. The packet size distribution of this trace is shown in Figure 3.7. The packets have then different size but the average packet size is the same as in the previous simulations (Average Packet Size = 755.572 bytes and Standard Deviation = 674.05). The simulation results for the case of 8 orthogonal channels are illustrated in Figure 3.8. In Figure 3.8a, we show that the overall throughput is improved in average of about 20% with a peak of 36% at 50 STAs with our optimization. Jain's index is plotted in Figure 3.8b where we observe an improvement of 90% in average. We also used this trace to simulate flows with different average packet sizes. The mean, standard deviation, minimum value and the maximum value of the packet sizes are summarized in Table 3.2. The simulation results are illustrated in Figure 3.9. Figure 3.9a shows that the overall throughput is improved of 22% in average after the association optimization. Figure 3.9b shows that fairness is improved of 88% in average. With these results, we see that our solution, relying on a model assuming equal mean packet size, is still efficient when this assumption does not hold anymore.

# of STAs	25	250
Mean	887.42	864.75
StDev	149.05	155.17
Min	594.96	396.14
Max	1088.12	1186.38

Table 3.2: Size properties of the transmitted packets in case of variable packet sizes

3.4.3 On-Line Optimization

Our optimization can be used, in practice, in an on-line way. More precisely, it may be run at regular interval or trigger when an event occurs, to take into account STAs that have left or joined the Wi-Fi network or that have moved. To illustrate the dynamic behavior of our approach, we simulated on the same network topology another scenario in which we randomly remove 50 STAs among the 250 and we replace it with 50 new ones at each interval. The new locations follow the same distribution as the initial one. The new STAs first associate to APs according to the RSSI value, and then our optimization is applied. We repeat this

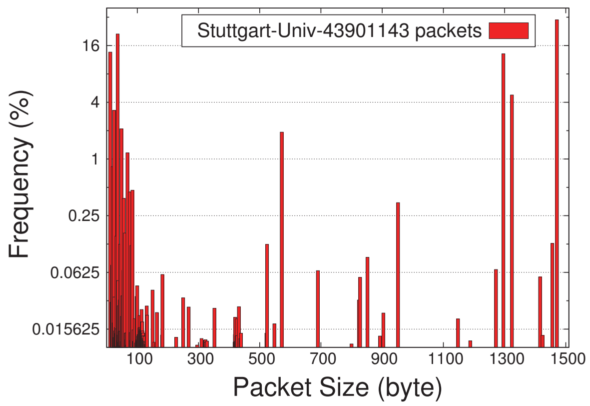
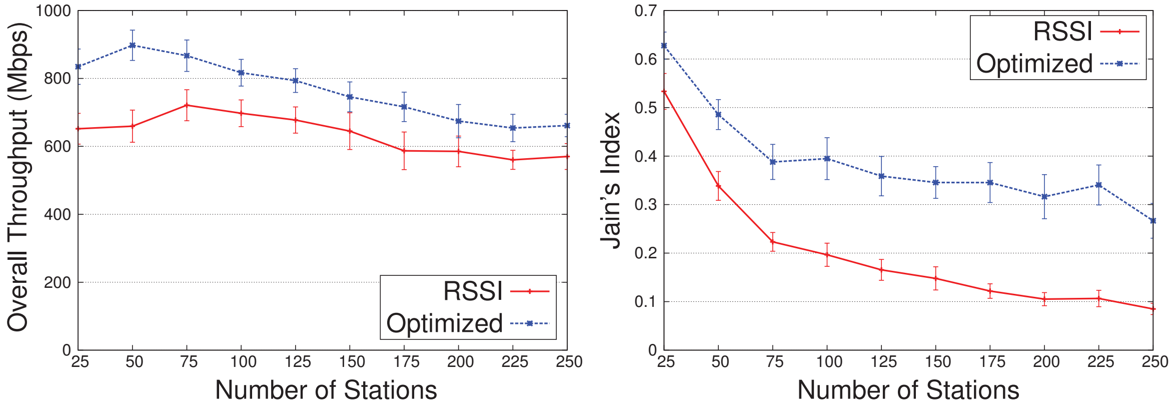
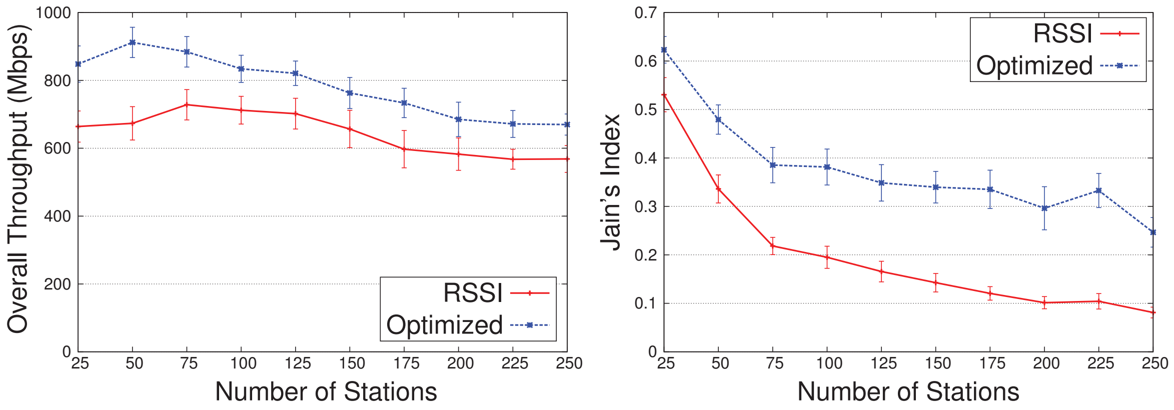


Figure 3.7: Packet size histogram from a trace of the University of Stuttgart.



(a) Overall Network Throughput. (b) Jain's Fairness Index.

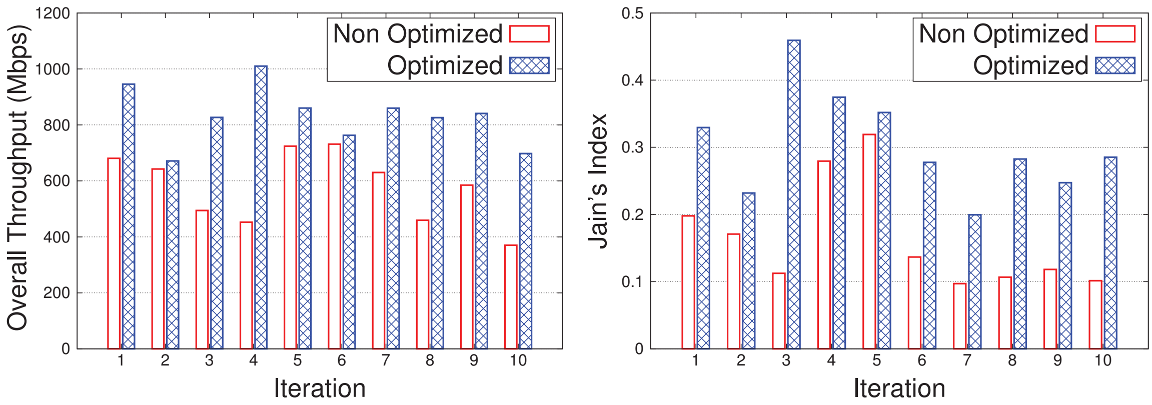
Figure 3.8: Packets with different sizes but with an identical mean for each STA (8 orthogonal channels)



(a) Overall Network Throughput. (b) Jain's Fairness Index.

Figure 3.9: Packets with a different average size for each STA (8 orthogonal channels)

scenario 10 times. The results with 8 orthogonal channels are shown in Figure 3.10. The “Non optimized” evaluation corresponds to a configuration where the association of the 200 STAs that remain come from the previous optimization and the 50 new STAs are associated in function of the RSSI. Figure 3.10a shows the improvement of the overall network throughput



(a) Overall Throughput.   (b) Jain's Fairness Index.

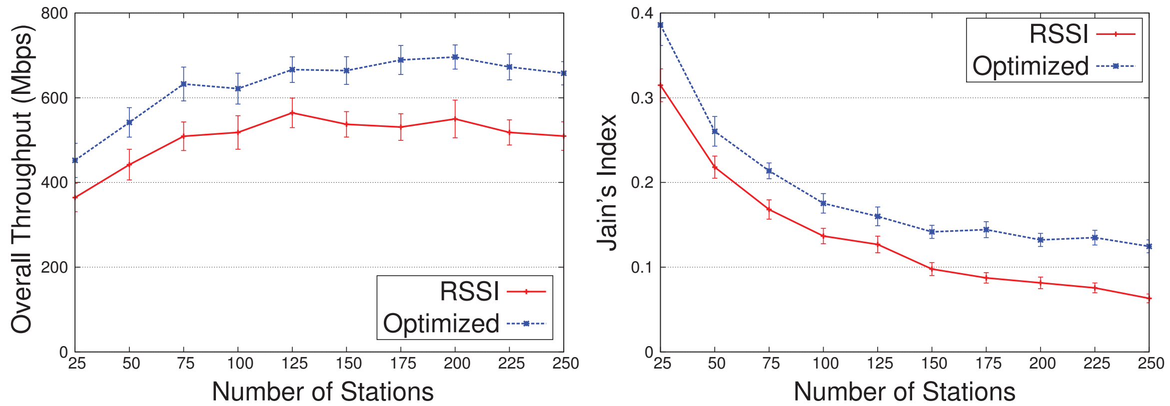
Figure 3.10: Overall throughput and Jain's index for dynamic association optimization

after the optimization at each iteration. This improvement varies between 5% and 120%. Figure 3.10b shows an average fairness improvement of 110% by optimization.

From these results, we note that even if only 20% of the STAs change of position in an already optimized configuration, our solution still allows a significant improvement of the overall network performance.

3.4.4 TCP traffic

We also tested our solution with TCP traffic. The scenario is the same as the one evaluated with CBR traffic, except that traffic is replaced by TCP flows. The use of TCP allows us to test the robustness of our model with respect to TCP when STAs are in competition for accessing the medium for the TCP acknowledgment transmissions. In Figure 3.11, we can observe that the results are qualitatively the same: there is an increase of 20-30% of throughput, and 20-90% for fairness. It empirically shows that the different TCP mechanisms (congestion/flow control) and the upload traffic generated by the TCP acknowledgments do not affect the accuracy of our model.



(a) Overall Network Throughput. (b) Jain's Fairness Index.

Figure 3.11: 8 Orthogonal Channels with TCP flows

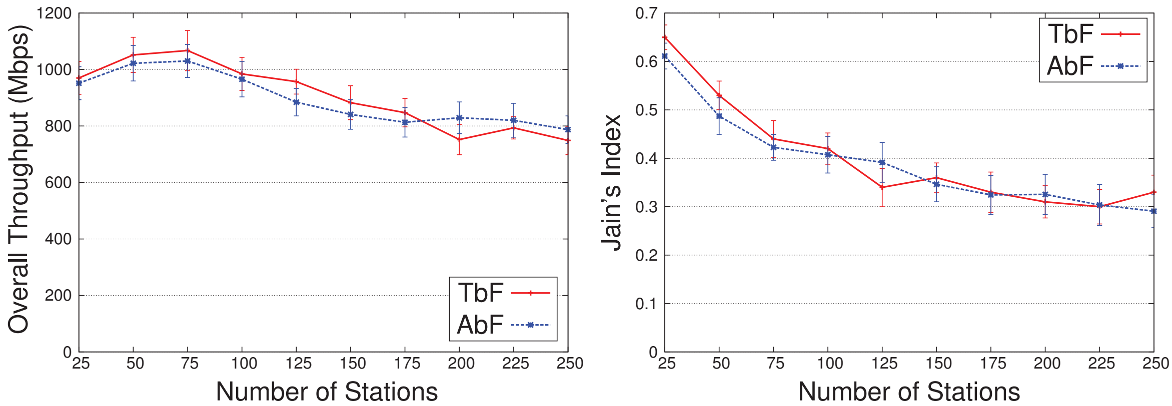
3.5 Comparison with the time-based fairness approach and discussion

In this section, we compare the results, obtained with ns-3, given by our optimization model which assumes an access-based fairness (named AbF hereafter) with the model proposed in [20, 21, 18, 19] that is based on a time-based fairness (named TbF hereafter). We remind that in the TbF model, the medium is equally shared in time between the STAs associated to a same AP while the medium is equally shared in access between APs in conflict. We made the required modification in ns-3 to integrate this model.

Figures 3.12 and 3.13 present the results obtained by the two models in case of 8 orthogonal and 3 orthogonal channels respectively. We observe almost the same overall throughput for 8 orthogonal channels (Figure 3.12a). For 3 orthogonal channels, results are equivalent until 75 STAs but the difference can reach up to 30% for 250 STAs in favor of the AbF algorithm (Figure 3.13a). Jain's index is almost equivalent whatever the number of orthogonal channels.

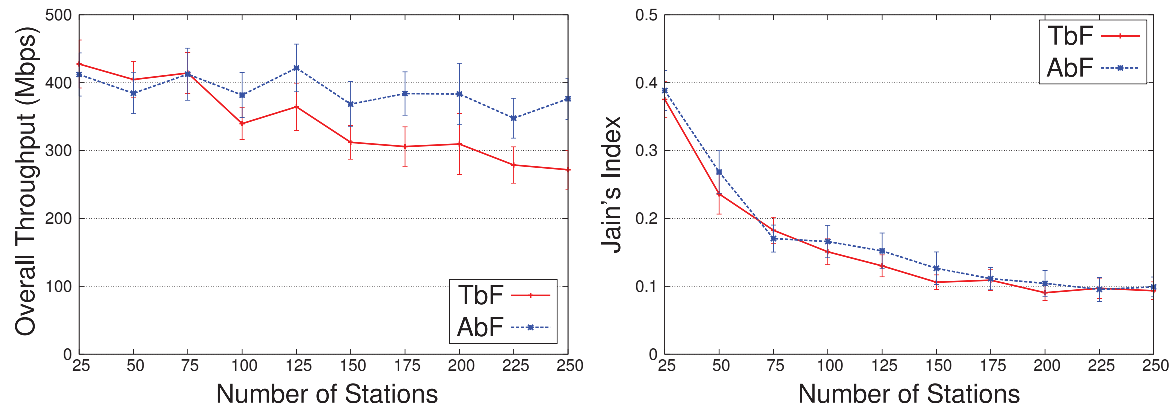
These results are surprising, because, intuitively, the TbF model should lead to a better throughput in all cases since, with the TbF model, STAs with high physical rates are not penalized by the presence of low physical rate STAs (unlike the AbF model). That may be explained by the increase of the number of APs in conflict when the number of orthogonal channels decreases. Indeed, with 3 orthogonal channels, each AP is very likely in conflict with at least one another AP and the gain expected with the use of TbF model is reduced.

