CHAPTER 2

LITERATURE REVIEW

A survey of literature has been carried out in order to identify the research problem, formulate the objectives, and determine the methodology and to evaluate the proposed solution of this research. It was noticed that the literature pertaining to the chosen area of research are certainly large in number, however information on association between AP and MU, and Handover between APs in WLAN are precise, concise and less focused. Hence, this research tries to resolve the lacunas that are predominantly observed in the reported literature. The major contributions by researchers reported are briefly discussed here.

2.1 A BROAD OVERVIEW OF LOAD BALANCING IN WLAN

2.1.1 Association Schemes

The Distributed Coordination Function (DCF) MAC protocol has been used by 802.11 is able to give equal long-term channel access provisions to all contending MUs by Koksal et al. (2000) and Tay et al. (2001).

The IEEE 802.11b standard defines four modes of bit rates (1, 2, 5.5 and 11 Mbps). APs and MUs select a specific mode based on the WLAN conditions. In methods and advantages of multi-rate protocols have been exposed in Holland et al. (2001): Usually, a MU with a relatively low signal to noise and interference ratio chooses a low transmission rate to improve its bit error rate. However, the 802.11 medium access control (MAC) protocol
provides per-frame fairness, meaning that in the long term MUs have the same chance to access the medium and send their frames (all MUs should transmit with an equal average frame rate over a longer time horizon). As the time duration required to transmit a frame with a low transmission rate is much longer than the duration for the same frame size with a higher transmission rate, a low transmission rate MU will occupy the channel for a longer time. This phenomena degrades the throughput of high rate MUs if they are associated to the same AP.

In Papanikos and Logothetis (2001), used Probe Response messages sent by the APs provide the number of associated MUs and information about the received signal value RSSI from the requesting MU. MUs use this information to calculate a weight for all reachable APs and then associate with the AP with lowest weight. Authors determine MU and AP association based on RSSI and the number of MUs associated with each AP.

Judd and Steenkiste (2002) and Fukuda and Oie (2004) propose that APs maintain a measurement of their load and broadcast beacons containing this load to MUs in the cell. New MUs receive beacons from multiple APs and use this information to associate with the least loaded AP.

Balachandran et al. (2002) provide the best AP and MU association scheme, and the network also suggests roaming to APs located further away if nearby APs are considered unable to cover the MU’s requirements. In order to implement these solutions, it is necessary to modify the MUs: firstly, they have to send new management frames before they are actually going to associate; secondly, they will no longer be responsible for association or roaming decisions. The first issue can be solved by using new radio measurements (future IEEE 802.11k devices). There is no standardized procedure for solving the second issue as yet, but it is expected to be revised by the IEEE 802.21 group, which will provide new mechanisms intended to
assist handovers, and by IEEE 802.11v, which is in the development stage and will include management capabilities to allow network-directed roaming. It is not vital to solve the second of these issues, since it is also possible to perform implicit admission control/association management. This involves actions taken on the network side that induce the desired MU behavior, and therefore leave the roaming and association decisions to MUs so that hardware/software modifications are not required. And they suggest that MUs associate with the AP that can accommodate its minimum bandwidth requirements.

For example, in Heusse et al. (2003) it has been shown that if a MU with a transmission rate of 11 Mbps shares the channel with a MU at a transmission rate of 1 Mbps, the throughput of the 11 Mbps MU is about the same as that of the 1 Mbps MU (assuming an equal traffic load of each MU as well as the saturation mode). Consequently, aggregate throughput and throughput of high bit rate MUs may be dramatically brought down; because low bit rate MUs will occupy more channel access time to transmit an equal amount of data. Such “performance anomaly” of 802.11 WLANs has been reported by Heusse et al. (2003).

It is obvious that this rather and simple. Selection process can lead to problems concerning the network performance of larger areas with many MUs and several APs by Arbaugh et al. (2003) and Bejerano et al. (2004).

