Throughput Management for CSMA/CA Networks: IEEE 802.11e Wireless LAN

Shafiqul Karim

Thesis submitted for the degree of
Doctor of Philosophy
in
Electrical and Electronic Engineering
at
The University of Adelaide
(Faculty of Engineering, Computer and Mathematical Sciences)

School of Electrical and Electronic Engineering

November 12, 2012
Chapter 1

Introduction

Over the last decade the popularity of deploying IEEE 802.11 based WLANs [3] has increased rapidly at an unprecedented rate. WLANs just as their wired LANs counterpart provide end-users access to specific network services. For example, the ability to run applications that facilitate transmitting/receiving of data within or outside of the WLAN domain.

WLANs were initially viewed as a temporary low cost, reduced performance alternative to traditional wired Ethernet LANs. However over time, the advancements made in WLAN technology have come to a point where WLANs are now a serious and valid infrastructure install alternative to wired Ethernet LANs. As a mark of both the popularity and improvement in performance, WLAN deployments can be found in residential, commercial, industrial and public areas.

WLANs are fundamentally centered around the use of a shared transmission medium. As a result this requires some form of management and coordination to share access to the transmission medium. The IEEE 802.11 [3] Medium Access Control (MAC) layer standard, originally described two alternative medium access mechanisms. A centralized approach that provides contention-free medium access known as the Point Coordination Function (PCF), and a decentralized distributed approach that provides contention-based medium access known as the Distributed Coordination Function (DCF). Currently the most widely implemented and used IEEE 802.11 medium access mechanism is DCF. DCF is a relatively simplistic mechanism that provides only best-effort service for data transmission.

With the number of users and deployments of WLANs rising, this undoubtedly leads to an increase in network bandwidth demand. As a result, we have observed the progressive introduction of WLAN devices that support higher bandwidth data transmissions. The original IEEE 802.11 Physical (PHY) Layer [4] was only capable of supporting a maximum 2Mbps net data transmission rate within the WLAN. Additional revisions to
the IEEE 802.11 PHY layer configuration have been standardized that further increase the supported maximum net data transmission rate in a WLAN. The IEEE 802.11b PHY layer [5] supports a maximum data transmission rate up to 11Mbps, IEEE 802.11a [6] and IEEE 802.11g [7] both provide support for rates up to 54Mbps, and the most recently introduced IEEE 802.11n [8] has been shown to be capable of providing rates up to 600Mbps.

As a consequence of the increasing bandwidth, end-users are able to push more data through the WLAN and more importantly deploy a wider variety of applications. Applications such as those that provide real-time services, for example Voice over IP (VoIP), teleconferencing and live high quality video streaming. Also too, non real-time applications such as web browsing, individual file transfers or Peer-to-Peer (P2P) file sharing have benefited immensely from the increased bandwidth provided by newer generation WLANs. However, an oversight still exists within the infrastructure of IEEE 802.11 WLANs.

Generally speaking, in an over-provisioned network, a best-effort data transmission scheme is sufficient in meeting Quality of Service (QoS) requirements for deployed applications (an explanation of QoS is provided in Section 1.1). However, as network utilization rises, it is to be expected that a best-effort approach will gradually become unable to ensure that QoS requirements are fulfilled. Therefore based on this understanding, it is expected that the continued use of the simple best-effort DCF scheme may not always be able to ensure that QoS requirements are fulfilled within a WLAN.

In light of this shortcoming in the pre-existing IEEE 802.11 MAC layer, an updated IEEE 802.11 MAC layer with QoS enhancements has been proposed and standardized. Referred to as IEEE 802.11e [1], it introduces a new medium access scheme known as the Hybrid Coordination Function (HCF). HCF consists of two medium access mechanisms, Enhanced Distributed Channel Access (EDCA) and HCF Controlled Channel Access (HCCA).

The operation of EDCA is a direct extension of DCF (distributed and contention-based), with the key enhancement of supporting service differentiation across 4 different Access Categories (voice, video, best-effort and background). Each Access Category (AC) has specific tunable parameters associated to it, which affect the level of probabilistic service differentiation it obtains against the other access categories. While HCCA is a direct extension of PCF (centralized and contention-free), and is aimed at providing strict QoS guarantees for traffic flows within the WLAN.

Unlike IEEE 802.11 DCF and PCF, which both have strict clearly defined implementation and configuration information presented within their related standard documentation, the standard documentation for IEEE 802.11e EDCA and HCCA only describes the key framework structure for both mechanisms. Therefore the implementation and
configuration of components within the EDCA and HCCA framework are left open to interpretation.

In this thesis, we focus specifically on the IEEE 802.11e HCF and its access mechanism EDCA. We investigate and develop a complete implementation and configuration proposal for the EDCA mechanism. Our proposal aims to provide the ability to manage throughput across all range of network conditions, while retaining optimal performance and being fully compliant with the designated IEEE 802.11e EDCA framework. Given that throughput management is the primary goal, this narrows the focus specifically to controlling the behaviour of ACs that benefit the most from this objective. Thus controlling the behaviour of access categories AC0 and AC1 (background and best effort) will be discussed in this thesis.

1.1 Quality of Service (QoS)

The term Quality of Service (QoS) essentially defines a range of quality requirements placed on the overall performance provided by one or more components within a network [9, 10]. The actual quality requirement is a subjective matter, since different end-users will have different views on what constitutes quality to them.

Even so, network service providers try to utilize several key network performance measures in order to infer what the perceived level of QoS end-users are receiving. Some example of the network performance measures used are, delay, jitter, throughput and data loss rate. Besides these network performance measures, network service providers may also take into account other things in gauging overall QoS satisfaction. Such as the network service uptime and downtime, security and reliability.

QoS, itself, is directly tied to what service requests are being made on the network by end-users. Therefore each type of application executed by end-users will each have its own specific QoS requirements. Applications that provide real-time services such as Voice over IP (VoIP), teleconferencing and live video streaming, place strict QoS requirements on network performance measures in delay, jitter, data loss rate and throughput. While, applications that provide non real-time services such as, file transfer, email and web browsing, all have much more flexible QoS requirements for the same network performance measurement parameters.

As a result, provisioning of QoS in networks can be characterized as being either parameterized or prioritized QoS [11, 12, 13]. Parameterized QoS ensures that quantitative QoS network-related metrics, such as throughput, delay and jitter are always maintained to a required value. Prioritized QoS generally does not imply that QoS network-related quantitative metrics are maintained to a required value, instead the goal is to provide
1.2. MAJOR RESEARCH CONTRIBUTIONS

service differentiation. Access to network resources is dictated based on the priority associated to different types of traffic. Generally, traffic classified as high priority is expected to gain access to network resources before low priority traffic.

Therefore, the ability to have QoS control mechanisms at hand is a valuable asset for network service providers. Network performance improvements may be achieved leading to for example, faster web page loading, faster file transfer rates, better real-time audio and video services, and low latency online gaming.

1.2 Major Research Contributions

This section outlines the key research contributions introduced in this thesis.

- We describe a simple relationship that can be used to determine what the resultant throughput proportion allocation sharing will be amongst the ACs, provided the WLAN is under saturation load and the collision probability is low. This relationship takes into account the specific configuration used for the tunable parameters of EDCA, and the number of saturated stations for each AC.

- Using the relationship identified above, in conjunction with an analytical model [2], we devised a unique simple and fast search mechanism. The search mechanism allows us to identify the specific EDCA parameter configuration that will maximize overall throughput performance within the WLAN, while achieving a certain throughput proportion allocation between ACs.

- Based on this capability, we designed and developed a centralized EDCA configuration based around a control mechanism which resides within the AP. Control Scheme-1 introduced the ability of allowing a WLAN administrator to specify what target throughput proportion allocation each AC should obtain within a saturated WLAN environment (independent of saturated station count).

  - The developed Control Scheme-1 retains a fully autonomous operation, other than requiring an input value for the target throughput proportion allocation for each AC. It is able to maintain the required proportion within the WLAN seamlessly for any known number of saturated stations.
  