We simulate a network of 10 APs and 100 STAs with AbF and TbF schedulers to under-



(a) Overall Network Throughput. (b) Jain's Fairness Index.

Figure 3.12: 8 Orthogonal Channels

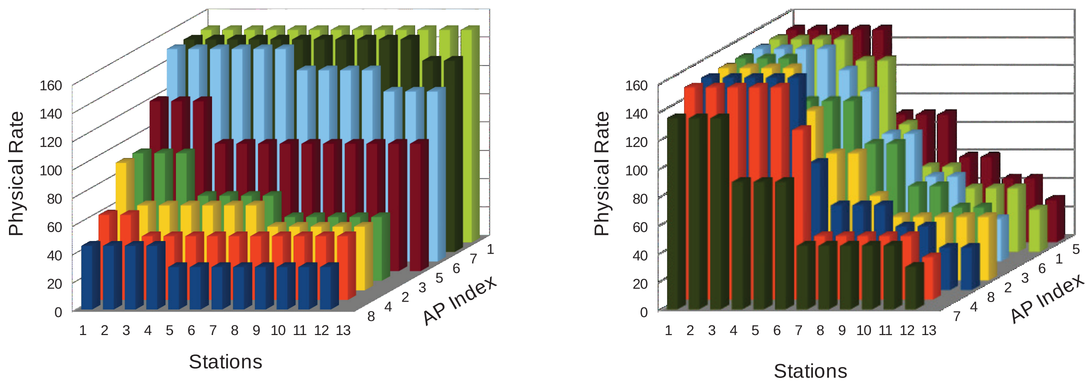


(a) Overall Network Throughput. (b) Jain's Fairness Index.

Figure 3.13: 3 Orthogonal Channels

stand how our optimization algorithm allocates the STAs according to their physical rate. So as to take into account the physical rate of STAs regardless of their positions relative to the APs, we install all the APs at the same location and the STAs are uniformly distributed around. Simulation results plotted in Figure 3.14 show the distribution of STAs on the APs ordered by their physical rates.

For the case of AbF (Figure 3.14a), it appears that our algorithm tends to associate STAs with the same physical rates to the same AP. STAs with high physical rates are consequently not penalized by low rates STAs and the global throughput is increased. Also, to ensure fairness, it fairly distributes STAs between the 8 orthogonal channels. Instead, in the case of TbF (Figure 3.14b), we notice that the STAs are distributed homogeneously between APs so they have the same number of STAs with high and low physical rates.



(a) AbF. (b) TbF.

Figure 3.14: Per-AP STA physical rate for AbF and TbF (8 orthogonal channels).

3.6 Conclusion

In this chapter, we address the association problem in saturated Wi-Fi networks. Our solution, based on an optimization model, aims to improve the overall network throughput while achieving a better fairness between STAs, compared to the classical association based on RSSI.

Our proposed optimization solution is based on a mathematical formulation of the problem and a local search algorithm. The benefit of this algorithm is a convergence in a few iterations when the starting point is the default RSSI association. Moreover, the algorithm can be stopped at any time and always gives a feasible and better association. It can be easily tuned according to the CPU or time constraints of the WLAN controller. Simulation results show that the proposed optimization significantly improves the network performance. In case of orthogonal channels, our optimization increases the overall throughput up to 40% and the fairness up to 120%, and for non-orthogonal channels we observe an improvement varying from 15% to 40% for the overall throughput and from 25% to 300% for the fairness. This improvement is due to a better distribution of STAs among the APs, and an improvement throughput for most of the STAs.

Our solution has also been validated when TCP traffic are transmitted and for different types of traffic (packet sizes and saturated). Moreover, we have shown that the use of an access-based fairness scheme, in addition to be more realistic, leads to better results when the number of orthogonal channels is limited compared to the use of a time-based fairness scheme.

In the next chapter, we formulate a new association optimization model that includes the users' traffic demand. Rather than providing fairness to STAs independently of their current load, it would lead to association that allocate resources to STAs that need them for a given period.

Chapter 4

An Association based on the Channel Busy Time Estimation

Sommaire

4.1 Introduction

In the previous chapter, we have proposed a solution to optimize the association of STAs to APs in WLAN in order to improve the overall network throughput and/or the fairness between STAs in a saturated network scenario. This corresponds to a case where devices (STAs and/or APs) have always a frame to send. Such an assumption is unrealistic but allows to express the minimum amount of throughput a device can obtain. Nevertheless, it does not take into account real traffic demands and the proposed association may be inaccurate. For example, if a part of the STAs have very low traffic, the unused bandwidth can be reused by other STAs with higher demands. A fair distribution of the resources may then be counter-productive: low traffic STAs may be associated to the same AP which may be consequently idle whereas STAs requiring high throughput are associated to overloaded APs.

In this chapter, we propose an association optimization based on traffic demands. These traffic demands are defined for the downlink traffic from APs to STAs and are measured in real time. The load of an AP is estimated through the channel busy time fraction (BTF). The BTF for an AP corresponds to the fraction of time during which it senses the channel (medium) busy due to its own transmissions or the ones from the other APs in its sensing range. In order to forecast its values for any association, we propose an analytical model that estimates BTF. It is based on a Markov network, a conflict graph and the current traffic demands. Our optimization problem consists in finding the association that minimizes the greatest BTF in the network. This original approach allows the Wi-Fi controller to unload the most congested AP and offer more bandwidth to STAs that need it.

The chapter is organized as follows. The BTF is defined in Section 4.2. The model estimating its value for any configuration is presented in the same section. Section 4.3 introduces the optimization problem and the objective function. Simulation results are shown and discussed in Section 4.5.

4.2 Association Model

4.2.1 Assumptions and Scenario

The considered network model and the generic assumptions are defined in Section 2.4. In this chapter, we consider a WLAN in unsaturated mode, where each STA requires a certain throughput. Indeed, at given point in which there may be some APs that do not have any frame for transmission. This assumption allows us to evaluate the performance gain obtained by our association optimization in more realistic scenarios.

A variety of metrics are used, in the literature, for optimizing the association in WLANs. These metrics are chosen according to the desired objective. In the context of unsaturated networks, a classical goal is to unload the most loaded APs by balancing the load between them. The load of an AP can be expressed via different metrics. Metrics such as the number of STAs per AP and the sending throughput of an AP do not give an accurate indication of the real load of an AP, because the load depends on the demand of its associated STAs, the data rate of the links and the activity of the neighboring APs. On the other hand, metrics like the channel utilization or BTF seem more appropriate. This quantity is easily collected from the local Wi-Fi card statistics obtained on the current configuration. From a protocol point of view, it can be collected from the channel load request/report defined in IEEE 802.11k amendment [52]. Therefore, the BTF of the current configuration is known for each AP, but it has also to be predicted for the other configurations that can be considered for a new association. Our prediction model relies on the following assumptions:

- **Data rate**: APs are able to determine the best data rate for all the STAs in its transmission range (associated or not).

- **Throughput**: we assume that a STA, associated to a new AP, will request at least the same throughput as in the current configuration.

- **Probability of success**: the probability of success for each STA (probability that a frame is correctly received) is measured between the AP and STA. Its prediction for another association is difficult. In our model, we assume that this probability remains the same if the STA does not change its channel when it reassociates. In case of a channel change, the success probability is set to the smallest probability of success among the STAs already associated with the new AP.

The objective function based on BTF and the heuristic used to minimize it are presented in Section 4.3. We introduce, in the next section, the analytical model used to estimate BTF for all APs.

4.2.2 Busy Time Fraction estimation

BTF for an AP is defined as the fraction of time the channel is sensed occupied. This measurement can be obtained from the measurement reports of IEEE 802.11k or directly from the physical registers of the interface that measures the busy time according to the CCA mechanism [53]. This quantity is generally available for the current association. But in the context of our optimization, it is necessary to estimate this fraction for any other configuration.

In our model, we define b_j the busy time fraction for an AP j. This time consists of two quantities: the local busy time fraction and the neighbor busy time fraction. The local busy time fraction, denoted b_j^L, corresponds to the time, per second, during which the channel is occupied by its own transmissions. This time takes into account the physical occupation of the channel and the access method times (back-off, DIFS, etc.) The neighbor busy time fraction, denoted b_j^N, is the proportion of time the channel is occupied by APs in its sensing range. It considers only the physical occupation of the channel corresponding to transmissions. We get,

$$b_j = b_j^L + b_j^N$$

Local Busy Time Fraction

This time includes the time to transmit data on the physical channel (T_{PHY}) to one of its STAs and the time of the access method (T_{MAC}). It is necessary to consider T_{MAC} in the computation of the BTF because these times are mandatory for the access method that manage access to the medium. Indeed, the node cannot begin a transmission on the physical link during these times. Consequently, they are part of the busy time fraction and must be taken into account to have a BTF close to 1 at saturation. Thus, the local BTF of an AP j is the sum of the busy time fractions due to its transmissions to all of its STAs.

$$b_j^L = \sum_{i \in S_j} b_{ij}^L$$

where S_j is the set of STAs associated with AP j, and b_{ij}^L the BTF corresponding to the transmissions from AP j to STA i. b_{ij}^L can be computed as follows:

$$b_{ij}^L = \overline{T_{ij}} \times \lambda_i$$

where $\overline{T_{ij}}$ is the average time required for AP j to transmit one datagram to STA i with data rate R_{ij}. λ_i is the average number of datagrams transmitted to STA i in one second. It does not take into account retransmissions.

Nevertheless, a frame is subject to transmission errors and may require one or more retransmissions. According to the IEEE 802.11 standard, the time required for AP j to successfully transmit a frame of size L_i to STA i at data rate R_{ij} after k attempts is given by:

$$T_{ij}(k) = \overbrace{T_{data} + T_{Ack}}^{T_{PHY}} + \overbrace{T_{DIFS} + T_{SIFS} + T_{BO}(k)}^{T_{MAC}}$$

where:

- T_{data} is the duration of the data frame, given by:

$$T_{data} = T_P + T_H + T_{sym} \times Ceiling(\frac{PLCPServiceBits+8\times L_i+PadBits}{N_{DBPS}})$$

- T_{Ack} is the duration of the ACK frame, given by:

$$T_{Ack} = T_P + T_H + T_{sym} \times Ceiling(\frac{PLCPServiceBits+ACK+PadBits}{N_{DBPS}})$$

- $T_{backoff}(k)$ is the average back-off after k unsuccessful successive transmission attempts and is given by:

$$T_{BO}(k) = \frac{min\left(2^k(CW_{min}+1)-1, CW_{max}\right)}{2} \times T_{slot}$$

Note that the value of the parameter N_{DBPS} is different for data and Ack transmission, and may be different at each data retransmission, but is constant during a retransmission. It is consistent with current implementations of data rate adaptation algorithms like Minstrel.

Other related parameters are listed in Table 4.1.

The average time that AP j requires to correctly transmit to STA i or discard a single datagram is [54]:

$$\overline{T_{ij}} = p_{ij}T_{ij}(0) + \sum_{k=1}^{m}\left(p_{ij}(1-p_{ij})^k\left(\sum_{l=0}^{k-1}T^c_{ij}(l) + T_{ij}(k)\right)\right) + (1-p_{ij})^{m+1}\sum_{l=0}^{m}T^c_{ij}(l) \quad (4.1)$$

where m is the maximum number of retransmissions, p_{ij} is the probability of success to transmit from APj to STA i and $T^c_{ij}(l) = T_{BO}(l) + T_{DIFS} + T_P + T_H + T_{data} + T_{ATO}$ is the time between two consecutive transmissions if the frame transmission fails (T_{ATO} is the acknowledgment timeout).

Symbol	Description
T_{DIFS}	DCF Inter Frame Space
T_{SIFS}	Short Inter Frame Space
$T_{backoff}(k)$	Average back-off after k unsuccessful successive transmission attempts
CW_{min} & CW_{max}	The minimum and maximum sizes of the contention window
T_{slot}	Duration of a slot
T_P	Duration of the preamble of the physical layer
T_H	Duration of the header of the physical layer
T_{sym}	Symbol interval
$Ceiling$	Returns the smallest integer value greater than or equal to its argument value
$PLCPServiceBits$	Service field containing 16 bits of zeros to initialize the data scrambler
$PadBits$	Tail (padding) size for each encoding stream
N_{DBPS}	Number of data bits per symbol and is derived from the data rate parameter R_{ij}
ACK	Acknowledgment frame size

Table 4.1: PHY and MAC characteristics of IEEE 802.11

Neighbor Busy Time Fraction

In this section, we present the model to estimate the fraction of time during which the channel is sensed busy by an AP due to the activity of the other APs. An AP senses the channel busy if at least one AP in its vicinity, i.e. in its sensing range, is transmitting.