Bejerano and Bhatia (2004) propose different AP selection method by allowing MUs to switch to others APs in order to balance workload among APs. However, although such load balancing techniques can effectively improve fairness, they are still restricted to use MU-AP links and thus achievable throughput improvement is limited.
Gambiroza et al. (2004) attempt to address the fair bandwidth allocation problem in the context of wireless backhaul networks, where the network topology is nearly static. In their work, the authors only provide an optimal solution for bandwidth assignment in the special case where all the links interfere with each other, i.e., the link contention graph is a circle. However, in many situations not all APs and MUs are in direct interference range of each other. Finding an optimal solution in this type of case is very difficult.

In Fukuda et al. (2005) an AP selection mechanism is proposed whereby the selection metric is based on wireless channel conditions rather than the received signal strength. However, the authors have assumed the same bit rate for all MUs which do not occur in most practical settings. We identify the core problem of AP selection to be the choice of metric to consider and whether the choice should be (periodically) reevaluated or not. Potentially, a more effective AP selection mechanism might significantly improve the overall network throughput while also improving the individual MUs rates and delays.

In particular, each node has (approximately) the same number of opportunities to transmit a data frame, regardless of its bit rate and hence the amount of channel access time needed. If MUs transmit packets of similar sizes and experience similar loss rates, they achieve approximately the same throughput irrespective with their bit rates. This is referred to as throughput-based fairness by Tan et al. (2004).

When multiple such APs are available, the AP with highest RSSI is chosen. Network-wide max-min fair bandwidth allocation algorithms were developed by Bejerano et al. (2004). In their scheme, each MU deploys appropriate MU software to monitor the wireless channel quality it experiences from all its nearby APs. The MU then reports the information to a
network control center, which determines MU and AP association. Their algorithms are the first that provide worst-case guarantees on the quality of the bandwidth allocation.

The algorithm proposed in Velayos et al. (2004) is more sophisticated but follows a similar logic. There are three possible AP states: under-loaded (will accept any request), balanced (will not accept extra load) and over-loaded (will expel the MU on the assumption that it will automatically request a less loaded neighboring AP).

In Wang et al. (2004), APs could even build a custom radiation pattern to balance load, but besides a very specialized RF hardware, this solution relies on the APs’ perfect knowledge of their own coverage and the exact position of MUs, which is hardly feasible.

For example, in Sang et al. (2004) proposed to base of the AP selection decision on the number of MUs associated per AP. While this is easy to implement, specific effects in IEEE 802.11 lead to severe problems with this approach as well.

In Fanglu et al. (2005), a dedicated wireless load balancer has been proposed, which distributes MUs among APs to mitigate the low rate MUs effect on high rate ones. A potential drawback of this solution is that it requires a centralized entity in the WLAN such as an Access Controller (AC) or a Centralized Switch, which is dedicated to manage the network resources. Between this central unit and the MUs some signaling traffic has to be conveyed, which consumes some resources.

In Bazzi et al. (2005) the APs accept or deny new association requests depending on the respective load. When the first choice is rejected,
the MUs will send association requests to the next AP in the signal strength-arranged list until they are admitted.

Another way of implementing network-directed MU-driven load balancing could be achieved by using the concept of cell breathing. Cell breathing techniques consist in dynamically modifying cell dimensions by increasing or reducing transmission power. Cell breathing is a side effect in CDMA networks that reduces the cell coverage when more MUs are supported, but this could be advantageous in load balancing techniques if optimal strategies are applied. The concept of cell breathing for load balancing in WLANs is explained in Brickley et al. (2005): a highly congested AP reduces its coverage radius so that the furthest MUs lose connectivity and try to roam to less loaded APs. An under-utilized AP may increase it’s transmit power in order to expand its coverage. Consequently, new MUs will roam to this AP and the load on neighboring APs will decrease.

A purely MU-driven and non-intrusive association scheme is proposed in Vasudevan el al. (2005): a MU observes a skewed time period of beacon frame receptions to estimate the available bandwidth. This method for estimating available bandwidth introduces large delays due to the channel observation time (several beacon intervals).

Velayos et al. (2006) propose to combine their load balancing with explicit admission control, based on medium access delay measurements. While these techniques provide good admission control and load balancing, they do not guarantee that all MUs with network access rights will be fairly served.

Bejerano el al. (2006) provides an in-depth analysis of cell breathing in IEEE 802.11 WLANs and proposes a centralized solution based
on two different algorithms: one is aimed at reducing the load of the most congested AP and the other tries to find the Min-Max load balanced solution, but they are not able to provide better performance when most of the MUs are associated with same AP.