  - Control Scheme-1 utilizes a simple and innovative parameter switching technique to facilitate the ability to administer a wider range of through proportion allocations for each AC.
- Control Scheme-1 had been tested and shown to be able to maintain the required proportion allocations between ACs and still maximize throughput performance even when all AC traffic was TCP-based.

- We developed a simple and effective means of guaranteeing the successful transmission of TCP acknowledgment packets through the use of the HCCA medium access mechanism. This feature allowed for more TCP flows to effectively transmit data together and most importantly it stabilized the TCP throughput performance within the WLAN.

- In order to overcome the inherent throughput sharing anomaly that exists between Downlink (DL) and Uplink (UL), we identified an innovative solution by allowing the AP to have its own set of configurable EDCA parameters. As a result, a new relationship is used to determine the throughput proportion allocation between DL and UL, based on the configured EDCA parameters for the AP and the wireless client stations. Thus there were now three throughput proportion allocations that could be evaluated, namely overall DL to UL, internal DL AC-based and internal UL AC-based.

- A second search mechanism was defined, which still retained the features of being simple and fast. The search process identifies the EDCA parameter settings for the AP and stations that will maximize throughput performance, while achieving a certain overall DL to UL, internal DL AC-based and internal UL AC-based throughput proportion allocation.

- The revised control mechanism, Control Scheme-2 is shown to allow the ability to now ensure that throughput performance was still maximized while maintaining the three required target throughput proportion allocations within a WLAN which is saturated for DL (AP) and UL (a known number of stations for each AC).

- The continued requirement of saturation load conditions within the DL and UL in a WLAN was clearly an unrealistic assumption. As a first pass towards handling all possible non-saturation load conditions within the WLAN, we assumed only the DL could be non-saturated. Based on this initial assumption, Control Scheme-2 was modified so that it could now operate effectively under any given traffic load scenario for DL. The modified Control Scheme-2 still maintained the goal of maximizing throughput performance and whenever possible enforce the required target throughput proportion allocations or subsequently chose a new throughput proportion allocation that was deemed acceptable.
• Identifying the traffic load condition for the DL was made easy given the fact that the control mechanism would reside within the AP. However in the case of the UL, a similar approach was infeasible. As a result we developed an innovative load detection mechanism that could be used to identify when the UL became non-saturated and to what degree it was unsaturated. The key innovation of the UL traffic load detection mechanism, was its fast execution and completely non-intrusive/passive operation. The detection mechanism is executed solely from within the AP, and thus no modifications were required for stations. Furthermore, its operation required no form of cross layer interaction.

• Finally, using the UL traffic load detection mechanism the modified Control Scheme-2 was further modified one last time to now also support operation in scenarios where the UL is non-saturated. Thus the completed control mechanism, Control Scheme-3 is now fully capable of handling various scenarios where different traffic load conditions exist in both the DL and UL. In all possible scenarios, Control Scheme-3 is still designed to try and achieve maximum throughput performance and whenever possible maintain the required target throughput proportion allocations. Furthermore all of these features are achieved when AC traffic is TCP-based and with the knowledge of the number of active stations present within the WLAN.

1.3 Thesis Structure

In Chapter 2, we begin by defining the two common forms of WLAN deployments, followed by a description of the currently available set of IEEE 802.11 WLAN standards. We mention briefly the different PHY layer implementations specified in the IEEE 802.11 standards. The MAC layer specifications defined within the standards are described in greater detail, and we specifically identify if, how and what QoS provisions are offered within the MAC layer. Through the provided descriptions, the IEEE 802.11e MAC QoS enhancement standard is identified as offering direct QoS provision support within a WLAN, in particular the IEEE 802.11e EDCA medium access mechanism. As a result we then highlight some of the relevant work that deals directly with the IEEE 802.11e EDCA medium access mechanism. Finally, we discuss the key motivations of this thesis in addressing some of the identifiable open areas of research related to IEEE 802.11 EDCA.

In Chapter 3, we start the process of addressing the research motivations discussed earlier in Chapter 2. We mention two primary objectives that need to be addressed, maximizing throughput performance and providing throughput management services within the WLAN. Based on these objectives, we mention specific related work in the literature and point out their critical shortcomings. We then begin the task of determining how
to design and develop a control mechanism capable of achieving the required objectives, while making sure it does not suffer from any of the highlighted shortcomings of the work present in the literature. Once the control mechanism, Control Scheme-1 has been finalized, we test its performance through the use of various simulation test case scenarios.

In Chapter 4, we highlight an inherent problem that exists within the legacy IEEE 802.11 DCF medium access mechanism. This particular problem is the throughput sharing anomaly that occurs between Downlink (DL) and Uplink (UL). We point out that the EDCA-based control scheme developed in Chapter 3 unfortunately, also suffers from this problem. Based on the discussed anomaly, we highlight some of the related work that attempts to directly solve this problem. However as the discussions in Chapter 4 reveal, the solutions presented within the related work do not achieve our defined objective goals. Thus, we propose our own innovative solution, which only requires minor modifications to Control Scheme-1 defined in Chapter 3. Using various simulation test case scenarios we aim to show the revised control, Control Scheme-2, is able to overcome the described anomaly. In addition, Control Scheme-2 is expected to continue to satisfy the objectives of maximizing throughput performance and introducing throughput management capabilities within the WLAN.

In Chapter 5, we discuss the unrealistic assumption that has been used in the construction of Control Scheme-1 and Control Scheme-2 in Chapters 3 and 4. This assumption is the need for the WLAN to be continually saturated. Therefore, the objectives discussed earlier were only achievable in a saturated WLAN. In light of this, a reassessment of the required objectives is discussed where the new assumption used is that only DL traffic load can be unsaturated. In order to achieve the updated objectives, three components are highlighted in Chapter 5 as being vital additions to the Control Scheme-2. The importance of each component is discussed in detail through the use of simulation test results.

In Chapter 6, we begin by pointing out that the configuration used for the three components described previously in Chapter 5, are insufficient alone in achieving the revised objectives. As a result, we propose in Chapter 6 critical modifications to certain components in order to truly achieve the revised objectives. In addition to the critical modifications, we propose some enhancements to improve the control mechanism’s overall performance capabilities. With the implementation of both modifications and enhancements to the pre-existing Control Scheme-2, simulation results are used to discuss and show it achieving the revised objectives.

In Chapter 7, the final assumption of requiring continually saturated UL traffic load in the WLAN is removed. However as discussed in the chapter, a direct translation of the approach used to handle non-saturation in DL cannot be applied to the UL. In
1.3. **THESIS STRUCTURE**

In order to develop a complete control mechanism that is able to handle non-saturation load conditions in the UL, we need a means for determining the actual load conditions within the UL. As a result, we define the specific requirements needed in an effective UL load detection mechanism. Based on these requirements, a review is carried out to identify any possible solutions currently available in literature. Unfortunately, no viable solution was identified as being able to meet our requirements. Therefore the main purpose of Chapter 7, is the discussion and description of our proposed UL load detection mechanism.

In Chapter 8, we propose the last required modifications needed to *Control Scheme-2* in order for it to now also handle non-saturation load within the UL. Using the UL load detection mechanism from Chapter 7, an extensive and in-depth description is provided for all the specific modifications required in order to create the final complete control mechanism *Control Scheme-3*. In all simulation tests performed, the DL is configured as being continually saturated, while different non-saturation UL traffic load scenarios are used in each individual test carried out.

In Chapter 9, we discuss and describe the final testing and verification procedure used on *Control Scheme-3*. The previous assumptions of requiring continually saturated DL and/or UL are omitted. As a result, in the extensive suite of tests performed in simulation, realistic network scenarios are configured whereby both DL and UL traffic load conditions can vary independently. The data collected from all the simulations are used in compiling statistical results, which we use to verify with confidence the effectiveness of the complete control mechanism *Control Scheme-3*. Finally, as a last highlight of the beneficial use of our complete control mechanism, we compare its performance to that of the reference IEEE 802.11e EDCA implementation.