As presented in Chapter 1, there are two methods to sense the medium busy in IEEE 802.11 standard. One is the *virtual carrier sense* in which NAV is set by APs within its transmission ranges. The second type corresponds to the *physical carrier sense* performed by the CCA mechanism. The Neighbor BTF computation is simple when all APs are in the sensing range of each other and they coordinate their transmission using the DCF mode. But computing the Neighbor BTF becomes very challenging when not all APs are in the same carrier sense range. In other words, some APs cannot detect other APs in the network. This means that on each such a conflicting pair of APs, at a given point in time, at most one AP can transmit. It is natural to represent these symmetric conflict relations by means of an undirected conflict graph [55], where vertices correspond to APs and edges indicate conflicts between them.

To compute the Neighbor BTF of AP j, noted by b_j^N, we introduce a set of notations. We define a random binary variable X_i that indicate if AP i transmitting or not:

$$X_i = \begin{cases} 1 & \text{if } AP_i \text{ is transmitting} \quad (4.2) \\ 0 & \text{otherwise} \quad (4.3) \end{cases}$$

To simplify the notations, we define the event A_i as follows:

$$A_i = \{\text{AP } i \text{ is transmitting}\} \tag{4.4}$$

We define also b_{ij}^N as the fraction of time the channel is sensed busy by AP j due to the transmission of AP i. It is easy to notice that this fraction is the probability that AP i transmitting:

$$b_{ij}^N = Pr(X_i = 1) = Pr(A_i) \tag{4.5}$$

So, the fraction of time that the channel is sensed busy by AP j due to the transmissions from its neighbors is the probability of union of their transmissions (A_i) and given by:

$$b_j^N = Pr\left(\bigcup_{i \in N_j} A_i\right) \tag{4.6}$$

where N_j is the set of neighbors of vertex j in the conflict graph. The events A_i are not disjoint and the computation of the union is consequently not trivial. To compute this probability, we propose to use the **Inclusion-Exclusion Principle** [56], which is defined as:

$$Pr\left(\bigcup_{i \in N_j} A_i\right) = \sum_{k=1}^{|N_j|}\left((-1)^{k-1} \sum_{\substack{I \subset N_j \\ |I|=k}} Pr\left(\bigcap_{l \in I} A_l\right)\right) \tag{4.7}$$

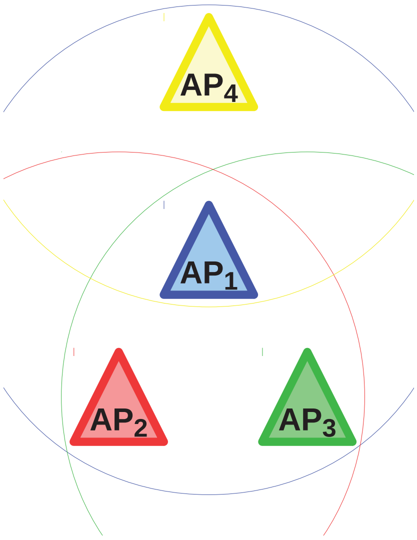
where $I \subset N_j$ with $|I| = k$ describes all the subsets of N_j with cardinal k.

If $|I| = 1$, the computation $Pr\left(\bigcap_{l \in I} A_l\right)$ is trivial. When $|I| > 1$, we have to take into account the conflict graph. Indeed, there are two possible cases that we illustrate through the example given in Figure 4.1. We consider BTF of AP 1. It has three neighboring APs. As there is a link between AP 2 and AP 3, they cannot transmit at the same time and $Pr(A_2 \cap A_3) = 0$. As there is no conflict between AP 3 and AP 4, transmissions from these APs can overlap and $Pr(A_3 \cap A_4) \neq 0$.

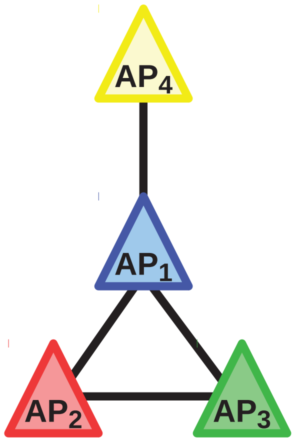
Consequently if two APs in I are neighbors, then their transmissions are exclusive and the intersection is zero:

$$Pr\left(\bigcap_{l \in I} A_l\right)_{\substack{I \subset N_j \\ |I|=k}} = 0, \text{ if } \exists (p,q) \in I^2 \text{ s.t. } p \in N_q (q \in N_p) \tag{4.8}$$

Otherwise, transmissions are not exclusive and this probability may be > 0.



(a) APs Topology



(b) APs conflict graph

Figure 4.1: Topology with 4 APs and its conflict graph

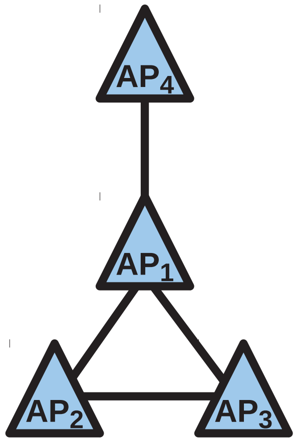
But the events $(A_l)_{l \in I}$ are not independent and the probability of their intersection (overlap) cannot be computed as their product. In order to compute this probability, we consider this problem as an **Undirected Graphical Model** or **Markov Network**. It is based on a graph where the vertices correspond to the events A_i and the edges represent the dependencies between them.

The considered graph is then the same as the APs conflict graph. In Figure 4.2 we show the previous example with a topology with 4 APs conflict graph and the corresponding Markov Network.

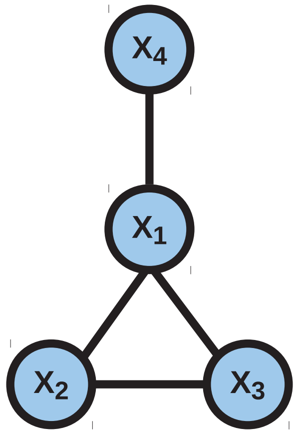
Markov network relies on the Global Markov Property [57], which is defined as follows:

Definition: For any disjoint subsets of the vertices A, B, and C in the graph G such that C separates A and B (i.e. every path between a node in A and a node in B passes through a node in C), the random variables X_A are conditionally independent of X_B given X_C, i.e. $X_A \perp X_B / X_C$, where $X_A = \{X_v\}_{v \in A}$.

In our context, we assume that the transmissions of non-neighboring APs are independent if the set of all their neighbors (union of neighbors) does not transmit:



(a) APs conflict graph



(b) Markov Networkh

Figure 4.2: 4 APs conflict graph and its Markov Network. Formally, the Markov Network is defined as function of the correlations between random variables.

$$\begin{aligned} Pr\left(\bigcap_{l\in I} A_l\right) &= Pr\left(\bigcap_{l\in I} A_l \middle| \bigcap_{l'\in I'} \overline{A_{l'}}\right) Pr\left(\bigcap_{l'\in I'} \overline{A_{l'}}\right), \\ & I \subset N_j, |I| = k, I' = \bigcup_{i\in I} N_i \end{aligned}$$

Applying this property, we obtain:

$$\begin{aligned} Pr\left(\bigcap_{l\in I} A_l\right) &= \prod_{l\in I}\left(Pr\left(A_l \middle| \bigcap_{l'\in I'} \overline{A_{l'}}\right)\right) Pr\left(\bigcap_{l'\in I'} \overline{A_{l'}}\right) \\ &= \left[\prod_{l\in I} \frac{Pr\left(A_l \cap \left(\bigcap_{l'\in I'} \overline{A_{l'}}\right)\right)}{Pr\left(\bigcap_{l'\in I'} \overline{A_{l'}}\right)}\right] Pr\left(\bigcap_{l'\in I'} \overline{A_{l'}}\right) \\ &= \frac{\prod_{l\in I} Pr\left(A_l \cap \left(\bigcap_{l'\in I'} \overline{A_{l'}}\right)\right)}{\left(Pr\left(\bigcap_{l'\in I'} \overline{A_{l'}}\right)\right)^{|I|-1}} \end{aligned} \quad (4.9)$$

Moreover, we have,

$$Pr\left(A_l \cap \left(\bigcap_{l' \in I'} \overline{A_{l'}}\right)\right) = Pr\left(A_l \cup \left(\bigcup_{l' \in I'} A_{l'}\right)\right) \tag{4.10}$$

$$- Pr\left(\bigcup_{l' \in I'} A_{l'}\right) \tag{4.11}$$

and,

$$Pr\left(\bigcap_{l' \in I'} \overline{A_{l'}}\right) = 1 - Pr\left(\bigcup_{l' \in I'} A_{l'}\right) \tag{4.12}$$

By substituting (4.10) and (4.12) in Equation (4.9), we obtain:

$$Pr\left(\bigcap_{l \in I} A_l\right) = \frac{\prod_{l \in I}\left(Pr(\bigcup_{l' \in I' \cup \{l\}} A_{l'}) - Pr(\bigcup_{l' \in I'} A_{l'})\right)}{\left(1 - Pr(\bigcup_{l' \in I'} A_{l'})\right)^{|I|-1}} \tag{4.13}$$

To sum up, b_j^N, given by Equation (4.6), is obtained from the union of the events A_i (Equation (4.7)), itself obtained from Equation (4.13).

We obtain a system of nonlinear equations where each term (variable) is the probability of union of the events $\{A_i\}$. As the number of possible combinations between all events is finite then the system contains a finite number of equations. This system can be solved by any numerical method.

4.3 Association optimization

Our association scheme is based on BTF. This quantity describes the saturation level of an AP. If an AP is saturated, its BTF is close to 1, and the associated STAs are likely restrained in terms of throughput and thus unsatisfied. On the other hand, if BTF is lower, STAs necessarily obtain the required throughput since a part of the bandwidth is available and not used. STAs are then assumed satisfied in terms of their throughput demand. The optimization problem aims to minimize the maximum BTF in the network. Formally, it is given by Equation (4.14).

$$minimize \max_{j \in A} [b_j] \tag{4.14}$$

where A is the set of APs and b_j the BTF of AP j. This objective function has been designed to:

- share the load among APs as it will try (if such solutions exist) to unload the most loaded APs,
- satisfy a maximum number of STAs in terms of throughput, as it will try to decrease BTF of saturated APs,
- increase the STA throughputs, as unsatisfied STAs will be moved to unsaturated APs.

The evaluation of BTF of each AP relies on the model proposed in Section 4.2.2 which predicts b_j (i.e. BTF of AP j) for any association. To solve our optimization problem, we propose to use the same iterative heuristic based on the local search principle described in Chapter 3.

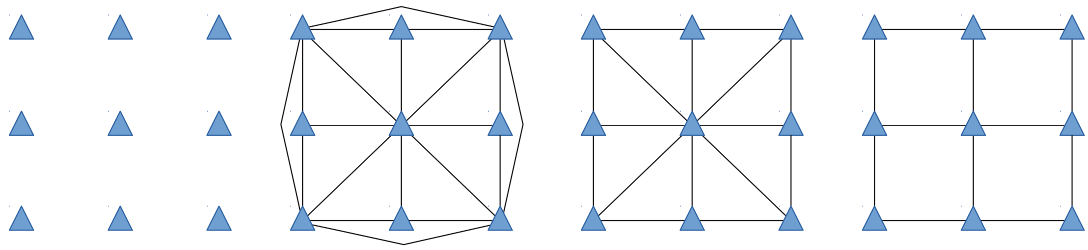
4.4 BTF run time

To evaluate the effectiveness of our approach in terms of computation time, we tested two scenarios. The first one is used to compare the run time of the BTF heuristic with those of the local search presented in Chapter 3. To this end, we considered the same development environment and topology used to evaluate the local search heuristic in Section 3.3.3. This topology is composed of 4 APs and 20 STAs. The run times of the 100 configurations tested are shown in Table 4.2.

Configuration	**Run Time (ms)**	
	AbF	**BTF**
1	0.233	3.690
2	0.185	2.250
3	0.149	1.645
4	0.202	1.096
5	0.154	3.018
6	0.196	2.630
7	0.207	1.791
8	0.338	2.815
9	0.165	2.439
10	0.152	2.564
...	...	...
100	0.142	0.931
Mean	**0.183**	**2,727**

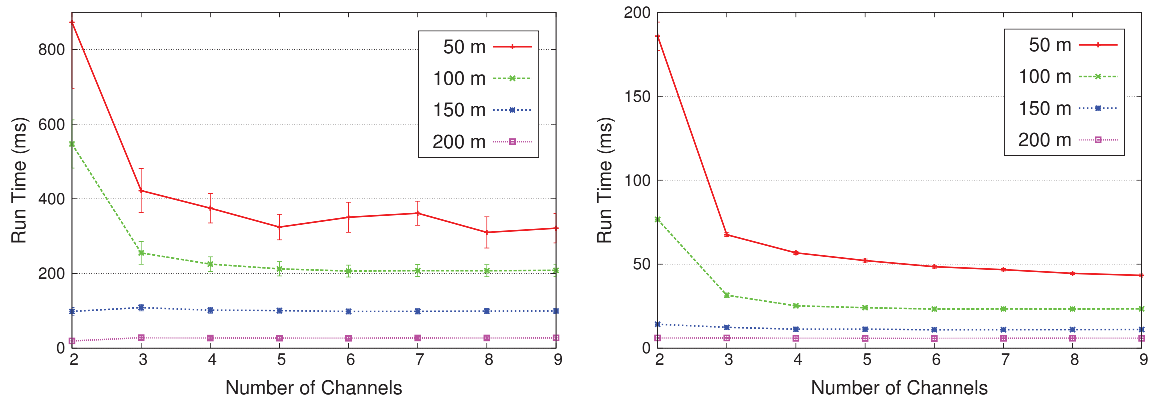
Table 4.2: Comparison of run times between BTF and AbF heuristics

These results shows us that the local search with the AbF model is 15 times faster than



(a) Topology. (b) (100m). (c) (150m). (d) (200m).

Figure 4.3: 9 APs Topology and corresponding conflict graphs for each distance.



(a) Total run time for BTF heuristic. (b) Run time per iteration for BTF heuristic.

Figure 4.4: Total and per iteration run time for BTF heuristic in WLAN with 9 APs.

with the BTF heuristic. However, with an average run time of less than $3ms$, the BTF heuristic remains fairly fast and feasible for a real implementation.