In the case of continuous power assignment, linear programming is used to maximize throughput, while in the case of discrete values a greedy algorithm is used. However, as stated in Garcia et al. (2006), the furthest MUs may sometimes be expelled arbitrarily in spite of the fact that they may be contributing an insignificant load, depending on the applications they run.

Reference Garcia et al. (2006) also gives an overview of different load balancing techniques that can be applied by using IEEE 802.11k measurements and statistics.

If the MUs were able to gather more information they would be able to perform smarter associations and therefore maximize the network efficiency. In Nicholson et al. (2006), MUs perform different tests on all APs within range in order to determine the most suitable association.

A preferable and less intrusive solution would be to introduce a trade-off between the received signal strength and cell utilization. MUs must choose an AP that is close enough to allow frame exchange with the minimum received signal quality that guarantees correct transmission at the highest possible bit rate. However, it is advisable to avoid cell saturation by selecting the AP with the lowest load.

Bahl et al. (2007) provide five different versions of a cell breathing scheme, depending on the presence of heterogeneous or non-heterogeneous traffic demands, and depending on the availability of continuous power assignments as against a discrete set of power levels.
Load balancing in WLANs has been intensely analyzed in Bejerano et al. (2007) association control algorithm was proposed for bandwidth allocation with constant bit rate.

The authors in Gong et al. (2008) and Fukuda and Oie (2007) also propose including new load information in beacon frames so that association decisions remain on the MU side. Different commercial products use similar systems in which the APs announce their utilization in beacon frames. However, as with all proprietary solutions, there are interoperability problems that could be solved by the new IEEE 802.11k standard.

The authors of Bejerano et al. (2004) and Jabri et al. (2008) propose network-controlled schemes in which MUs send the required information to a central unit, which also has access to the load information for each BSS.

In addition to the load metric chosen, the load balancing techniques found in the literature can be classified according to the element in charge of taking association decisions by Yen et al. (2009).

2.1.2 Handover Decision Algorithms

Considerable work has been done in literature to determine the appropriate decision mechanism for Handover.

The popular work of Wang et al. (1999) also employed the use of a cost function that involved offered bandwidth, power consumption of MUs and financial cost. The performance of their work was evaluated with respect to the handover latency experienced. It is essential to mention that all algorithms that employ cost functions require manual inputs by the MU especially for each weight factor. This can become a bottle neck for both experienced and inexperienced MUs and could result in poor handover in the
event of any input mistake. It is necessary that VHO algorithms should be more independent and seamless with respect to the MU.

Chuanxiong et al. (2004) employed available bandwidth and delay as decision metrics in their work and measured the performance of their work against throughput and unnecessary handover rate experienced during handover.

Wei and Qing-An (2006) considered traffic load, RSS and variation of RSS. They made use of a cost function that normalized these metrics to enable comparison. The normalized RSS value in their cost function was only added to traffic load. This indicated the inability of their cost function to include other metrics like power consumption, financial cost as obtained in other popular cost functions. Their results were measured against the average blocking probability the algorithm offered.