In Chapter 10, we conclude this thesis, by firstly presenting a summary and overview of the work accomplished from earlier chapters. Finally we finish by providing an outline of some of the possible avenues for potential future research directly related to our work.
Chapter 2

Background

IEEE 802.11 are the primary standards used that defines all aspects of the implementation and configuration of devices designed for use in a wireless local area network (WLAN) environment. In this chapter, we describe the specific components of IEEE 802.11:

- Supported WLAN Architecture
- Physical (PHY) layer
- Medium Access Control (MAC) layer

Most importantly, we focus on what QoS provisions are provided within the IEEE 802.11 standards. Based on this information, we then highlight briefly, the various types of related work. Finally, we discuss the motivation and the open areas of research that we aim to address in this thesis.

2.1 IEEE 802.11: WLAN Architecture

The building block of an IEEE 802.11 WLAN is referred to as the Basic Service Set (BSS), this simply refers to a group of wireless stations that communicate with one another. The area encompassed by a BSS is referred to as the Basic Service Area (BSA), this area is essentially defined by the propagation characteristics of the wireless transmission medium. Wireless stations within a given basic service area can communicate with other wireless stations that are members of the same BSS. IEEE 802.11 WLAN supports two forms of BSS topology, independent BSS (IBSS) or infrastructure BSS.

IBSS:

An IBSS is a WLAN that is deployed without using any fixed infrastructure that assists in communication. All wireless stations that are part of the same BSS will communicate
directly with each other and thus must be within direct transmission range. Typically, an IBSS is a temporary short-lived network that consists of a small number of wireless stations. A common term generally used for IBSS networks is *ad-hoc* networks.

**Infrastructure BSS:**

Infrastructure BSS are a *WLAN* characterized by the use of specialized wireless station referred to as an access point (*AP*). The *AP* is used for all communications within an infrastructure BSS. This means for one wireless station to communicate with another, it has to first transmit the data to the *AP*. The *AP* after receiving the data will then relay this onto the destination wireless station. Through this process it is clear that the originating wireless station does not have to be within direct transmission range of the destination wireless station. As a result of this feature, network coverage is simply dictated by the communication range that exists between the *AP* and individual wireless stations.

Another feature of an infrastructure BSS, is the need for wireless stations to *associate* with an *AP*. Wireless stations always initiate the association process, and it is up to the *AP* to choose if it grants or denies access based on the specific type of association request exchange used. IEEE 802.11 enforces the rule that each wireless station can not be associated with more than one *AP* at the same time. However, there are no strict regulations in place for the reverse case of the number of wireless stations that can be associated with an *AP*.

### 2.2 IEEE 802.11: Physical (PHY) Layer

The IEEE 802.11 *WLAN* standard specifies a number of different PHY layer configurations that can be used within *WLAN* ready devices. The PHY layer itself, covers the process of how digital information is physically transported over the wireless transmission medium. For example, the PHY layer defines what wireless frequency band to transmit information in and what signal encoding process should be used. Ultimately, the PHY layer imposes a hard limit on the achievable net data transmission rate within the *WLAN*.

The original IEEE 802.11 standard defined in 1997 [4], specified two net data transmission rates of 1Mbit/s and 2Mbit/s. Data could be transmitted using either Infrared (IR) signals or over the Industrial Scientific Medical frequency band of 2.4GHz. Within the 2.4GHz band two signal modulation techniques were used, either *frequency-hopping spread spectrum* (FHSS) or *direct-sequence spread spectrum* (DSSS). The IEEE 802.11 PHY layer implementation has since been rendered obsolete with the release of newer IEEE 802.11 PHY layer implementations.
Currently, four PHY layer implementations have been standardized and are in use today, IEEE 802.11a, IEEE 802.11b, IEEE 802.11g and IEEE 802.11n [6, 5, 7, 8].

IEEE 802.11a:
The IEEE 802.11a standard [6] specifies operation within the 5GHz frequency band. Each WLAN transmits wirelessly across an individual channel within the 5GHz band. IEEE 802.11a employs a wireless channel bandwidth of 20MHz.

The standard specifies the use of Orthogonal Frequency Division Multiplexing (OFDM) as the method of encoding digital data for transmission over a wireless medium. The maximum achievable net data rate, inclusive of frame overhead and error correction padding is 54Mbit/s.

IEEE 802.11b:
The IEEE 802.11b standard [5] specifies operation within the 2.4GHz frequency band. As with IEEE 802.11a, IEEE 802.11b also employs 20MHz bandwidth channels. Transmission of data over the wireless medium employs the signal modulation technique known as Direct-sequence Spread Spectrum (DSSS). The maximum achievable net data rate is specified at 11Mbit/s.

IEEE 802.11g:
The IEEE 802.11g standard [7] just as IEEE 802.11b, also operates within the frequency band of 2.4GHz. Once again, 20MHz bandwidth channels are employed. It supports backwards compatibility with IEEE 802.11b and thus is also able to utilize DSSS signal modulation. However, IEEE 802.11g also includes the use of the same OFDM signal modulation as IEEE 802.11a. This allows for the same high data rate of 54Mbit/s to be achieved.

IEEE 802.11n:
The IEEE 802.11n standard [8] provides a significant increase in the maximum data rate capability for WLANs over IEEE 802.11a/b/g. The increased data rate is achieved through the use of Multiple-input Multiple-output (MIMO) technology and the use of a wider bandwidth channel. IEEE 802.11n supports both operational frequency bands of 2.4GHz and 5GHz, and specifies the use of OFDM signal modulation. IEEE 802.11n also continues to maintain backwards compatibility with the previous IEEE 802.11a/b/g standards.
MIMO technology essentially refers to the technique of utilizing multiple antennas to transmit and receive data. A process referred to as Spatial Division Multiplexing (SDM) is used, where data that needs to be transmitted, is broken up into multiple data streams and each stream is then transmitted and received via a particular transmit and receive antenna pair. IEEE 802.11n specifies the use of 40MHz channels, doubling the channel bandwidth from 20MHz used in IEEE 802.11a/b/g.

The IEEE 802.11n standard defines a number of different possible configurations of coupling both MIMO technology and a wider bandwidth channel. The most aggressive configuration is shown to be theoretically capable of achieving a maximum net data rate of 600Mbit/s.

2.3 IEEE 802.11 Medium Access Control (MAC)

IEEE 802.11a/b/g/n each define a specific type of PHY layer implementation that can be used within wireless devices. However, in order to coordinate and manage wireless data transmissions effectively, an additional layer is required on top of the PHY layer known as the MAC layer.

Currently the most common and widely implemented MAC layer configuration is still based on the legacy IEEE 802.11 MAC standard that was originally ratified in 1999 [3]. The legacy IEEE 802.11 MAC standard, originally defined two different ways of effectively managing wireless transmissions amongst wireless stations. These two are the Distributed Coordination Function (DCF) and the Point Coordination Function (PCF).

Only DCF has seen widespread implementation within wireless devices. This has mainly been due to its implementation simplicity and ability to operate in IBSS and infrastructure BSS scenarios. While in the case of PCF, its implementation complexity and restricted use only in infrastructure BSS scenarios inevitably forced it to be relegated to just an accepted MAC layer configuration on paper.

2.3.1 IEEE 802.11: Distributed Coordination Function (DCF)

Wireless data transfer is effectively a broadcast medium, where transmitted data is generally received by all members of the network. As part of the development of IEEE 802.11 WLANs, a key assumption taken was that, being able to transmit and listen to the wireless medium was not possible at the same time. As a result, a station within the WLAN must share its time on the transmission medium with other stations, just as the case with wired Ethernet [14].

Therefore just as with Ethernet, IEEE 802.11 DCF utilizes a Carrier Sense Multiple Access (CSMA) mechanism to facilitate data transmissions wirelessly. Where CSMA
generally refers to a station having to sense when the medium is idle before commencing transmission. In the case of wired Ethernet, it is possible to transmit and listen to the medium at the same time, therefore collisions can be detected directly. As a result, wired Ethernet utilizes CSMA with Collision Detection (CSMA/CD) to perform data transmissions.