In the second scenario, we considered an even larger and more complex topology. This topology consists of 9 grid-shaped APs and 40 STAs as shown in Figure 4.3a. In this scenario, to show the effect of the network density and the complexity of the conflict graph on the run time, we considered two parameters: the grid step (distance between two adjacent APs) and the number of used orthogonal channels. Figures 4.3b, 4.3c and 4.3d show, in the case of a single channel, the obtained conflict graph corresponding to each distance ($100m$, $150m$, $200m$). The graph for a distance of $50m$ is not shown as it corresponds to a full-mesh. The obtained results are illustrated in Figure 4.4.

Figure 4.4a shows the average run times for the BTF heuristic. We note that the run time decreases when as the distances between APs increase (i.e. WLAN is less dense). Also, as the number of channels, used by the APs, increases, the running time decreases, especially for $50m$ and $100m$. But for $150m$ and $200m$ the run time is almost the same regardless of the number of channels. Despite this, for the densest topology ($50m$ and 2 channels), the run time is less than 1 second.

To further show the correlation of the run time with the network density (the distance between APs and the number of used channels) independently of other parameters such as the number of STAs in the network and their position (data rate used), we have computed these run times by iteration. In our heuristic, an iteration corresponds to a change of association for a single STA. The results obtained are illustrated in Figure 4.4b. These results lead to the same conclusion as with the total run time.

There are two reasons that explain the increase in execution time with increasing network density:

- In a dense topology, each AP has more neighbors, then more terms to compute in the formula of the Inclusion-Exclusion principle (yet, most of these terms are null because of the exclusivity of the transmissions).
- In a dense topology, a STA has more APs in its reception range, so more association possibilities to test during the search procedure.

The obtained results, through these two scenarios, shows that, for a real implementation and even with a very dense WLAN and a small number of orthogonal channels (we have at most 3 orthogonal channels on the 2.4 GHz band), the run time of the BTF heuristic remains reasonable.

4.5 Evaluation

In this section, we present the simulation environment, the performance metrics, and the different simulation scenarios. We then discuss the simulation results.

4.5.1 Simulation configuration

We used a fixed point method [58] to solve the system of nonlinear equations. The optimization heuristic is implemented using C++ programming language. In order to evaluate the proposed approach, we use the network simulation tool "Network Simulator-3 (ns-3)". The radio channel between a STA and its AP is modeled through the ns-3 log-distance path loss model. The transmission power is set to 16 dbm. The number of APs and STAs is fixed for each topology. STAs are associated according to the RSSI value in the initial configuration. The ideal Wi-Fi manager algorithm of ns-3 is used to determine the data rates between APs and STAs. This algorithm sets the data rate between an AP and a STA as a function of the signal over interference plus noise ratio (SINR) value at the destination. For each scenario, we increase the input rates for all STAs (the flow rates between APs and STAs). For each average input rate, simulations are repeated 30 times with different STAs position. Flow rates are constant during a simulation but set randomly for each STA with a given average. Therefore, data rates and flows are different from a STA to another for each simulation. A 95% confidence interval is computed over these 30 samples. The duration of each simulation is 60 seconds.

The simulations were conducted on two different network topologies. The first topology is the WLAN of "ENS de Lyon" at a given floor of the building. This network is composed of 15 APs as shown in Figure 4.5. APs use the ISM frequency band (2.4 Ghz). In this band the number of non-overlapping channels is limited (three orthogonal channels: 1, 6 and 11). This figure shows also the three conflict graphs (one for each orthogonal channel).

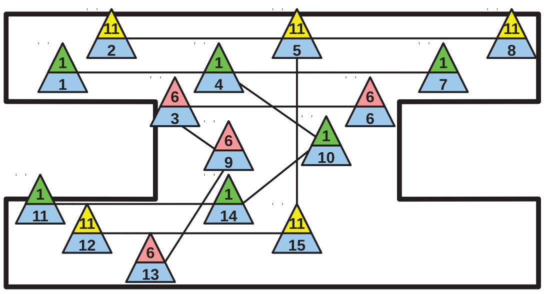


Figure 4.5: WLAN topology at one floor in "ENS de Lyon". The upper number corresponds to the used channel and the lower number corresponds to the AP number.

The second topology is random topology. This topology composed of 25 APs uniformly deployed in a square of size $500m \times 500m$. 100 STAs are distributed in the coverage area of these APs as illustrated in Figure 4.6. APs are configured with 8 orthogonal channels. In this scenario, APs location is changed at each simulation. This randomness allows us to consider an important number of different topologies (30 for each set of parameters).

4.5.2 BTF Estimation

In order to estimate the accuracy of the BTF model proposed in Section 4.2.2, we compare its values obtained by simulation and from the model. The considered topology is "ENS-WLAN". The simulated scenario consists of 60 STAs randomly distributed in the coverage area of the 15 APs. We plot in Figure 4.7 the BTF values according to the average input rate

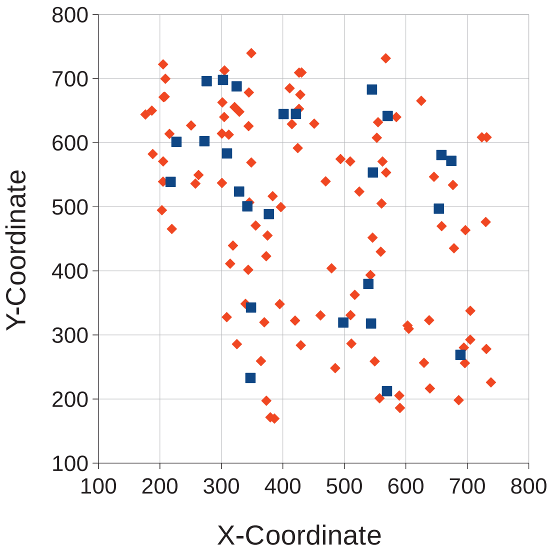


Figure 4.6: WLAN random topology. The blue points represent the APs and the red ones represent the STAs.

(mean of the flow rates). To evaluate the effectiveness of the approach in dense environment, we consider BTF of APs 4 and 5 of our topology (each one is in conflict with 3 other APs).

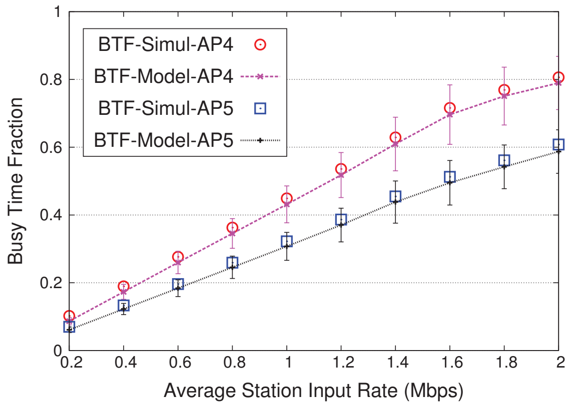


Figure 4.7: Simulation vs Model BTF values for APs 4 and 5 in ENS-WLAN

For AP 4, the difference between the measured BTF (simulation) and model is 7% when the load is low ($BTF < 0.2$). This difference decreases to 2% when the load increases ($BTF > 0.7$). For AP 5 the difference varies in average between 3% and 6%. According to these results, it appears that the model provides a very good estimation of BTF. The model slightly underestimates the BTF as we do not take into account acknowledgments transmitted by the STAs. It is neglected because, to include them in the BTF computation, a full knowledge of the conflict graph is required (in particular conflicts between STAs) whereas in our model only conflicts between APs are considered. It is more realistic from an implementation point of view, but it greatly complicates the model. These conflicts/acknowledgments can be easily integrated in the model if the controller is able to infer them.

4.5.3 Association Optimization

In order to evaluate the improvement brought by our approach, we have considered the two topologies described above. From the simulations results, we compute the following performance parameters:

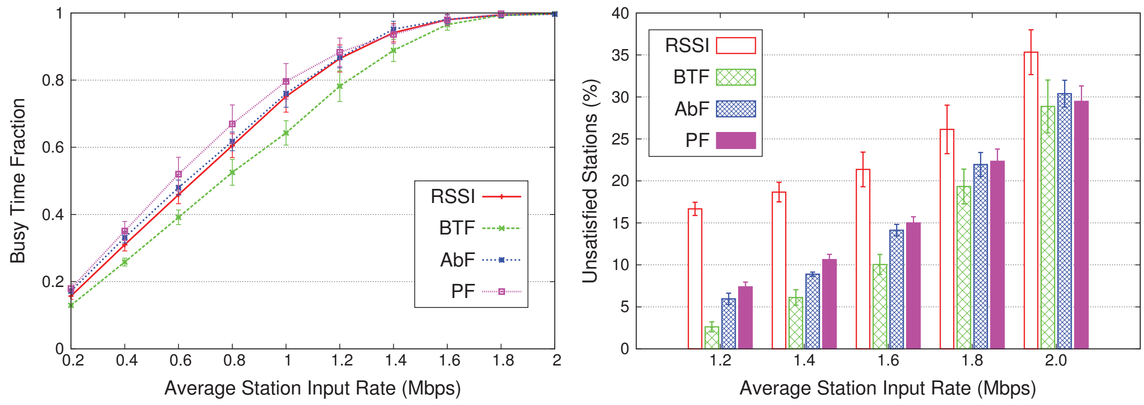
- Busy Time Fraction: for each simulation we consider the greatest BTF in the network.
- Number of unsatisfied STAs: it represents the proportion of STAs that are not satisfied in terms of throughput (i.e. when the ratio between the obtained throughput and the demand is less than 98%).
- Throughput Satisfaction Ratio: it is the ratio between the throughput obtained and the throughput requested for each STA.

Our solution, denoted BTF in the figures, is compared to three existing approaches:

- Initial configuration: STAs associate to APs according to the value of the RSSI. It is denoted as $RSSI$ in the figures.
- Access based Fairness [59]: STAs associated to the same AP have the same opportunity of service in saturated mode. It is denoted AbF in the figures (solution presented in Chapter 3).
- Proportional Fairness [20]: the saturated mode is also considered with an access opportunity to the medium which is proportional to the data rate of each STA (time based fairness). It is denoted PF in the figures.

In the performance evaluation we consider different flow types, as follows:

- UDP: all the packets have the same size (1500 bytes) and the inter-packet time is constant for each STA.
- Real trace: packet sizes vary according to a distribution obtained from a real trace [60] (Average Packet Size = 755.572 bytes and Standard Deviation = 674.05).
- TCP: constant bit rate flows are installed over TCP connections.



(a) BTF of the most loaded AP. (b) Unsatisfied STAs.

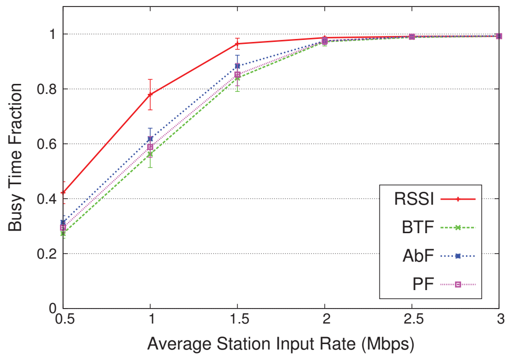
Figure 4.8: BTF of the most loaded AP and proportion of unsatisfied STAs according to the WLAN load for the ENS-WLAN topology.

ENS topology. The first scenario considers the topology of our university (ENS) with UDP flows. Figure 4.8a illustrates the BTF of the most loaded AP as function of the WLAN load. For AbF, the busy time fraction is approximately the same as the one observed for the RSSI association. For PF association, the busy time fraction is even increased of 11% in average. With BTF optimization, the busy time fraction is decreased by approximately 15% when the network is not heavily loaded. This will allow STAs to request more traffic without saturating APs. It clearly appears that AbF and PF approaches are unable to decrease BTF in unsaturated WLAN.

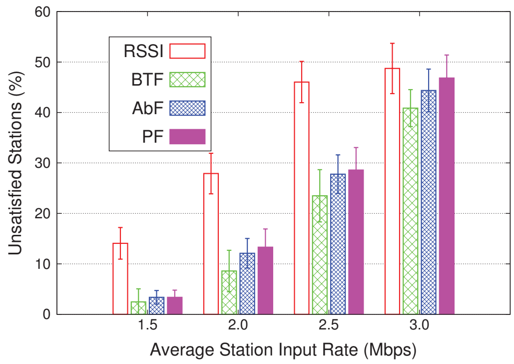
Nevertheless, when the WLAN reaches saturation, the three approaches provide similar results in terms of busy time fraction. To show the benefit of our approach when the WLAN becomes loaded (more than 1 $Mbps$ per STA in the figure), we measure and compare the number of unsatisfied STAs before and after the optimization for the three approaches. Results are shown in Figure 4.8b. Our solution reduces the number of unsatisfied STAs by more than 84% for an average load of 1.2 $Mbps$ per STA, and 18% for an average load of 2 $Mbps$ per STA. For AbF, the gain varies between 64% and 14%, and for PF it varies between 55% and 16%. Even in saturated conditions, our solution still presents a lower number of unsatisfied STAs.

Random topologies. To evaluate our approach with denser topology and more complex conflict graphs between APs, we performed simulations on random topologies.

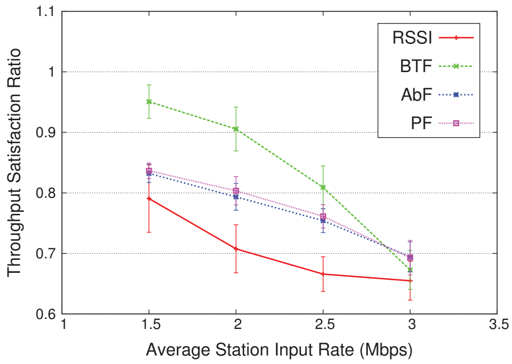
Figure 4.9 illustrates the results of the simulations with the real trace. It allows us to evaluate the performances for different packet sizes. Figure 4.9a shows that the BTF approach



(a) BTF of the most loaded AP.