Stevens-Navarro and Wong (2006) presented a platform for the analysis and comparison of the four most prominent decision algorithms in literature, that is, simple additive weighting (SAW), technique for order preference by similarity to ideal solution (TOPSIS), multiplicative exponent weighting (MEW) and the grey relational analysis (GRA). Due to the ability of all algorithms to accept different attributes, the authors selected parameters such as bandwidth, delay, jitter and bit error rate (BER) to conduct their comparisons. The results showed that MEW, SAW and TOPSIS provided similar performance to all four traffic classes while GRA provided a slightly higher bandwidth and lower delay for interactive and background traffic classes. This work is a commendable effort owing to the absence of standard performance comparison platforms in literature. The authors showed that different algorithms can be simulated, analyzed and their performances compared. It will be beneficial if further studies of this nature are carried out to accommodate more decision making parameters.
For example, Pramod and Saxena (2008) proposed a decision algorithm which they called dynamic decision model. This model adopts a three phase approach comprising priority phase, Normal phase and decision phase. The ‘priority phase’ discovers all available networks, filters out ineligible networks based on RSSI and velocity of the MU, and then assigns priorities to all eligible networks using the difference between RSSI and its threshold value, RSST. The network with the highest difference is assigned the priority. The ‘normal phase’ records the system information and MU preferences for offered bandwidth, power consumption, and network usage in terms of respective weight factors. It then calculates a cost function for each network. Finally, the ‘decision phase’ calculates a score function by multiplying the priority from priority phase and cost function from normal phase, for each candidate network. It then selects a network having the highest value of score function as “best” network to handover to and all current transmissions are transferred to the selected network if it is different from the current network. It is observed that this decision model is quite simple and “seemingly” easy to implement. However, the authors believe that it is dynamic because it considers the RSSI and the velocity of the MU. It should be noted here that most decision algorithms do consider RSSI and as such could also be considered dynamic to an extent; therefore, using RSSI here does not really make it more dynamic than others. This work also claims to consider the velocity of an MT in handover decision; it however did not say how this metric should be measured. Furthermore, the authors compared their results against a standard decision model (SDM). They argue that SDMs do not use RSS and velocity in making decisions. It is, however, not clear how the authors determined that SDMs do not use RSS and velocity in making decisions.

It will be helpful to know how the standard decision model was developed and what metrics are supported by it. SuKyoung et al. (2008) tried
to highlight the metrics, best suited for the VHO decision phase. They proposed a generalized VHO algorithm that optimizes a cost function. The cost function included the battery lifetime of an MT and the load balancing between APs. The authors referenced certain metrics such as, “network latency, congestion, battery power and service type” that could be considered in future algorithms. It is then obvious that no algorithm could be said to be absolutely robust until every single minute metric for VHO has been considered. Metrics used in this paper were both dynamic and static in nature making the algorithm relatively good, except for the high computation overhead involved in the cost function and load balancing process.

Ezil and Srivatsa (2008) used a cross layer approach to decision making and employed the following metrics: connection status, RSS, speed of MU and QoS requirement of certain applications. They estimated the performance of their work against the offered throughput and handover delay experienced by each application.

2.1.3 Handover Implementation Algorithm

With the growing need for mobility, handover became a hot topic in the scientific literature in the past few years. Most of the proposed solutions are able to shrink the handover latency to an acceptable value, but they usually give up a bit of security for the sake of fast handover, or they require special features not present in today's average networks.

The Inter Access Point Protocol (IAPP) (IEEE Std. 802.11F, 2003) is another possible solution, proposed by the IEEE too. The protocol came to life as an IEEE recommended practice 802.11f in 2003 but was revoked in 2006. However the main idea behind its design is still considerable: instead of re-authentication, the current AP sends a so called security context (basically the PMK) to the new AP. The solution is weaker than pre-authentication since
it requires secure communication between the APs, and it is impossible to use for inter operator handovers where the authentication servers are not the same.

IEEE itself addressed the problem in (IEEE Std. 802.11i, 2004), where pre-authentication with key caching is proposed as the solution. The idea of the pre-authentication approach is to execute the entire authentication procedure between the MU and the new AP while the MU is still associated with the old AP (i.e., before the handover takes place). The MU and the new AP then cache the resulting Pair wise Master Key (PMK), and they run only the four-way handshake protocol when the handover actually takes place. This solution has its advantages: the authentication can take as much time as needed (within reasonable limits), and the solution does not depend on the used authentication mechanism. Moreover, pre-authentication is part of the WPA2 standard ensuring that most new APs and MUs will implement this feature. The downside is that pre-authentication requires link layer communication between the participating APs. This requirement is not so easy to meet if the APs are located in different networks (possibly controlled by different operators).

A pro-active key distribution mechanism is proposed by Arbaugh et al. (2004). The idea is to use a neighbor graph to determine which APs are possible targets of a handover, and to distribute keying material to those APs from the AA server. The scheme's only disadvantage is that the MU is required to use the PMK shared with the current AP to construct the new PMK meaning that if an adversary can somehow break an old PMK, then she will be able to follow the MUs move in the network and easily calculate the new PMKs in use.