Given the difficulty of detecting collisions, IEEE 802.11 DCF requires that all data frames (with the exception of broadcast type frames) be acknowledged by the receiving wireless station. The receiving wireless station is required to send an acknowledgment (ACK) frame to the transmitting wireless station. Thus if the ACK frame is not received by the transmitting wireless station it assumes a collision has taken place.

The occurrences of collisions within a wireless medium inevitably leads to reduced overall network performance, therefore reducing the number of collisions is critical. As a result, IEEE 802.11 DCF employs a contention-based mechanism in order to reduce and avoid the possibility of collisions. This particular type of approach is referred to as CSMA with Collision Avoidance (CSMA/CA).

IEEE 802.11 DCF as the name implies, is a completely distributed system with no centralized control and thus operates independently within each wireless station and AP. As a result each instance of the DCF mechanism within a wireless station and AP determines for itself, when it can access the wireless medium to transmit data. DCF provides two contention-based modes of operation, the first being a basic CSMA/CA approach and the latter the additional/optional use of Request to Send (RTS) and Clear to Send (CTS) frame exchanges.

**DCF: Basic CSMA/CA**

Within a wireless station (or AP), data packets from the higher layers arrive into a queue within the MAC layer. If a packet arrives with the queue empty, and the physical carrier sensing determines the wireless medium has been idle for an interval of time longer than a distributed inter-frame space (DIFS). The DCF mechanism within the related station adds the necessary MAC framing structure to the data packet and then passes down the data frame to the PHY layer for immediate transmission.

Upon successfully receiving the data frame, the destination station waits for a period of time to elapse referred to as a short inter-frame space (SIFS which is less than a DIFS) and then transmits an ACK frame back to the source station to indicate a successful data frame transmission.

With a data frame transmission exchange taking place, it is important other stations know that they need to defer their own transmission attempt. Aside from the physical carrier sensing, the DCF mechanism also utilizes a virtual carrier sensing process.
Within the MAC header of the transmitted data frame is a piece of information referred to as the duration field entry. The duration field reports the amount of time the wireless medium will be busy transmitting the data frame and the corresponding ACK frame exchange. Each non-transmitting station uses the value of the duration field entry to adjust an internal DCF timer parameter referred to as the network allocation vector (NAV).

The NAV timer serves the direct purpose of allowing DCF to perform its required virtual carrier sensing actions. Therefore as the NAV timer counts down, while it is non-zero the virtual carrier sensing indicates the medium is busy, and when it is zero, the virtual carrier sensing indicates the medium is idle. The complete process described is shown illustratively in Figure 2.3.1.

Once the data and ACK frame exchange is completed and the NAV value indicates zero, stations are required to once again sense the medium being idle for a DIFS before attempting a transmission. However if a station had previously deferred their transmission attempt due to activity on the medium they are no longer allowed to attempt an immediate transmission, likewise the same for the station that was previously transmitting. Instead now, stations are required to start the additional task of executing the DCF back-off process.

The DCF back-off process requires stations to sense the medium being idle for an additional random period of time after the elapsed DIFS interval before attempting a transmission. Thus the station that selects the smallest random period of time will get the first opportunity to transmit. The random period of time is selected from an array within DCF referred to as the contention window (CW).
The CW is divided into a number of slots, with each slot representing the same fixed period of time (slot time). Thus stations randomly select a number of slots to specify the additional random period of time. Stations effectively count down the elapsed number of slots the medium has been idle, and once it reaches zero they can commence transmission. If during the count down process the medium becomes busy, stations will pause the count down process and only restart it from where it left off, after the medium is idle for a DIFS interval.

The effective number of slots within the CW is governed by two DCF parameters, $CW_{\text{min}} = 31$ defines the smallest size for CW, while $CW_{\text{max}} = 1023$ defines the largest size for CW. When a station first starts its back-off process, it selects a uniformly distributed random number of slots using $CW_0 = CW_{\text{min}}$. Whenever a frame transmission fails (i.e. a collision), the stations involved will attempt a retransmission by restarting a new back-off process, however this time using a CW that is twice the size of its previous value, specifically $CW_i = 2 \times CW_{i-1} + 1$ (where, ‘$i$’ refers to the retransmission attempt count).

Therefore with each consecutive retransmission attempt, the size of CW is exponentially increased up to largest valid value specified by $CW_{\text{max}}$. The maximum number of actual retransmission attempts is set by the value of the DCF parameter referred to as the short retry limit. The default value as specified by the IEEE 802.11 standard is 7, with the count beginning from 0. Once a successful transmission is completed the short retry count is reset to 0, and likewise the CW is also reset to $CW_{\text{min}}$.

Aside from the above described transmission process of sending complete packets encapsulated within a MAC frame, DCF also supports the feature of fragmented frame transmissions. In essence, if a higher layer packet is too large to fit within a single MAC frame payload, DCF will automatically divide it into smaller fragments which can then be transmitted via multiple frames. Fragmentation can also be executed if higher layer packet size exceeds a user-defined DCF fragmentation threshold parameter value ($\text{Frag.\_threshold}$), which ranges between 256-2312 bytes [3].

All fragments that comprise a complete packet, are transmitted during a fragmentation burst sequence, this process is illustratively shown in Figure 2.3.2.

In Figure 2.3.2, fragment frames and their respective ACK frames are separated by the SIFS, as a result the transmitting station retains access to the medium during the whole fragmentation burst sequence. Each fragment sets the NAV to the hold the medium until the end of the ACK frame transmission of the next fragment frame, thus Fragment 0 sets the NAV to hold the medium until ACK 1. Likewise each ACK frame will also set the NAV to hold the medium until the next ACK frame. Note that the last ACK frame in the burst sequence is not required to specify a value for NAV.

In the case of an error during the transmission of a particular fragment (aside from the
initial fragment), a station does not have to perform the exponential back-off procedure, instead it can retransmit the lost fragment immediately after a SIFS interval following the time alloted to the unreceived ACK frame. As a result, all previously successfully transmitted fragments need not be retransmitted.

Stations are required to keep a retry count for each fragment, if all retransmissions fail for a particular fragment, stations will discard all fragment frames. This effectively means, dropping the packet and starting a new back-off process for the next packet within its MAC queue.

**DCF: CSMA/CA with RTS/CTS**

Within a WLAN, there can be cases where certain wireless stations may not be able to communicate directly with every other wireless stations within the WLAN. This creates the hidden node problem scenario, an example of this is depicted in Figure 2.3.3.

In Figure 2.3.3, node 2 is able to communicate with both nodes 1 and 3, however nodes 1 and 3 are unable to communicate directly with each other. As a result, from the perspective of node 1 (node 3), node 3 (node 1) is a hidden node. Therefore it is quite possible for both node 1 and node 3 to transmit simultaneously, which would then render node 2 unable to receive any valid data. Furthermore, with nodes 1 and 3 being outside of each others coverage range, neither would have any indication of a transmission error occurring since the collision is localized to node 2.

In an attempt to overcome this problem, IEEE 802.11 DCF allows the option of using the Request to Send (RTS) and Clear to Send (CTS) frame exchange process. A simplified RTS and CTS frame exchange process is illustrated in Figure 2.3.4.

In Figure 2.3.4, node 1 has a data frame to transmit to node 2; the first step in the process is to transmit the RTS frame. The transmission of the RTS frame serves
the purpose of both reserving the transmission medium and subsequently notifying other
nodes within range, that node 1 wants to transmit a data frame.

Now with node 2 receiving the RTS, it responds by transmitting a CTS frame. Sim-
ilarly again the CTS frame serves the purpose of both acknowledging node 1 that it can
begin data transmission, and also notifying surrounding nodes that data transmission is
about to begin. Therefore, we can see that by node 3 observing the CTS frame, the issue
of the hidden node and simultaneous transmissions is avoided. Thus with the completion
of the RTS/CTS exchange data transmission can then take place.