(b) Unsatisfied STAs.



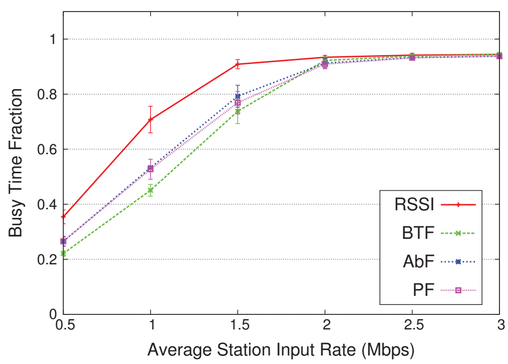
(c) Throughput Satisfaction Ratio.

Figure 4.9: BTF association optimization using random topology with the Real Trace flows.

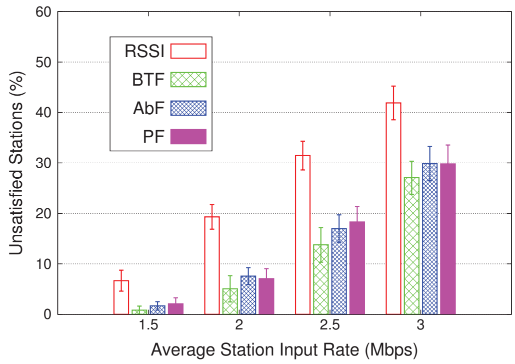
can offload 35% of the most charged AP. AbF and PF approaches allow a load reduction of about 25% and 30% respectively. In Figure 4.9b, the number of unsatisfied STAs is decreased of 54% with BTF optimization with regard to to the RSSI association. For AbF and PF, improvements are 45% and 42% in average. The satisfaction ratio is shown in Figure 4.9c. For the BTF approach, it is increased in average between 3% and 27%. On the other hand, AbF approach allows a gain between 6% and 13%, and between 5% and 14% for PF.

Figure 4.10 shows results with TCP flows. In Figure 4.10a our approach decreases the load of the most loaded AP up to 37%. For the AbF and PF approaches the decrease is almost the same and does not exceed 25%. Regarding the number of unsatisfied STAs shown in Figure 4.10b, BTF allows a gain between 87% and 35%. For AbF, the decrease of the number of unsatisfied STAs varies between 79% and 28%. For PF, the decrease is between 69% and 29%. Figure 4.10c plots the throughput satisfaction ratio. With BTF the STAs gain in throughput on average between 25% and 15%. The AbF and PF approaches allow an average gain of 17% and 16% respectively.

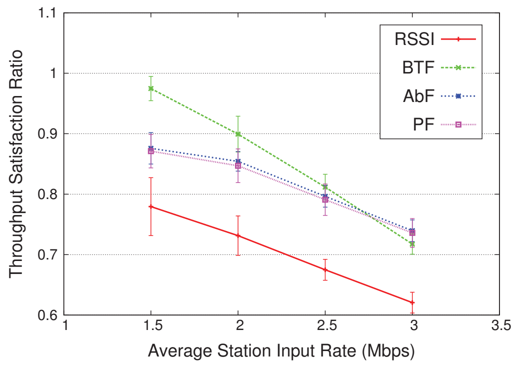
In order to illustrate the impact of the BTF approach in a more realistic implementation context where the optimization process is executed whenever needed (at regular interval for



(a) BTF of the most loaded AP.



(b) Unsatisfied STAs.



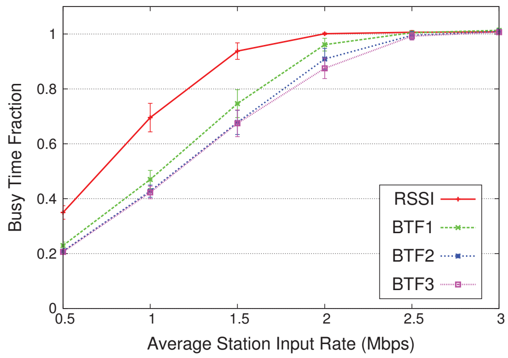
(c) Throughput Satisfaction Ratio.

Figure 4.10: BTF association optimization using random topology with TCP flows.

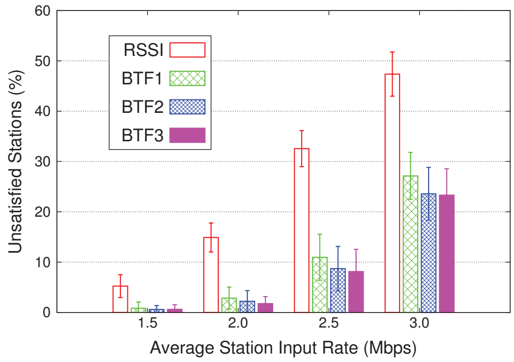
instance), we have simulated the same scenario with UDP flows in which our optimization method is applied 3 times.After each optimization we evaluate the performance and then collect the necessary measures for the next optimization. Results for this scenario are shown in Figure 4.11.

Even if the first optimization allows significant improvements for all performance parameters, the second and third optimizations can further improve these performances. For the greatest BTF in the network, the improvement for the first, second and third optimizations is in average 30%, 35% and 36% respectively. For the unsatisfied STA number, the improvement is 68%, 74% and 75% respectively. For the throughput satisfaction ratio, the improvement is 22%, 27% and 29% respectively.

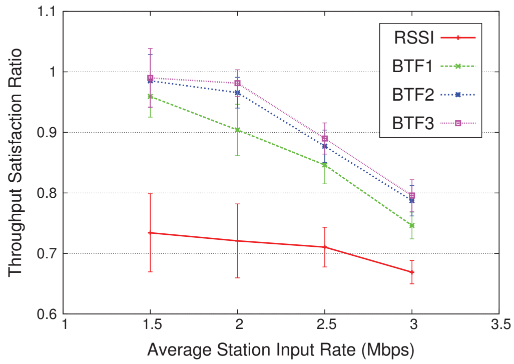
All these results tend to show that our solution generally offers better performance whatever the load of the network. Nevertheless, when the network is very loaded (average STA input rate > 2.5 $Mbps$) the AbF and PF approaches allow to have results close to BTF.



(a) BTF of the most loaded AP.



(b) Unsatisfied STAs.



(c) Throughput Satisfaction Ratio.

Figure 4.11: BTF association optimization using random topology with UDP flows and applied 3 times.

4.6 Conclusion

In this chapter, we propose an original approach for the association optimization in Wi-Fi networks. Our solution is based on a model predicting BTF at each AP and aims to associate STAs in order to minimize the most loaded AP. We have shown through simulations that the model allows an accurate estimation of BTF in the considered configurations. Moreover, the performance evaluation study has shown that such an approach reduces the congestion in the network as it decreases the most loaded AP. This improvement can reach 18% in average when the network is not heavily loaded. Also, our solution decreases the number of unsatisfied STAs, up to 80% when the network becomes saturated and improves throughput of the unsatisfied STAs. When the network is unsaturated, which corresponds to the normal conditions of a Wi-Fi network, approaches based on models that rely on saturated conditions are significantly less efficient that our proposition.

In the next chapter, we extend this model for new versions of Wi-Fi such as IEEE 802.11n/ac. We adapt the BTF estimation model to take into account new enhancements of physical and MAC layers of Wi-Fi such as channel aggregation, frame aggregation, and block acknowledgment.

Chapter 5

Busy Time Fraction Estimation Approach for High Throughput WiFi Networks

Sommaire

5.1 Introduction

The majority of Wi-Fi commercial products, operating on the 2.4 and 5 GHz frequency bands, now implement the IEEE 802.11n and IEEE 802.11ac amendments. These IEEE 802.11 solutions offer high to very high throughput thanks to major modifications of the initial versions of 802.11. These modifications include, among others, the use of the MIMO (Multiple Input / Multiple Output) technology, the possibility to aggregate channels to transmit on a wider channel as well as to aggregate frames in a single frame [53].

Although these versions offer high throughput (up to 600 $Mbps$ for IEEE 802.11n and up to almost 7 $Gbps$ for IEEE 802.11ac), the design and management of these WLANs should still be optimized. Indeed, even if it is very likely that these high throughput Wi-Fi networks will remain mainly not saturated, the load may be very different from one point to another in the network and over time. For instance, some STAs may require low traffic while some others may be unsatisfied due to a high traffic demand in some parts of the network. Balancing the load with a change in the STAs' association is a possible solution to allow unsatisfied STAs to use the remaining bandwidth not used by the satisfied ones.

In the previous chapter, we designed an association mechanism using the channel busy time estimation in order to improve the users' performance in non saturated Wi-Fi networks. This solution offers good performance by reducing the number of unsatisfied STAs in the network and by improving the STAs' throughput. This work only considers the initial versions of IEEE 802.11, like 802.11a/b/g and does not integrate the new features of IEEE 802.11n/ac. It exists studies [23, 26] that test their association solution with IEEE 802.11n/ac, but none of these works integrates in their models and algorithms the specific features of IEEE 802.11n/ac.Nevertheless, these features have an important impact on the STAs' throughput and consequently on the performance of a given association.

In this chapter, we propose an association solution that embeds the main new features of IEEE 802.11n/ac. These features are integrated in the model of our solution that estimates and predicts the share of the radio medium between the APs (APs) and the STAs through the channel busy time fraction. If the MIMO property and the channel aggregation can be easily considered in the data rate parameter of our model, the frame aggregation is more difficult to estimate and to predict. We propose a model based on the conflict graph between APs and the STAs' traffic demand. For this model, we compute a metric, named Hypothetical

Busy Time Fraction (H-BTF), that combines the classical Busy Time Fraction (BTF) and the aggregation mechanism. The metric allows us to express the network load at an AP through a single quantity. Based on this model, we design an association algorithm that aims to distribute the network load among the APs.

The chapter is organized as follows. We discuss, in Section 5.2, the challenges of AP association in the new generations of WLANs that are based on recent standards like IEEE 802.11n. The H-BTF metric is defined in Section 5.3. The model estimating its value for any configuration is presented in the same section. Section 5.4 introduces the optimization problem and the heuristic used to propose approximate solutions. Simulation results are shown and discussed in Section 5.5.

5.2 Challenges with association in aggregation-based WLANs

For a given destination in IEEE 802.11n/ac networks, the AP/STA aggregates in a single MAC Protocol Data Unit (A-MPDU) the frames present in the buffer at the time it gets access to the medium. Consequently, the number of aggregated frames is not fixed and depends on the state of the buffer at the transmission time. This quantity is named *aggregation rate* in the rest of this chapter, and is formally defined as the mean number of frames aggregated in a A-MPDU frame. In our model, this aggregation rate is computed from the current association for each STA and it can change when the STA is associated with another AP. The efficiency of the aggregation mechanism is directly related to the aggregation rate. We illustrate the impact of the aggregation rate on the performance of the network in Figure 5.1. We plot the aggregation rate and the BTF as a function of the input rate for a simple scenario with an AP and a STA. Simulations are performed with ns-3. The throughput is not shown in the figure, but corresponds exactly to the input rate until 136 $Mbps$ where the network saturates.

The aggregation rate varies between 1 (no aggregation) and the maximum number of frames in a single A-MPDU (41 for this configuration). The BTF varies from 0 to 1. We notice that the BTF reaches 1 rapidly. Once it is close to 1, frames cannot access the medium immediately. The buffer is filling up and the aggregation rate increases. But, the aggregation mechanism still allows an enhancement of the throughput from approximately 64 $Mbps$ to 136 $Mbps$.

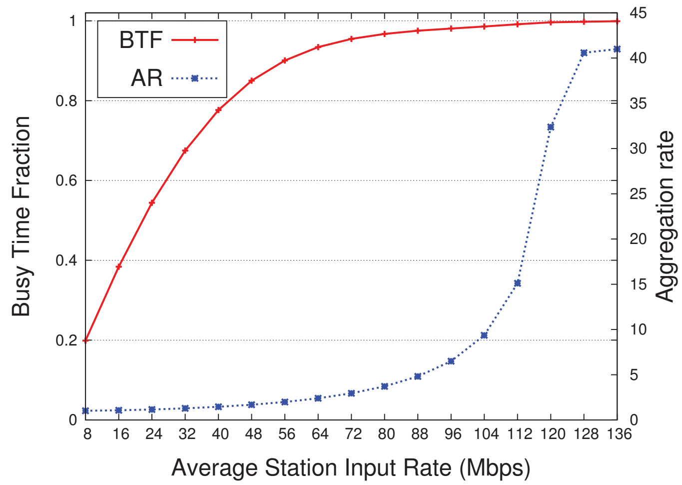


Figure 5.1: BTF and AR (Aggregation Rate) as function of the input rate.

These results show that BTF is not an appropriate metric to reflect the network load when frame aggregation is enable. Therefore, association optimization with the purpose of load balancing between APs cannot be based only on the BTF but must also consider the aggregation rates.

However, it is very difficult to forecast the aggregation rate for a new association as it depends on the statistical nature of the traffic and the contention between the different APs/STAs.

In order to keep the model tractable, and to avoid introducing too many assumptions about the traffic nature, we systematically consider the maximum aggregation rate in the model/computation. The resulting BTF at a given AP represents a Hypothetical BTF (H-BTF) which is less than 1 until the network becomes saturated and that no throughput increase is possible. This metric has the benefit to express the network load through a single quantity. The optimization problem may then be simply expressed without considering complex combination of BTF and aggregation rates.