Alimian and Aboba (2004) examined the possible latencies caused by the above described phases of the handover process (they also examined phases in upper layers that are not covered in this paper), and established a
problem space showing that it is physically possible to achieve seamless handover if the MU is moving with the velocity of a pedestrian. They also showed that the authentication phase is responsible for a large (if not the largest) part of the overall latency and it is, therefore, a good idea to speed up this phase.

Yet another approach to the same problem is proposed by Aura and Roe in Aura and Roe (2005). Their solution uses a very fast but somewhat weaker authentication algorithm during the handover that initiates a strong and potentially slow authentication protocol run (e.g., EAP-TLS) right after its successful execution. If the strong second protocol fails (meaning that he MU cheated in the first authentication), then the AP denies further access to the network. This allows unauthorized access to the network if the weak protocol is broken, but only for a few seconds as the strong protocol supposedly cannot be broken. This scheme does not require inter AP communication, it is easy to implement, and it can handle enter operator handovers too. On the other hand, the optimistic approach means at least some loss of security, which might be unacceptable for some business applications (it is unlikely that any major company would allow a possible one second access to its intranet).

The detailed description of the EAP protocol itself and its use in RADIUS environment is beyond the scope of this paper; good documentation of these can be found in Aboba et al. (2003), Aboba et al.(2004) and Stanley et al. (2005).

AP-SIM is described in Haverinen and Salowey (2006). Its authentication mechanism is based on the scheme used in GSM networks or subscriber authentication. GSM networks are handling millions of seamless handovers every day which makes EAP-IM a good candidate for a fast
handover mechanism in WiFi networks. But that is not able to provide a fast handover for video call.

GRE Farinacci et al. (2000), Ethernet bridging (IEEE Std. 802.1d, 2004) or Ethernet over IP could provide a solution to this problem, but the usage of these requires serious trust between the operators (typically, they must share their LANs), and special firmware on the APs. In addition, there are scenarios where necessary connectivity is not available to support "make before break" communications (IETF Working Group, 2007). In conclusion, pre-authentication is the prevailing solution when the handover happens between APs of the same network, but it is not suitable or not easily adoptable for inter-operator handovers.

2.2 NEED AND JUSTIFICATION OF THIS RESEARCH WORK

It is a generally known fact that the more the parameters considered when making any decision, the better the outcome. This also applies to MU association and Handover decision making. So many metrics need to be considered when making decision. Some of these metrics sometimes cannot be easily measured.

The increasing number of decision metrics in recent research findings continues to make the decision more of a complex task. However, algorithms need to become simple and yet robust enough to handle these metrics. Consideration must also be given to the adaptation of these decision metrics to existing infrastructure. This means that algorithms should not be too complex for practical implementation, which will impact on financial cost. At the same time, decision metrics should not be too few, as this makes the decision outcome less accurate and can result in poor performance.
Owing to these conflicting requirements, the necessity for a balance between extreme complexity and simplicity becomes obvious. We have identified from literature, certain metrics that can form a general basis for decision making. In considering these metrics, we classified them as follows:

**Dynamic metrics**

1) Available bandwidth
2) Throughput (data rate)
3) Received signal strength
4) Bit error rate

**Static metrics**

1) Financial access cost
2) Security features for example, authentication
3) NIC power consumption rate
4) Battery power status

Fair load balancing is achievable only over a large number of MUs with less number of APs. Difficulty is experienced if there is considerable variation in no of MUs associated with APs. These difficulties have necessitated us to develop a new association scheme, which gives a high level of throughput, bandwidth and low bit error rate. Hence, in this research these aspects have been taken into consideration and methodology based on association of MU with AP is developed and simulated.

### 2.3 OBJECTIVE OF THIS STUDY

1. To design a fair load balancing scheme for WLAN by access point selection scheme.
2. To design an efficient handover scheme for between access points for shortens the delay of authentication.

3. To develop and verify the throughput and bandwidth allocation of individual MUs.

2.4 SUMMARY

In this chapter, we have analyzed the most significant contribution published in the recent years pertaining to load balancing scheme and handover scheme in WLAN. However, it is clearly noticed that though the contributions are highly valuable and effective, there are more approaches available which are based on the existing methodologies. The current load balancing models fail to enable efficient use of AP bandwidth and throughput. Evidence indicates that the throughput and balance index of APs are unstable and the above form the foundation for the work presented here.