In Figure 2.3.5, we illustrate the proper atomic timing operation sequence of the
RTS/CTS exchange. As with the basic CSMA/CA case, if a higher layer packet arrives with the MAC queue empty and subsequently the wireless medium has been idle for at least a \textit{DIFS}, then a RTS frame is transmitted straight away. The RTS frame sets the NAV duration for the complete period of time required to transmit the RTS itself, CTS frame and the data frame plus its ACK frame. Once the RTS frame is received by the destination station, it waits for a \textit{SIFS} to elapse and then immediately sends a CTS frame. With the CTS frame transmitted, the NAV duration is set for the period of time required to transmit the CTS frame itself and the data frame plus its ACK frame. Once the source station receives the CTS frame, it waits for a \textit{SIFS} to elapse and then starts transmission of the data frame. Finally, the NAV duration set by the data frame will be the period of time to transmit the data frame itself plus its ACK frame.

![Diagram of DCF CSMA/CA using RTS/CTS](image)

Figure 2.3.5: DCF CSMA/CA using RTS/CTS

The RTS/CTS exchange in the scenario where fragmentation is involved is depicted in Figure 2.3.6. RTS/CTS frames are only exchanged at the start of a fragmentation burst sequence. The NAV durations specified by both RTS and CTS only accounts for the transmission time for the first fragment frame and its ACK. As a result as described previously, each fragment frame sets a NAV duration that accounts for the next fragment to be transmitted.

As mentioned before, RTS/CTS is an optional feature within DCF, its execution or non-execution is dictated by the value assigned to the DCF RTS threshold parameter (\textit{RTS\_threshold}). The range of \textit{RTS\_threshold} is 0-2347 bytes, if a higher layer packet size exceeds the value set to \textit{RTS\_threshold}, RTS/CTS will be used. Furthermore when using RTS/CTS, instead of the \textit{short retry limit} retransmission condition, another retransmission condition referred to as the \textit{long retry limit} is used, by default this is set to 4 \text{[3]}. 


2.3.2 IEEE 802.11: Point Coordination Function (PCF)

The Point Coordination Function (PCF) defined within the legacy IEEE 802.11 standard provides an alternative method of accessing the wireless medium. PCF is an optional feature within the 802.11 specifications, its implementation is not mandatory like DCF. However, interoperability is maintained between devices that only implement DCF and those that incorporate both DCF and PCF.

DCF is a contention-based mechanism that manages the transmissions of data on the wireless medium. If contention-free management and transmission of data is required, then PCF may be used. An important thing to note is, continuous full-time contention-free service is not provided. The wireless medium is expected to have periods of contention-free service while alternating with standard contention-based service provided by DCF.

Contention-free service provided by the PCF is a centralized medium access control mechanism. Medium access is delegated through the used of a specialized PCF entity referred to as the point coordinator. The point coordinator (PC) role is expected to be implemented within an AP. Wireless stations associated with the AP are only allowed to transmit when instructed to do so via the PC. As a result, the need for a dedicated PC means PCF is only suitable for use in infrastructure BSS network scenarios.

PCF: Operational Behaviour

Time on the wireless medium is divided into a contention-free period (CFP) and the contention period (CP), with access in CFP controlled by PCF and access in CP controlled by DCF. A CP is required to be at least long enough such that a maximum size data frame and its related ACK frame can be transmitted. It is expected that the medium...
will consist of alternating CFP and CP, which will usually repeat at regular intervals referred to as a contention-free repetition interval. Figure 2.3.7, is an illustration of a contention-free repetition interval and the data transmission process used by PCF within a CFP.

![Contention-free Repetition Interval](image)

**Figure 2.3.7: Data Transmission Process using PCF**

The start of a CFP is designated by the transmission of a beacon frame via the AP. The maximum duration of the CFP is announced through setting the beacon frame field \( CFPMaxDuration \). Wireless stations receiving the Beacon will then set their NAV duration according to this information, and thus effectively disable DCF-based access to the wireless medium. In order to further guarantee the PC gains complete control of the wireless medium, transmissions within a CFP are all separated by an SIFS and an additional interval referred to as the PCF inter-frame space (PIFS). The PIFS interval is specifically set to a duration that is shorter than the DCF-based DIFS interval.

As the AP gains control of the wireless medium, the PC begins to poll all the wireless stations on its polling list to allow them to transmit data. The polling list designates the list of wireless stations that wish to be able to transmit data frames during the CFP. Wireless stations get onto the polling list through the initial association request process with an AP. The association request process essentially allows the AP to determine if a wireless station is capable of responding to polls during a CFP.

Wireless stations may only transmit once they receive a polling frame (CF-Poll) from the PC, each CF-Poll only allows a single data frame transmission. In Figure 2.3.7, the PC sends a CF-Poll to station 1, station 1 has a data frame to send to station 3 and it
transmits the data frame along with an acknowledgment for receiving the CF-Poll (CF-ACK). After receiving the data frame from station 1, the PC then transmits a CF-Poll to station 2 after an SIFS. During this process at the same time, it also includes a CF-ACK receipt for station 1 that it has successfully received the data frame. Since station 2 has no data to transmit, the PC after waiting a PIFS moves on to poll station 3. The PC sends a CF-Poll along with the data frame from station 1 to station 3. Upon receiving the CF-Poll and data frame, station 3 simply sends back CF-ACK after an SIFS. The PC designates the end of the CFP through a CF-End broadcast frame, this signals the start of the CP for DCF-based transmissions.

As depicted in Figure 2.3.7, PCF incorporates the ability to piggyback numerous frame types together and thus combining them into a single frame transmission. This process enables PCF to improve medium usage efficiency. The various forms of piggybacked frame transmissions are discussed in [3].

Generally, CFP and CP are of predefined length and thus occur at regular specified intervals, however this is strictly not correct. The start and end time of a CFP can be altered, the first due to an un-foreshadowed overrun of the CP, and the latter due an earlier transmission of a CF-End frame by the PC. An early termination of the CFP can be instigated by the PC, based on the decision of the size of the polling list, the traffic load and possibly other factors that the PC determines relevant.

In case of the start of a CFP, this can be delayed due to a frame transmission occurring during the CP that extends beyond the designated end of the CP (CFP start). When this occurs, the CFP is effectively foreshortened, an illustration of this is depicted in Figure 2.3.8.

![Figure 2.3.8: CFP foreshortened](image-url)
2.3.3 QoS in Legacy IEEE 802.11 MAC

Generally speaking, in order to achieve end-to-end QoS provisioning it has to be supported throughout the complete network path. WLANs are a specific component within a complete network path, and therefore QoS provisioning within WLANs is a relevant issue. The two MAC layer mechanisms defined within the legacy IEEE 802.11 standard each provide different approaches to managing access to the wireless medium. As a result, we review each independently in terms of what QoS provisions they provide.

**DCF: QoS Issues**

Fundamentally, DCF does not support or provide any formal means of *parameterized* or *prioritized* QoS. It is unable to ensure that QoS network-related quantitative metrics can be maintained to a required value. Only best effort service for transmission of data frames is possible with DCF. All wireless stations and the AP equally contend for access to the wireless medium. All traffic types are also treated equally, therefore no form of service differentiation is provided. This is due to the fact that all packets from the higher layer arrive into a single MAC queue. Finally, the DCF mechanism as defined in the legacy IEEE 802.11 standard, does not support the scope for any modifications being made to specific DCF-based parameters such as the CW, $CW_{min}$ and $CW_{max}$.

**PCF: QoS Issues**

The PCF mechanism unlike DCF is able, under certain circumstances, to deliver a level of *parameterized* QoS. With a centralized medium access mechanism, the PC is able to dictate when to allow certain wireless stations to transmit data. The *polling list* contains the list and ordering sequence of wireless stations that the PC should poll.