5.3 System model

5.3.1 Network model

In this chapter we propose a model that generalizes the solution presented in the previous chapter for the new IEEE 802.11 versions. Therefore, this new model is based on the same assumptions and measurements considered in the previous chapters. The improvements made at the physical layer (MIMO, channel aggregation, guard intervals, etc.) are taken into account in the computation of the data rate. We compute for each MCS the corresponding date rate. The frame aggregation mechanism is expressed through the maximum aggregation rate. It is the maximum number of frames aggregated in the same A-MPDU. Its practical computation is linked to three constraints. First, the maximum number of frames acknowledged within a single block acknowledgment is 64 frames. Second, the size of an A-MPDU is limited. The maximum size ranges from 8 to 64 kB (it depends on the Wi-Fi card). Third, the maximum A-MPDU transmission duration is 10 ms. As the transmission duration depends on the data transmission rate, the maximum aggregation rate can change when the STA changes its association.

As already mentioned, our model aims to predict the performance of the network through the H-BTF quantity for different association schemes. The H-BTF computation is presented in the next section.

5.3.2 Hypothetical Busy Time Fraction estimation

The BTF of a given AP is defined as the proportion of time the channel is sensed busy by this AP. It is composed of its own transmissions, and transmissions from its neighbors. This quantity is generally available on the current products through logical registers of the Wi-Fi card, and may also be obtained with the IEEE 802.11k measurement reports. But, in the context of our association solutions, we need to estimate this fraction for any other configuration. Consequently, we propose the following model. As we have seen in the previous chapter, the busy time fraction of an AP j is formally defined as:

$$b_j = b_j^L + b_j^N. \tag{5.1}$$

b_j^L refer to the local busy time corresponding to the transmissions from the AP itself (AP j). b_j^N corresponds to the transmissions from the other APs and detected by the local AP according to the CCA mechanism [53]. Thus, it considers only the physical occupation of the channel corresponding to transmissions.

The local busy time b_j^L is the sum of the transmission times over the set of STAs associated with AP j (denoted S_j):

$$b_j^L = \sum_{i \in S_j} b_{ij}^L$$

b^L_{ij} is thus the BTF corresponding to the transmissions from AP j to STA i. It may be expressed as the product of the average number of datagrams transmitted to STA i per second (denoted λ_i) and the average time $\overline{T}$ required for AP j to successfully transmit a frame of size L to STA i with data rate R_{ij}:

$$b^L_{ij} = \overline{T} \times \lambda_i$$

To compute $\overline{T}$, we consider the time for AP j to transmit a frame of size L in the case it is properly acknowledged (this time is denoted $T_{ij}(k)$) and in the case the aggregated frame is not acknowledged ($T^c_{ij}(k)$) due to transmission errors. When the system is in the k^{th} backoff stage, we get:

$$T_{ij}(k) = (T_{DIFS} + T_{BO}(k))/\tau + T_{data} + (T_{SIFS} + T_{BA})/\tau \tag{5.2}$$

where:

- T_{data} is the duration of the data frame, given in Section 4.2.2.
- T_{BA} is the duration of the block acknowledgment frame, given by:

$$T_{BA} = T_P + T_H + T_{sym} \times Ceiling(\frac{PLCPServiceBits+BlockAck+PadBits}{N_{DBPS}})$$

- $T_{BO}(k)$ is the average back-off after k unsuccessful successive transmission attempts and is also given in Section 4.2.2.

Eq. 5.2 takes into account the different times to access the medium (T_{DIFS}, T_P, T_H, $T_{BO}(k)$) and the times relating to the acknowledgment (T_{SIFS}, T_{BA}).

The times to access the medium and send the acknowledgment are divided by the aggregation rate τ since they are shared by several aggregated frames. When the frame is not acknowledged due to transmission errors, the transmission time becomes:

$$T^c_{ij}(k) = (T_{DIFS} + T_{BO}(k))/\tau + T_{data} + T_{BATO}/\tau \tag{5.3}$$

It takes into account the block acknowledgment timeout T_{BATO} at the AP.

The parameter N_{DBPS} derived from the data rate parameter R_{ij} may be different at each retransmission, but is constant during a retransmission. It is consistent with current Wi-Fi manager implementations like Minstrel.

Finally, we condition the computation of $\overline{T}$ by the number of times the AP j retransmits the same frame to STA i and the initial backoff stage. Indeed, a new frame can be sent with a backoff stage $k > 1$ if the previous frames (present in the previous aggregated frame) have undergone errors. In the formula below, $\overline{T}$ is thus a function of the initial backoff stage h with which the new frame (from AP j to STA i) is sent, and the number of errors that occur for this frame (variable k).

$$\begin{aligned}
\overline{T} \;=\; & \sum_{h=0}^{m} \Big[p_{ij} T_{ij}(h) + \sum_{k=1}^{m} \Big(p_{ij}(1-p_{ij})^k \Big(\sum_{l=0}^{k-1} T_{ij}^c(l+h) + T_{ij}(k+h) \Big) \Big) \\
& + (1-p_{ij})^{m+1} \sum_{l=0}^{m} T_{ij}^c(l+h) \Big] . q_{ij}^h (1-q_{ij}) \\
& + \Big[p_{ij} T_{ij}(m+1) + \sum_{k=1}^{m} \Big(p_{ij}(1-p_{ij})^k \Big(\sum_{l=0}^{k-1} T_{ij}^c(l+m+1) + T_{ij}(k+m+1) \Big) \Big) \\
& + (1-p_{ij})^{m+1} \sum_{l=0}^{m} T_{ij}^c(l+m+1) \Big] . q_{ij}^{(m+1)} \qquad (5.4)
\end{aligned}$$

We remind that p_{ij} is the probability of success to transmit one frame between AP j and STA i, q_{ij} is the probability of transmission failure of at least one frame in an A-MPDU and m is the maximum number of retransmissions.

The computation of the neighbor busy time b_j^N (given in Equation 5.1) in where the medium is sensed busy by AP j due to the transmissions of its neighbors is the same as in Chapter 4.

Hypothetical Busy Time Fraction

The formulas derived in the previous computations take into account the measured aggregation rate τ. In practice, this rate is difficult to forecast for a new association as it requires to model the buffer state at each AP for the new association.

The aggregation rate for each STA depends on the size of the buffer, the arrival rate of the frames for each flow and the time required to transmit a frame (waiting time and transmission time). The first two parameters are known but the third is difficult to estimate. The waiting time of a frame in the buffer depends on the channel state, on the arrival rate and the transmission times which themselves depend on the aggregation rates. Besides, it is easy to understand that if an AP is saturated and with the assumption of a large buffer size, for each transmission there is enough frame in the buffer of the AP to fill the A-MPDU at the destination of each STA. Therefore, the computation of the BTF with maximum aggregation rates indicates the actual load of an AP.

As already mentioned, we rely instead on the maximum aggregation rate denoted τ_{max}. The computation of the metric H-BTF consists in substituting the parameter τ by τ_{max} for all the equations. It expresses the current load with regard to the maximum BTF and the maximum aggregation rate through a unique quantity.

5.4 Association optimization

The model proposed in Section 5.3.2 is used to predict the H-BTF for any association configuration. The value of H-BTF indicates the load level at each AP. So, an H-BTF value less than 1 means that the AP is not saturated and the associated STAs are satisfied. Also, it means that a part of the bandwidth is available and more STAs can associate with this AP. On the other hand, an H-BTF value close to 1 indicates that the AP is saturated or close to saturation. This means that all the STAs associated with this AP are not satisfied and cannot request more throughput.

We propose an optimization approach based on two functions. A first objective function aims to minimize the H-BTF of the most heavily loaded AP:

$$minimize\ f_1(X) \tag{5.5}$$

$$\text{with } f_1(X) = \max_{j \in A} [h_j(X)] \tag{5.6}$$

where A is the set of APs and $h_j(X)$ the H-BTF of AP j for a given association X. This objective function unloads saturated APs and then allows STAs associated with these APs to have more throughput. Once the optimal is found for the first objective function, a secondary objective function is used to share the load between the APs:

$$maximize\ f_2(X) \tag{5.7}$$

$$\text{with } f_2(X) = \sum_{j \in A} log(1 - h_j(X)) \tag{5.8}$$

The first function is used to unload saturated APs and to satisfy a maximum number

Algorithm 2 H-BTF association algorithm

//Initialization
Collect measurements for current solution X_0
Infer the APs conflict graph for each channel
//The optimization loop
while (Convergence() = false) **do**
    $\mathcal{V} = \mathcal{N}(X_0)$;
    $X = \arg\min_{\mathcal{U}\in\mathcal{V}} f_1(\mathcal{U})$;
    if ($f_1(X) < f_1(X_0)$) **then**
        $X_0 = X$;
    else
        $X = \arg\max_{\mathcal{U}\in\mathcal{V}} f_2(\mathcal{U})$;
        if ($f_2(X) > f_2(X_0)$) **then**
            $X_0 = X$;
        end if
    end if
end while
end procedure

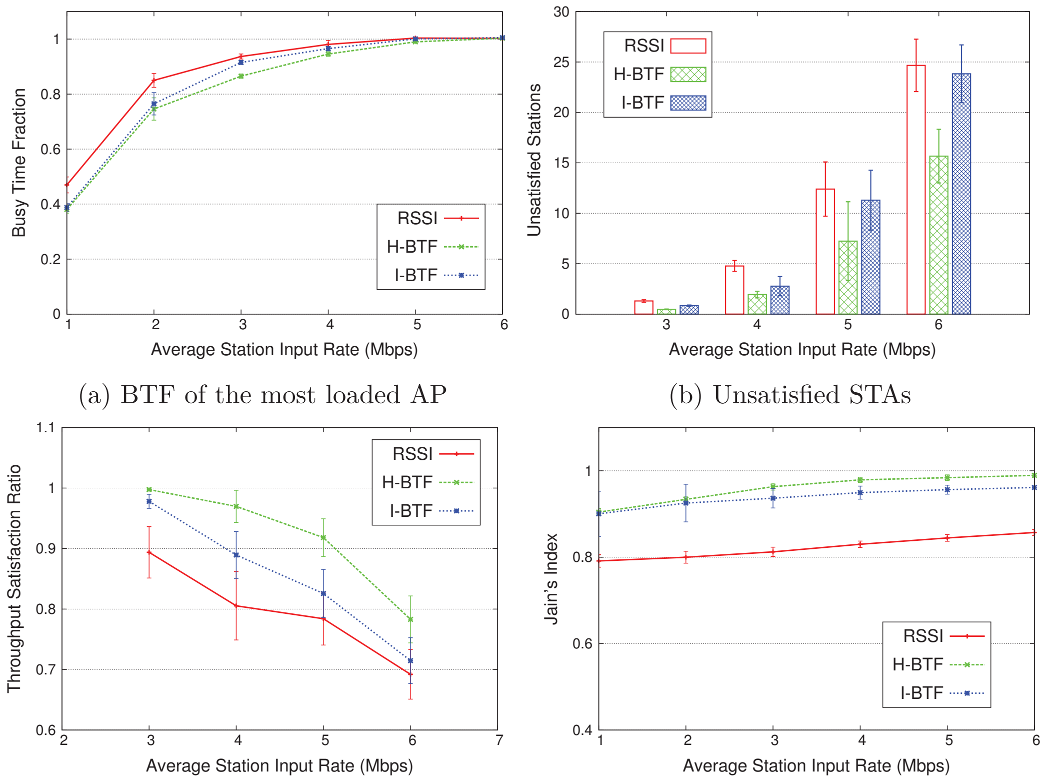
of STAs. But it is a min-max approach that is focused on the most loaded AP. On the other hand, the second function helps to share the load between the APs (once they are decongested). Moreover, note that with the min-max problem, several association schemes may lead to the same value of f_1. The second optimization is thus a way to choose a better association among the set of solutions minimizing $f_1(.)$.

To solve this optimization problem, we propose an iterative heuristic based on the local search principle presented in Section 3.3.

The controller runs the iterative local search Algorithm 2. In this algorithm, the search procedure starts from the current solution (association). The optimization is performed through the local search algorithm on the first objective function f_1 until it reaches a local minimum. Then, the neighborhood of this solution is explored with the second objective function f_2. It improves load sharing between APs and gives the opportunity to exit from local optimum with regard to f_1. The process is repeated until no improvement is observed.

5.5 Evaluation

To evaluate the improvements offered by our approach, we used the same simulation environment and scenarios, the performance metrics and parameters as Chapter 4. However, to evaluate the fairness in terms of load (load balancing) achieved in the network, we compute the Jain's Index [49] which is defined as follows:



(a) BTF of the most loaded AP

(b) Unsatisfied STAs

(c) Throughput Satisfaction Ratio

(d) Jain's Index

Figure 5.2: BTF association optimization using the ENS topology with UDP flows

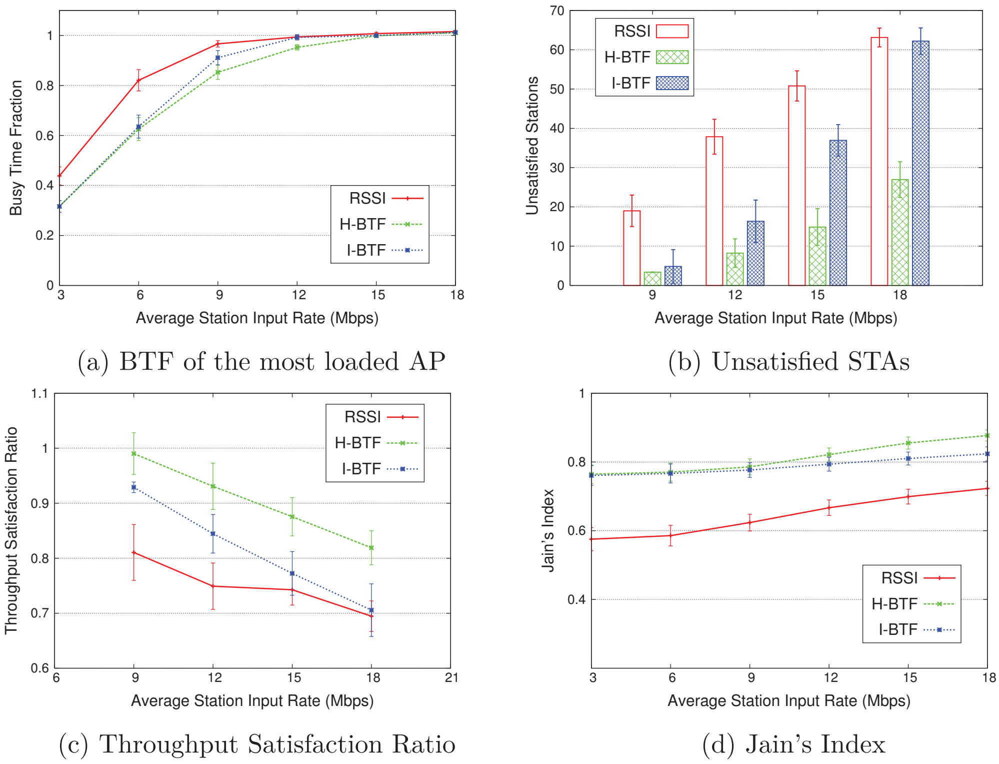
$$Jain = \frac{\left(\sum_{j=1}^{M} b_j\right)^2}{M \sum_{j=1}^{M} b_j}$$

where M is the number of APs in the network.