If a small size *polling list* is active, and small sized data frame transmissions are required, it is clear that the PCF can maintain good control over the QoS network-related quantitative metrics delay and jitter. However, with a larger size *polling list* and polled wireless stations able to transmit data frames of variable lengths this introduces a great degree of variation in transmission time, thus affecting the observed delay and jitter. Furthermore, wireless stations can connect at different PHY rates with the AP, thus again leading to variable transmission times.

A potential drawback of PCF, is the default use of *round robin scheduling* for the polling sequence. The transmission of frames between wireless stations adjacent to each other in the polling sequence will experience different transmission delay as compared to wireless stations further apart. This is due to the fact that all transmissions must go through the PC, and only when polled can wireless stations transmit or receive a data
An important and significant flaw within PCF, is the complete omission of any form of prioritized QoS. Wireless stations and the AP still utilize a signal MAC queue for all higher layer packets. Therefore during a CFP, no service differentiation is employed between wireless stations transmitting different types of traffic.

Finally, an obvious QoS flaw within PCF is the unavoidable effects that arise from CFP foreshortening. The fact that the duration and start time of the CFP can vary, undoubtedly leads to variable QoS performance in each CFP.

2.4 IEEE 802.11e Medium Access Control (MAC) Quality of Service Enhancements

In light of the clear nonexistent or limited parameterized and prioritized QoS offered by the legacy IEEE 802.11 MAC layer schemes DCF and PCF, the IEEE 802.11 working group proposed a revision to the MAC layer standard referred to as IEEE 802.11e [1]. IEEE 802.11e specified an improved medium access scheme known as the Hybrid Coordination Function (HCF).

2.4.1 IEEE 802.11e: Hybrid Coordination Function (HCF)

HCF defines two alternative access mechanisms, the HCF Controlled Channel Access (HCCA) and the Enhanced Distributed Coordination Access (EDCA) [1]. HCCA just as with legacy PCF, utilizes a centralized controller and polling mechanism to schedule and coordinate contention-free transmissions. While EDCA is an extensively revised form of DCF that now incorporates contention-based service differentiation transmissions for specific types of traffic. HCF continues to retain the principle of CFP and CP, where HCCA is always used during the CFP, however during the CP both EDCA and HCCA can be used.

When a packet arrives at the MAC layer, its corresponding MAC frame is tagged with a specific traffic identifier (TID) value spanning from 0 to 15, based upon its QoS requirement. Note that this intuitively assumes a form of information pass through from higher layers to the MAC layer to distinguish each specific QoS requirement of packets. For simplicity and ease of discussion, we do not address this aspect of higher layer/MAC layer interaction. However, it should be noted in the case of TID values to 0 to 7, this is a direct translation from the 8 user priorities (UPs) defined in IEEE 802.1D [15].

Within the MAC layer, frames assigned TID values 0 to 7 are mapped across to 4 MAC layer queues referred to as access categories (ACs) associated with the EDCA mechanism.
While frames assigned TID values 8 to 15 are then mapped into the required traffic stream (TS) queues associated with the HCCA mechanism. Through this process HCF ensures that parameterized QoS is provided to TS queues through HCCA, while prioritized QoS is provided to AC queues through EDCA.

An important concept introduced by HCF, is the term transmission opportunity (TXOP), this simply defines the period of time allowed for a particular wireless station (or AP) to transmit on the wireless medium. Therefore, based on the TXOP duration a series of frames can be transmitted. A TXOP is referred to as an EDCA-TXOP when using the EDCA mechanism, and likewise a Polled-TXOP when using the HCCA mechanism.

In infrastructure BSS WLANs, HCF enforces that beacon frame transmissions by the AP are not delayed and are always transmitted at its required target beacon transmission time (TBTT). With TBTT information included in each beacon frame broadcast, wireless stations are not allowed to transmit a frame if its transmission cannot be completed before the upcoming TBTT.

2.4.2 HCF: Enhanced Distributed Coordination Access (EDCA)

Figure 2.4.1 shows that within each wireless station and AP, there are 4 AC queues used to support the 8 different TIDs. As a result of this process, one or more TIDs are effectively mapped across to the same AC queue. Table 2.4.1 depicts the exact mapping scheme between TIDs and ACs employed by IEEE 802.11e EDCA [1].

<table>
<thead>
<tr>
<th>Priority</th>
<th>TIDs: Same as IEEE 802.1D UPs</th>
<th>IEEE 802.1D designation</th>
<th>IEEE 802.11e Access Category (AC)</th>
<th>AC Service Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest</td>
<td>1 Background (BK)</td>
<td>0</td>
<td>Background (AC_BK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 Not defined (-)</td>
<td>0</td>
<td>Background (AC_BK)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 Best Effort (BE)</td>
<td>1</td>
<td>Best Effort (AC_BE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 Excellent Effort (EE)</td>
<td>1</td>
<td>Best Effort (AC_BE)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 Controlled Load (CL)</td>
<td>2</td>
<td>Video (AC_VI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 Video (VI &lt; 100ms latency and jitter)</td>
<td>2</td>
<td>Video (AC_VI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 Voice (VO &lt; 10ms latency and jitter)</td>
<td>3</td>
<td>Voice (AC_VO)</td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>7 Network Control (NC)</td>
<td>3</td>
<td>Voice (AC_VO)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4.1: Mapping between User Priority (UP) and Access Category (AC)

Each AC queue defined within EDCA, behaves as a virtual independent DCF-like station and will contend for access to the transmission medium. As mentioned before, EDCA is a contention-based medium access mechanism that provides prioritized QoS in the form of service differentiation between ACs. Thus, in order to facilitate service differentiation amongst ACs, two main methods are introduced in EDCA.
The first method is using different inter-frame space (IFS) intervals for each of the 4 ACs. The introduced IFS is referred to as an arbitration inter-frame space (AIFS) and is evaluated for each AC using the relationship depicted in Eq.(2.4.1).

\[
AIFS[i] = AIFSN[i] \times Slot\_Time + SIFS,
\]

(2.4.1)

where:

- \(i\) designates the specific \(AC[i]\), \(i = 0, 1, 2, 3\).
- \(AIFS[i]\) is the IFS interval assigned to \(AC[i]\).
- \(AIFSN[i]\) is the arbitration inter-frame spacing number (AIFSN) assigned to \(AC[i]\).
- \(Slot\_Time\) refers to the time duration of a slot.
- \(SIFS\) refers to the time duration of a short inter-frame space (SIFS) interval.

If a frame arrives at an empty AC queue and the wireless medium has been idle for an interval equal to its related AIFS, a transmission attempt can commence immediately. Since 4 ACs exist within each wireless station and AP, if all observe a frame arriving at an empty queue, the AC with the smallest assigned AIFS will transmit first. With
the medium becoming busy, the remaining ACs will defer medium access and start their back-off process. In the advent of certain ACs being assigned the same AIFS, a strict priority access rule is enforced within wireless stations and the AP. The internal scheduler allows the highest priority AC to transmit immediately, while all remaining ACs observe a *virtual collision* phenomena and will begin the required exponential back-off process as if a physical collision occurred.

The second method of providing service differentiation is through the allocation of different CW sizes for each AC. Thus, allocating a smaller size CW to a high priority AC will ensure that in general, the high priority AC will be able to transmit frames before a low priority AC. In the case of two or more ACs within a wireless station or AP, observing their back-off process elapsing simultaneously, the internal scheduler allows the highest priority AC to transmit, while all remaining ACs observe a *virtual collision* and will enter their exponential back-off process as if a physical collision occurred.

Aside from the above two methods of service differentiation, a minor form of differentiation is provided by EDCA through the use of a TXOP duration referred to as TXOP\_limit for each AC. The TXOP\_limit defines the duration of time that EDCA frame bursting can be used. Once a wireless station or AP gains access to the wireless medium, it can transmit multiple frames for an AC sequentially as long as the total transfer time required does not exceed the TXOP\_limit associated to that AC. As a result, each AC can have differentiated frame bursting capabilities.