In this section, we compare the results obtained by our H-BTF algorithm with the initial configuration, where the STAs associate to the APs according to the value of the RSSI. It is denoted RSSI in the figures. Also, we compare our results with a model that uses the aggregation rate τ (not max aggregation rate τ_{max}) predicted from the initial configuration for each STA. This solution is denoted I-BTF for Initial BTF.

ENS topology The first scenario considers the topology of our university (ENS) with UDP flows.

In Figure 5.2a, we plot the values of the BTF of the most loaded AP as a function of the load of the WLAN for the three solutions. Both approaches H-BTF and I-BTF reduce the value of BTF when the network is not heavily loaded. But, when the average STA input



(a) BTF of the most loaded AP

(b) Unsatisfied STAs

(c) Throughput Satisfaction Ratio

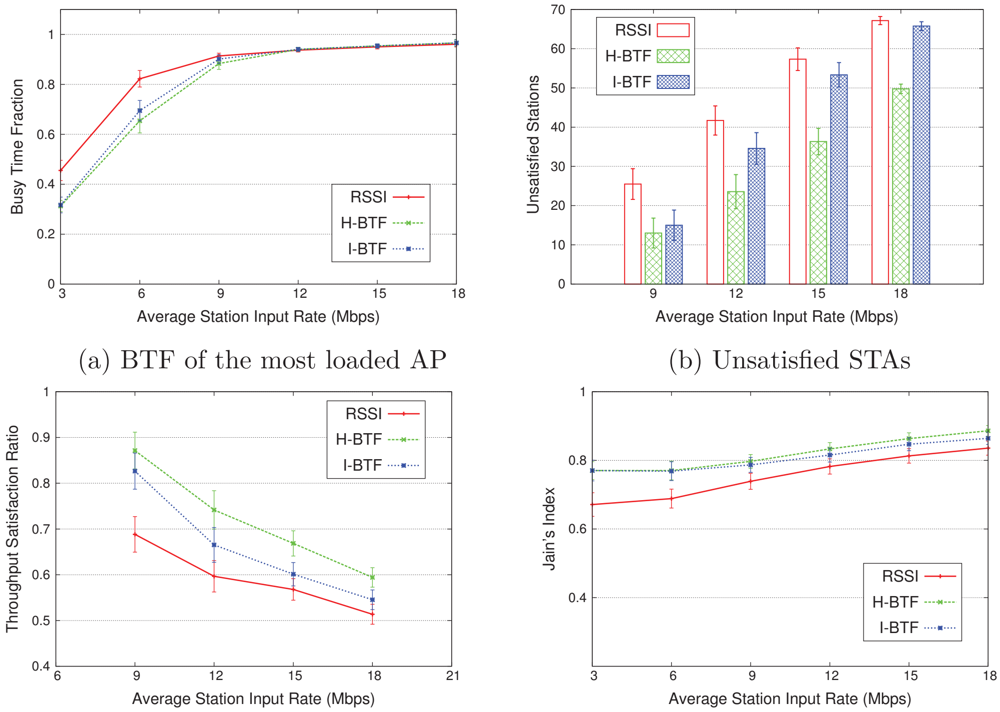
(d) Jain's Index

Figure 5.3: BTF association optimization using the random topology with UDP flows

rate ranges from 2 $Mbps$ to 5 $Mbps$, the H-BTF approach becomes better. For example, at 3 $Mbps$, the decrease in BTF for H-BTF is 8% and 2% for I-BTF. Figure 5.2b illustrates the number of unsatisfied STAs. The H-BTF approach reduces this number up to 64% compared to the RSSI association. For I-BTF the decrease does not exceed 35%. The gap between the two approaches increases to the benefit of the H-BTF approach when the load increases. Figure 5.2c plots the average throughput satisfaction ratio. With H-BTF, the STAs associated with saturated APs, gain in average between 12% and 20% in throughput after the reassociation. With I-BTF the average gain does not exceed 10% compared to the RSSI association. Figure 5.2d shows the Jain index. Both the H-BTF and I-BTF approaches provide close results in terms of fairness with a slight advantage for our approach H-BTF. The I-BTF approach also allows a good load balancing but without unloading the most congested APs.

Random topologies To evaluate our approach with denser topologies and more complex conflict graphs between APs, we performed simulations on random topologies.

The first simulated scenario with random topology consider UDP flows. Figure 5.3a plots the average BTF value of the most loaded AP according to the WLAN load. Our H-BTF



(a) BTF of the most loaded AP (b) Unsatisfied STAs

(c) Throughput Satisfaction Ratio (d) Jain's Index

Figure 5.4: BTF association optimization using the random topology with TCP flows

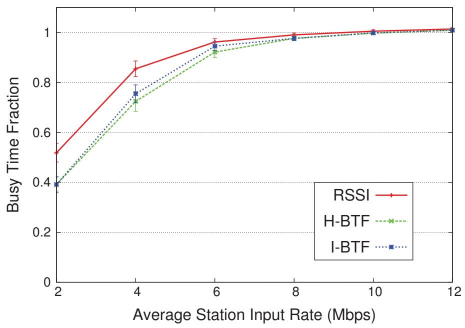
approach reduces this value up to 28% when the network is not very loaded and by 5% when the average rate per STA is around 12 $Mbps$. I-BTF has similar results with a low load, but, when the load increases, the gain is less important than with H-BTF.

In Figure 5.3b, the number of unsatisfied STAs with different load levels is given. The H-BTF approach decreases this number by 82% compared to the RSSI association and by 30% compared to I-BTF for an average input rate per STA of 9 $Mbps$. The performance of H-BTF decreases when the load increases to reach a reduction of 57% compared to RSSI at 18 $Mbps$. But the advantage over I-BTF rises from 30% to 56%.

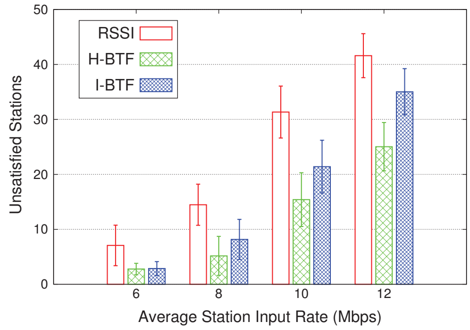
Figure 5.3c plots the STA satisfaction ratio according to the average input rate per STA. The H-BTF approach gives, to unsatisfied STAs, twice much throughput than with the I-BTF approach. The gain, compared to the RSSI solution, varies between 17% and 22% for H-BTF and between 2% and 14% for I-BTF.

The Jain index, shown in Figure 5.3d, indicates that the H-BTF and I-BTF approaches provide similar performance to share load between APs when the network load is low . But when the network load increases, the gap between the two approaches increases to reach 53% at 18 $Mbps$.

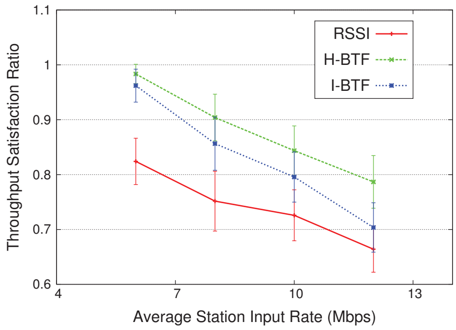
In this scenario, we use TCP flows with the random topology. Figure 5.4a shows the BTF



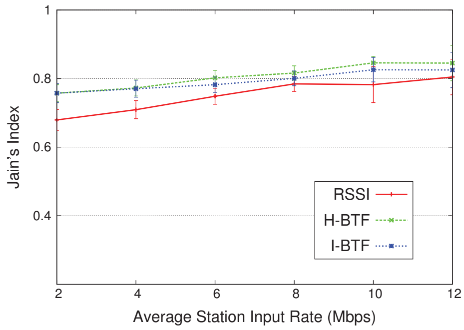
(a) BTF of the most loaded AP



(b) Unsatisfied STAs



(c) Throughput Satisfaction Ratio



(d) Jain's Index

Figure 5.5: BTF association optimization using the random topology with trace flows

value of the busiest AP in the network. When the network is not loaded (input rate less than 9 $Mbps$), this parameter decreases between 20% and 31% for the H-BTF solution and between 15% and 30% for the I-BTF solution. We note that in this scenario, the value of the BTF, even when the network is very loaded, does not exceed 0.95, this being due to the presence of TCP acknowledgments that are not taken into account.

Figure 5.4b illustrates the number of STAs not satisfied. With an average input rate of 9 $Mbps$, the H-BTF solution reduces this number by 50% and the I-BTF solution by 41%. For an average input rate of 18 $Mbps$, this decrease is 25% for the H-BTF and it is only 2% for the I-BTF.

Figure 5.4c shows the satisfaction ratio of the STAs in terms of the requested throughput. H-BTF improves this satisfaction by 21% on average, but the I-BTF only improves it by 10% on average.

Figure 5.4d shows the Jain index according of network load. When the network is not heavily loaded, the H-BTF and I-BTF solutions improve fairness by almost 14%, but as the load increases, this improvement decreases to 6% for H-BTF and to 3.4% for I-BTF.

The last tested scenario considers the real trace with the random topology. Figure 5.5a shows the BTF value of the most loaded AP in the network before and after optimization.

For an average load of less than 6 *Mbps*, the optimization allows to offload the most loaded AP in average by 15% for H-BTF and by 12% for I-BTF.

Figure 5.5b shows the number of STAs not satisfied. This parameter decreases by optimization on average between 39% and 64% for H-BTF solution and between 15% and 60% for I-BTF solution.

In Figure 5.5c, the satisfaction ratio of the STAs is plotted as function of the load of the network. With the H-BTF solution this ratio is improved by approximately 19% and the I-BTF shows an improvement of only 12% in average.

Jain's index is represented in Figure 5.5d. The H-BTF and I-BTF solutions improve Jain's index from 6% to 11% and from 2% to 11% respectively. These results show that our approach is better than the I-BTF approach regardless of the evaluation criteria and for the two topologies (ENS and random). The difference increases with the network load. It confirms the effectiveness of the H-BTF approach and the benefit/need of taking into account the aggregation ratio in the association optimization problem.

5.6 Conclusion

In this chapter, we propose to evaluate the load of a Wi-Fi AP through a new metric named H-BTF (Hypothetical Busy Time Fraction). This metric takes into account the classical busy time fraction and the frame aggregation mechanism offered by the latest IEEE 802.11 amendment. We propose a technological and mathematical framework to compute this metric at each AP of an extended service set. A controller is then able to optimize association between STAs and APs in order to share the load between APs and satisfy a maximum number of STAs in terms of throughput. Our proposal has been evaluated through a large set of simulations performed with ns-3. We have considered several performance criteria to compare this approach with the RSSI association and with a similar approach that uses the initial aggregation ratio. The obtained results illustrate the effectiveness of our proposed approach. It improves performance as STA satisfaction and load balancing between APs.

Conclusion

In this thesis we studied the problem of the association of wireless STAs to APs in Wi-Fi networks. Indeed, this key step has an impact on the performance of STAs and network. In order to propose solutions in accordance with the IEEE 802.11 standard, we have analyzed its main characteristics and modes of operation. This has allowed us to offer simple and effective solutions that are easy to implement in practice and without making modifications to network equipment (APs and STAs). These solutions consist of designing algorithms that allow a controller to centrally optimize and manage the association and reassociation operations.

In the first solution, we proposed a mathematical model to evaluate and predict the throughput that can be obtained by each STA for a given association. In this model, we considered the case of a network where all APs are always saturated. We have also assumed that each AP serves its STAs fairly. Even if these traffic conditions are rare, the optimization of association with these assumptions has the advantage of fairly sharing the bandwidth between the STAs. In the optimization problem, the best association is then defined as the one which maximizes a logarithmic utility function by using the STAs' throughput predicted by the model. The use of a logarithmic utility function allows the controller to achieve a good trade-off between overall throughput and fairness. We also evaluated this solution with the NS-3 simulator on a large set of scenarios and configurations. It empirically demonstrates that our proposal improves the overall throughput and fairness of the network.

The first solution does not take into account the fact that traffic demands can be very different from one STA to another. Therefore, we proposed a second solution for the association optimization problem that relies on real measurements such as STA throughput demands and frame error rates. In this solution, we have formulated an analytical model that allows to estimate the channel busy time fraction (BTF) for each AP and any association configuration. The model is based on a Markov network and a Wi-Fi conflict graph. Associations are optimized to minimize the greatest BTF in the network. This novel approach allows the Wi-Fi manager / controller to unload the most congested AP, increase the throughput of most STAs, and provide more bandwidth to the STAs that need it. NS-3 simulations with a large number of scenarios highlight the benefits of our approach to improve performance in congested and non-congested Wi-Fi networks. This evaluation showed also the accuracy of our BTF estimation, and its ability to balance the load between APs and to satisfy the STA throughput demands.

With the new versions of the IEEE 802.11 standard, the physical and MAC layers have become increasingly complex. Certainly, this has led WLANs to unmatched performance levels, but, at the same time, it makes the design of new models more difficult. Therefore, the main challenge is to propose models that take into account recent enhancements such as spatial multiplexing (MIMO) at the physical layer and the frame aggregation mechanism at the MAC layer. To this end, in our third solution, the new features of the physical layer are considered in the data rate parameter. We investigated the importance of the frame aggregation mechanism to evaluate the AP load. Indeed, we have described an association optimization approach based on the new metric, called H-BTF (Hypothetical Busy Time Fraction), which combines the classical Busy Time fractions (BTF) and the frame aggregation. The model estimates the H-BTF of each AP for current association and is able to predict H-BTF for other association schemes. The association is then optimized to minimize the load on the busiest AP. Thus, this allows to balance the load between the APs and to satisfy the needs of more STAs in terms of throughput. The results of ns-3 simulations showed the significant benefit of the new metric compared to the BTF metric in terms of users' throughput and satisfaction.

It should be noted that the three proposed association models are expressed as combinatorial optimization problems. A heuristic based on a local search algorithm is used to provide approximate solutions to these problems. This heuristic is based on an appropriate neighborhood structure between the associations, and a search procedure adapted to the objective function of each model. This approach has the advantage to be tuned according to CPU and time constraints of the WLAN controller. This choice is also justified by the effectiveness of local research to provide acceptable solutions in a reasonable time for complex combinatorial optimization problems.

Through the work conducted within the framework of this thesis, we illustrated, in this manuscript, how the association optimization can improve the overall performance of WLANs. We have also shown that there is not a single solution that works with any network scenario, traffic shape and standard versions. This proves that every situation requires fine-grained analysis in order to propose powerful models. We have also shown the importance of designing models combining efficiency, tractability, and simplicity of implementation.

Future work

With the emergence of the SDN concept, the control plane tends to be centralized in a single node in the network (one controller or several controllers), even in wireless networks. On the other hand, there is a need to manages network resources and services with more flexibility, efficiency and dynamically. A natural extension of this work is thus to integrate our solutions to the SDN concept.

However, one of the big challenges of centralized management, including SDN, is related to the problem of scalability and its constraints. Indeed, in a large WLAN, with a single controller in charge, new problems may arise such as the complexity of control algorithms that generally exponentially grow with the size of the network and the latency that may increase due to the geographic extent of the network. One of the possible approaches to address these issues is the concept of clustering. To ensure the scalability of a centralized optimization, it is possible to split the WLAN in several clusters and to restrict the optimization problem within each cluster. Since WLANs have a geographical limited coverage area, clustering is done on APs and STAs based on their locations. Optimization of the association is then centrally performed by the controller for each cluster.

Bibliography

[1] *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)*, pages 1–565, Oct 2009.

[2] *IEEE Std 802.11ac-2013 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012, IEEE Std 802.11aa-2012, and IEEE Std 802.11ad-2012)*, pages 1–425, Dec 2013.

[3] L. Zhang, L. Zhao, Z. Wang, and J. Liu. WiFi Networks in Metropolises: From Access Point and User Perspectives. *IEEE Communications Magazine*, 55(5):42–48, May 2017.

[4] A. Farshad, M. K. Marina, and F. Garcia. Urban WiFi Characterization via Mobile Crowdsensing. In *2014 IEEE Network Operations and Management Symposium (NOMS)*, pages 1–9, May 2014.

[5] P. Calhoun, M. Montemurro, and D. Stanley. Control And Provisioning of Wireless Access Points (CAPWAP) Protocol Specification. RFC 5415, RFC Editor, March 2009.

[6] *IEEE Std 802.11v-2011 — Amendment 8: IEEE 802.11 Wireless Network Management*, pages 1–433, Feb 2011.

[7] Mao Yang, Yong Li, Depeng Jin, Lieguang Zeng, Xin Wu, and Athanasios V. Vasilakos. Software-Defined and Virtualized Future Mobile and Wireless Networks: A Survey. *Mobile Networks and Applications*, 20(1):4–18, Feb 2015.

[8] Y. Bejerano, S. J. Han, and L. Li. Fairness and Load Balancing in Wireless LANs Using Association Control. *IEEE/ACM Transactions on Networking*, 15(3):560–573, June 2007.

[9] G. Bianchi. Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.

[10] Y. Lee. Throughput Analysis Model for IEEE 802.11e EDCA with Multiple Access Categories. *Journal of Applied Research and Technology*, 11(4):612 – 621, 2013.

[11] Cisco Visual Networking Index. *White Paper*, 2017.

[12] M. S. Gast. *802.11n: A Survival Guide*. O'Reilly, 2012.

[13] Mauricio G.C. Resende and Panos M. Pardalos. *Handbook of Optimization in Telecommunication*. Springer US, 2006.

[14] J. Kennington, E. Olinick, and D. Rajan. *Wireless Network Design: Optimization Models and Solution Procedures*. Springer-Verlag New York, 2010.

[15] Fabio D'Andreagiovanni, Carlo Mannino, and Antonio Sassano. GUB Covers and Power-Indexed Formulations for Wireless Network Design. *Management Science*, 59(1):142–156, 2013.

[16] Bernard Gendron, Maria Grazia Scutellà, Rosario G. Garroppo, Gianfranco Nencioni, and Luca Tavanti. A Branch-and-Benders-Cut Method for Nonlinear Power Design in Green Wireless Local Area Networks. *European Journal of Operational Research*, 255(1):151 – 162, 2016.

[17] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance Anomaly of 802.11b. In *INFOCOM 2003*, volume 2, pages 836–843, March 2003.

[18] W. Li, S. Wang, Y. Cui, X. Cheng, R. Xin, M. A. Al-Rodhaan, and A. Al-Dhelaan. AP Association for Proportional Fairness in Multi-rate WLANs. *IEEE/ACM Transactions on Networking*, 22(1):191–202, Feb 2014.

[19] Heyi Tang, Liu Yang, Jiang Dong, Zhonghong Ou, Yong Cui, and Jianping Wu. Throughput Optimization via Association Control in Wireless LANs. *Mobile Networks and Applications*, 21(3):453–466, Jun 2016.

[20] O. B. Karimi, J. Liu, and J. Rexford. Optimal Collaborative Access Point Association in Wireless Networks. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 1141–1149, April 2014.

[21] L. Li, M. Pal, and Y. R. Yang. Proportional Fairness in Multi-Rate Wireless LANs. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, April 2008.

[22] Huazhi Gong and JongWon Kim. Dynamic Load Balancing Through Association Control of Mobile Users in WiFi Networks. *IEEE Transactions on Consumer Electronics*, 54(2):342–348, May 2008.

[23] W. Wong, A. Thakur, and S. H. G. Chan. An Approximation Algorithm for AP Association Under User Migration Cost Constraint. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[24] S. Keranidis, T. Korakis, I. Koutsopoulos, and L. Tassiulas. Contention and Traffic Load-Aware Association in IEEE 802.11 WLANs: Algorithms and Implementation. In *2011 International Symposium of Modeling and Optimization of Mobile, Ad Hoc, and Wireless Networks*, pages 334–341, May 2011.

[25] Mario Collotta. FLBA: A Fuzzy Algorithm for Load Balancing in IEEE 802.11 Networks. *Journal of Network and Computer Applications*, 53:183 – 192, 2015.

[26] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas. Optimizing Client Association for Load Balancing and Fairness in Millimeter-Wave Wireless Networks. *IEEE/ACM Transactions on Networking*, 23(3):836–850, June 2015.

[27] S. K. Dandapat, B. Mitra, R. R. Choudhury, and N. Ganguly. Smart Association Control in Wireless Mobile Environment Using Max-Flow. *IEEE Transactions on Network and Service Management*, 9(1):73–86, March 2012.

[28] Suzan Bayhan and Anatolij Zubow. Optimal Mapping of Stations to Access Points in Enterprise Wireless Local Area Networks. In *Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '17, pages 9–18, New York, NY, USA, 2017. ACM.

[29] F. D'Andreagiovanni, R. G. Garroppo, and M. G. Scutellà. Power Savings with Data Rate Guarantee in Dense WLANs. In *2017 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, pages 1–8, May 2017.

[30] F. D'Andreagiovanni, R. G. Garroppo, and M. G. Scutellà. Green Design of Wireless Local Area Networks by Multiband Robust Optimization. In *International Network Optimization Conference (INOC)*, 2017.

[31] Jaeryong Hwang, Jaehyuk Choi, Joon Yoo, and Chong-kwon Kim. A Theoretical Approach to Optimal Association Control in Vehicular Wi-Fi Networks. *EURASIP Journal on Wireless Communications and Networking*, 2014(1):192, Nov 2014.

[32] Insoo Sohn. Access Point Selection Game with Mobile Users Using Correlated Equilibrium. *PLOS ONE*, 10(3):1–13, 03 2015.

[33] Z. Chen, Q. Xiong, Y. Liu, and C. Huang. A Strategy for Differentiated Access Service Selection based on Application in WLANs. In *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 317–322, April 2014.

[34] J. B. Ernst, S. Kremer, and J. J. P. C. Rodrigues. A Utility based Access Point Selection Method for IEEE 802.11 Wireless Networks with Enhanced Quality of Experience. In *2014 IEEE International Conference on Communications (ICC)*, pages 2363–2368, June 2014.

[35] E. Zola, F. Barcelo-Arroyo, and A. Kassler. Multi-Objective Optimization of WLAN Associations With Improved Handover Costs. *IEEE Communications Letters*, 18(11):2007–2010, Nov 2014.

[36] P. B. Oni and S. D. Blostein. AP Association Optimization and CCA Threshold Adjustment in Dense WLANs. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6, Dec 2015.

[37] A. Ozyagci, K. W. Sung, and J. Zander. Association and Deployment Considerations in Dense Wireless LANs. In *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, pages 1–5, May 2014.

[38] B. P. Tewari and S. C. Ghosh. Interference Avoidance through Frequency Assignment and Association Control in IEEE 802.11 WLAN. In *2014 IEEE 13th International Symposium on Network Computing and Applications*, pages 91–95, Aug 2014.

[39] D. Gong and Y. Yang. On-Line AP Association Algorithms for 802.11n WLANs with Heterogeneous Clients. *IEEE Transactions on Computers*, 63(11):2772–2786, Nov 2014.

[40] Arpit Gupta, Jeongki Min, and Injong Rhee. WiFox: Scaling Wi-Fi Performance for Large Audience Environments. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, pages 217–228, New York, NY, USA, 2012. ACM.

[41] F. Malandrino, C. F. Chiasserini, and S. Kirkpatrick. Cellular Network Traces Towards 5G: Usage, Analysis and Generation. *IEEE Transactions on Mobile Computing*, 17(3):529–542, March 2018.

[42] Qunfeng Dong, Suman Banerjee, and Benyuan Liu. Throughput optimization and fair bandwidth allocation in multi-hop wireless lans. *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–12, 2006.

[43] F. P. Kelly. Charging and Rate Control for Elastic Traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.

[44] Gilles Berger-Sabbatel, Andrzej Duda, Martin Heusse, and Franck Rousseau. Short-Term Fairness of 802.11 Networks with Several Hosts. In Elizabeth M. Belding-Royer, Khaldoun Al Agha, and Guy Pujolle, editors, *Mobile and Wireless Communication Networks*, pages 263–274, Boston, MA, 2005. Springer US.

[45] T. Bu, L. Li, and R. Ramjee. Generalized Proportional Fair Scheduling in Third Generation Wireless Data Networks. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–12, April 2006.

[46] David B. Shmoys and Éva Tardos. An Approximation Algorithm for the Generalized Assignment Problem. *Mathematical Programming*, 62(1):461–474, Feb 1993.

[47] J. P. Walser. *Integer Optimization by Local Search: A Domain-independent Approach*. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 1999.

[48] NS-3: A Discrete-Event Network Simulator for Internet Systems. `version3.23.https://www.nsnam.org/`, 2015.

[49] Raj Jain, Dah-Ming Chiu, and W. Hawe. A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems. *CoRR*, cs.NI/9809099, 1998.

[50] NS-3 code: Ideal HT Wi-Fi manager. `http://perso.ens-lyon.fr/mohammed.amer/Ideal-ht-wifi-manager`.

[51] Institute of Communication Networks and Computer Engineering. Trace of the Dormitory Network Selfnet. `http://www.ikr.uni-stuttgart.de/~sass/traces/Welcome.html`, 2004.

[52] *IEEE 802.11k-2008 — Amendment 1: IEEE 802.11 Radio Resource Measurement*, 2008.

[53] 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. *IEEE Std 802.11-2012*, pages 1–2793, March 2012.

[54] V. Atanasovski and L. Gabrilovska. Influence of Header Compression on Link Layer Adaptation in IEEE 802.11b. In *2005 International Conference on Wireless Networks, Communications and Mobile Computing*, volume 2, pages 1551–1556 vol.2, June 2005.

[55] Anthony Busson, Eric Fleury, and Ngoc Minh Phung. A Simple Method to Infer Wi-Fi Conflict Graph. In *CORES2017*, Quiberon, France, May 2017.

[56] Titu Andreescu and Zuming Feng. *Inclusion-Exclusion Principle*, pages 117–141. Birkhäuser Boston, Boston, MA, 2004.

[57] D. Edwards. *Introduction to Graphical Modelling*. Springer Texts in Statistics. Springer New York, 2012.

[58] V. Berinde. *Iterative Approximation of Fixed Points*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007.

[59] Mohammed Amer, Anthony Busson, and Isabelle Guérin Lassous. Association Optimization in Wi-Fi Networks: Use of an Access-based Fairness. In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWiM ’16, pages 119–126. ACM, 2016.

[60] The CAIDA Anonymized Internet Traces 2015 Dataset. `http://www.caida.org/data/passive/passive_2015_dataset.xml`, 2015.