The IEEE 802.11e EDCA standard specifies that for each AC, the value of AIFS\_N can range from 0 to 15, CW\_min and CW\_max can range from 0 to 32,767 at spacing intervals that satisfy the following exponential relationship in Eq.(2.4.2).

\[
CW\_\text{min} = 2^{\text{ECWmin}} - 1, \\
CW\_\text{max} = 2^{\text{ECWmax}} - 1, 
\]

where:

ECW\text{min} and ECW\text{max} range from 0 to 15.

While the duration of time TXOP\_limit represents, can range from 0s to 2.09712s at spacing intervals of 32\(\mu\)s. As in the standard, TXOP\_limit is expressed as integer value multiples of the unit time step of 32\(\mu\)s, therefore the integer value ranges from 0 to 65,535. Note that TXOP\_limit value of 0 signifies that only a single frame can be transmitted for each obtained EDCA-TXOP. The default reference configuration for these parameters as specified by IEEE 802.11e EDCA, is shown in Table 2.4.2.

In an infrastructure BSS WLAN, IEEE 802.11e EDCA specifies that the beacon frame broadcasted by the AP contain an information field referred to as the *EDCA parameter set*. The *EDCA parameter set* contains all the values for AIFS\_N, CW\_min(\text{ECW\_min}),
For PHY IEEE 802.11b
For PHY IEEE 802.11a/g
Other PHY

<table>
<thead>
<tr>
<th>AC</th>
<th>CWmin</th>
<th>CWmax</th>
<th>AIFS</th>
<th>TXOP_limit (Time Duration)</th>
<th>TXOP_limit (Time Duration)</th>
<th>TXOP_limit (Time Duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC_BK</td>
<td>31</td>
<td>1031</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AC_BE</td>
<td>31</td>
<td>1031</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AC_VI</td>
<td>15</td>
<td>31</td>
<td>2</td>
<td>6.016ms</td>
<td>3.008ms</td>
<td>0</td>
</tr>
<tr>
<td>AC_VO</td>
<td>7</td>
<td>15</td>
<td>2</td>
<td>3.264ms</td>
<td>1.504ms</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.4.2: Default IEEE 802.11e EDCA Parameter Set values

$CW_{\text{max}}(ECW_{\text{max}})$ and $TXOP_{\text{limit}}$ for each AC. Furthermore, IEEE 802.11e EDCA allows the EDCA parameter set to be altered dynamically as required by the AP. Note that IEEE 802.11e EDCA standard does not support this process in an IBSS WLAN [1].

### 2.4.3 HCF: HCF Controlled Channel Access (HCCA)

The HCCA mechanism is strictly designed to provide hard parameterized QoS support in an infrastructure BSS WLAN. HCCA like PCF utilizes a centralized control entity referred to as the hybrid coordinator (HC) to provide polling based channel access, this is expected to reside within the AP. Likewise, in order to guarantee the HC is able to control the medium, all its transmissions occur only after either a PIFS or SIFS interval has elapsed.

Unlike the PCF: Point Coordinator, the HCCA: Hybrid Coordinator is able to provide polling based channel access in both the CFP and CP. In HCCA, time on the medium is divided into beacon frame intervals, each broadcast beacon from the AP contains an information field that notifies wireless stations the duration of the CFP. The CFP follows directly after the beacon transmission, and after its completion the CP interval begins and continues through to the next TBTT. Figure 2.4.2, depicts an example of how a beacon frame interval may look like using HCCA.

As we can see from Figure 2.4.2, during the CP the HC is able to grab control of the medium in periods referred to as controlled access phases (CAPs). In order to achieve this, the HC needs to sense the medium being idle for a PIFS before it can begin a CAP. CAPs are effectively blocks within the CP, where contention-free transmission of frames take place.

The HC can initiate a CAP period, if as in the example shown, the HC decides to allocate Polled-TXOP to various wireless stations. Alternatively, the HC can also use a CAP to transmit downlink frames to individual wireless stations. During the CP, EDCA is used by both wireless stations and the AP to service AC based traffic, therefore it is expected that the AIFS durations should be chosen such that they are larger than a PIFS.
2.4. IEEE 802.11E MEDIUM ACCESS CONTROL (MAC) QUALITY OF SERVICE ENHANCEMENTS

In order to achieve parameterized QoS support in HCCA, a QoS guarantee is first negotiated between the HC and a wireless station. When a wireless station has traffic with the appropriate tagged TID (TID: 8 to 15), it first associates the traffic to one of its 8 available traffic stream (TS) queues. Note that the translation rules between the 8 TIDs and 8 TS queues are left open to implementation by the standard [1].

This active TS now has a set of traffic specification (TSPEC) parameters associated to it (average data frame rate, average frame size and delay bound are some examples of TSPEC parameters). The wireless station then notifies the HC of the TSPEC parameters associated to its active TS. Using this information, the scheduler employed by the HC will determine if it can meet the requested TSPEC. If it can, it will then generate the required Polled-TXOP schedule for the wireless station and notify the wireless station that its TS has been admitted. If it cannot, the HC will notify the wireless station that the requested TSPEC can not be supported.

Also as part of this process, wireless stations are required to notify the HC whenever its accepted TS is no longer active. This allows the HC to efficiently reschedule the existing Polled-TXOPs, and also use the freed resources to possibly accept new TSPEC requests for additional TSs. In addition to notifying the HC of inactive TSs, wireless
stations can also re-negotiate new TSPEC requirements for an already accepted TS with the HC.

2.4.4 QoS in IEEE 802.11e MAC

Providing QoS support in WLANs has clearly been made possible with the introduction of the revised MAC layer standard IEEE 802.11e. The IEEE 802.11e HCF incorporates two distinct medium access mechanisms, with EDCA providing prioritized QoS and HCCA providing parameterized QoS for the transmission of data frames within a WLAN. However, a critical issue remains in that the implementation and configuration aspect of both mechanisms are left open.

**EDCA: QoS Issues**

In EDCA, each AC’s associated EDCA parameter set can be configured differently in order to achieve a specific probabilistic service differentiation amongst the ACs. Furthermore, in an infrastructure BSS, the EDCA parameter set can be dynamically changed whenever necessary. However, the reference implementation and configuration of EDCA designates an always fixed and static EDCA parameter set.

It is clear that as network conditions change, the probabilistic service differentiation amongst ACs will also change since an always fixed and static EDCA parameter set is used. This can lead to cases where the actual priority nature between ACs is altered and most importantly, actual network performance may not be consistent across the different possible network condition scenarios.

Therefore, it is vital to understand what effects different EDCA parameter set configurations have on different network condition scenarios. Once this is well understood, another issue that remains is, deciding what specific target service differentiation performance goals should be provided in the network.

**HCCA: QoS Issues**

Generally speaking, HCCA is a deterministic medium access mechanism as opposed to the probabilistic medium access provided by EDCA. Additionally, HCCA does not suffer from any of the QoS shortcomings in PCF. As a result, HCCA is able to provide strict QoS guarantees within a WLAN.

As with EDCA, the critical issue in HCCA is with, how it should be implemented and configured. Such as determining what types of TSPEC parameters should be used for a different TSs. Based on this information the impact on the scheduling computation of Polled-TXOPs may vary. Likewise, a decision has to be made on what QoS goal
the actual scheduling mechanism is trying to achieve. For example fair allocation for all possible types of TSs, or specifically admit more TSs of a particular type over another (for example more voice-based TSs should be admitted than video-based TSs). As a result of this, the complexity level of HCCA can vary a great deal. Furthermore, since the HC resides in an AP, its computational capability and resources are limited. Therefore, this is another critical issue in determining what types of QoS goals can be achieved.

2.5 Overview of Related Work

The IEEE 802.11e standard only defines a framework of how the EDCA and HCCA mechanisms manage medium access. Unlike IEEE 802.11 DCF and PCF, a fixed complete implementation is not given and thus their respective implementations are left open for interpretation. In this thesis, our primary focus is on designing and developing an implementation of the IEEE 802.11e EDCA mechanism. In this section, we identify and provide a brief discussion of some of the relevant literature currently available that are specifically related to IEEE 802.11 EDCA.

2.5.1 Experimental Analysis of EDCA

The following [16, 17, 18, 19], all take an experimental approach to understanding the behaviour of the EDCA mechanism. Using both simulation tools and real-life hardware configurations, the effects of different configuration choices for the EDCA parameter set were observed. In particular the impact each configuration choice had on the level of service differentiation amongst ACs.

Through this process, the impact of selecting different AIFS values was shown to have the biggest impact on service differentiation between ACs. However, the level of service differentiation achieved was depicted as being fairly coarse and in extreme differentiation configurations it had the negative impact of reducing the throughput performance within a WLAN. In the case of CW parameters, the service differentiation achieved was depicted as being much more fine grained, and in particular CW\textsubscript{min} was shown to be the more effective differentiation parameter than CW\textsubscript{max}. Investigations into different TXOP\textsubscript{limit}, simply showed that TXOP\textsubscript{limit} did not have any effect on the actual access of the medium service differentiation amongst ACs. Instead, it simply provided a level of service differentiation amongst ACs once each had obtained access to the transmission medium.
2.5.2 Analytical Models

Numerous analytical models of IEEE 802.11e EDCA have been proposed and published in the literature, of which the following are some examples [2, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]. Each analytical model is essentially a direct extension of the well known analytical model of IEEE 802.11 DCF [30]. Therefore the majority are built on the primary assumption that all stations always have a frame awaiting transmission within their respective AC queue (this condition is generally referred to as saturation) [2, 20, 21, 22, 23, 26, 27, 28]. While in the case of [24, 25, 29], these models are able to account for non-saturation load conditions within stations.

In general, all analytical models serve a single purpose of being able to estimate as accurately as possible, what the expected result would be for some real-life entity. In case of the analytical models for EDCA, an estimate of the expected throughput, transmission and collision probability, and frame transmission delay can be obtained numerically for a given real-life WLAN configuration. As a result, this makes it easy to predict and understand the impact different EDCA parameter set configurations would have in an assumed WLAN topology.

2.5.3 Proposed EDCA Implementations

With the exact implementation of EDCA left open in the IEEE 802.11e standard, it comes as no surprise that numerous implementations have been proposed in the literature. These proposed implementations can generally be grouped together under two categories, Centralized Approaches and Distributed Approaches. Where essentially, Centralized Approaches rely on a single entity within the WLAN to make decisions that have network wide consequences (usually this role is designated to the AP). While Distributed Approaches specifically refer to each station making decisions independently that only affect their own individual medium access characteristics within the WLAN. In certain cases, some proposed implementations utilize both centralized and distributed approaches, however generally speaking the centralized entity is still the main driving force in determining what actual actions need to be taken within the WLAN.

Centralized Approaches:

There has been a number of implementations proposed, which take a centralized approach in managing a WLAN, some examples of which are [2, 24, 31, 32]. All of the mentioned implementations utilize the EDCA feature of allowing the AP to dynamically adjust the EDCA parameter set, which is applied within the WLAN.
In the case of [2, 24], the decision making process of determining what \( EDCA \) \( \text{parameter set} \) to use is dictated by the results obtained from an analytical model of \( EDCA \). However, both proposed implementations are built upon the assumption that all stations are saturated. This requirement is absolutely necessary for \( ACs \) in the case of [2], while in [24] this requirement is strictly only necessary for the \( ACs \) related to non real-time traffic. These requirements are not realistic as it is to be expected that the dynamics of the \( WLAN \) can change such that these conditions are not always present.

The proposed approaches in [31, 32] do not have such restrictions imposed on their operational environment. However, these approaches rely solely on heuristics to dictate what adjustments should be made to the relevant \( EDCA \) \( \text{parameter set} \). Therefore it is difficult to ascertain whether the choices made are truly optimal, since no analytical model results are used to support each decision making process.

**Distributed Approaches:**

There are numerous publications available, which describe implementations of allowing stations within a \( WLAN \) to independently adjust their own individual \( EDCA \) \( \text{parameter set} \) [33, 34, 35, 36, 37]. A major drawback and hinderance is the inherent requirement that each individual station must implement the same related configuration.

With the scenario of individual stations altering their own \( EDCA \) \( \text{parameter set} \), this raises the possible issue of unfairness between stations. Stations could alter their own \( EDCA \) \( \text{parameter set} \) in such a manner that they are more likely to obtain medium access than stations not running the same implementation. In each of the proposed solutions mentioned [33, 34, 35, 36, 37], this scenario is overlooked and it is assumed that all stations are running the same unmodified implementation.

A common trait amongst the mentioned proposals, is that all utilize a heuristic decision making approach in adapting the \( EDCA \) \( \text{parameter set} \) for each individual station. Therefore there is no indication if the choices made are optimal or not, since decisions are made without the consultation of analytical model results. Furthermore, this makes it harder to predict and gauge what the actual achievable performance might be within the \( WLAN \).

### 2.6 Motivation and Gap Analysis

In this thesis, our primary focus lies in designing and developing a complete configuration infrastructure for the IEEE 802.11e \( EDCA \) mechanism. Part of this process requires identifying what specific features should be introduced into an IEEE 802.11e \( EDCA \)-based \( WLAN \). To begin with, it is essential that each feature, is actually feasible and
has useful realistic implications in a WLAN. Therefore we identify the following set of features, which will all be used as a guideline for the design and development of our proposed complete configuration infrastructure for the IEEE 802.11e EDCA mechanism.

- **Complete Compliance to Standard Framework:**
  Adhering to the standard framework build, ensures that strict compatibility and operability is maintained between all IEEE 802.11 based devices. Therefore, regardless of possible configuration differences, the underlying interaction process remains fixed and consistent across all devices. The majority of the earlier highlighted proposed configurations in literature, in one way or form do not adhere to this principle.

  As mentioned earlier, the IEEE 802.11e EDCA framework describes strict guidelines on what parameters can be tuned, and what are the range of acceptable value settings for these parameters. In addition, the framework explicitly defines these parameters can be changed through beacon frame broadcasts from the AP. Finally, an explicit description is provided describing the required operational procedure of EDCA.

  In [2, 24], both proposed configurations specify the use of parameter settings that are not supported by the standard framework. In [33, 34], the underlying MAC frame structure is altered in a manner that is not supported by the standard framework. Additionally, individual stations are allowed to alter their own parameters, which is not supported by the standard framework. This is also a major issue with all three proposed configurations in [35, 36, 37] that specify individual station-based parameter adjustments. In the case of [32], the only issue with the proposed configuration is the unspecified nature of the information exchange process between the AP and stations, in particular if it adheres to the standard framework. In the case of [31], the proposed configuration is described as adhering to the standard framework.

  Therefore the feature *Complete Compliance to Standard Framework,* is seen as an absolute requirement in our proposed configuration. Through achieving this feature we expect that our proposed configuration will be compatible and operate smoothly with other configurations that adhere with the standard framework.

- **Centralized Approach:**
  The IEEE 802.11e standard formally specifies that all EDCA related parameter changes are to be initiated only through the use of beacon frames transmitted by the AP. This effectively dictates/mandates that any proposed configuration is centered
around the configuration of the AP and thus a Centralized Approach. Therefore we exclude all related work categorized as Distributed Approaches from consideration.

A benefit of a centralized approach is the minimal implementation requirements imposed on a WLAN. The implementation complexity is only reserved to a single device, usually this is the AP. As a result, this allows the implementation requirements for wireless client stations to remain as simple as possible. In addition, a completely centralized approach provides the benefit of being able to utilize bandwidth more efficiently given that none is wasted in the form of information frame exchanges between wireless client stations and the AP.

With control dictated by the AP, this can provide the ability for a network administrator to specify directly what management goals should be enforced within the WLAN. Furthermore with a centralized control, this can ensure that consistency and fairness is maintained across the WLAN. In order to provide this capability, the IEEE 802.11e standard specifies that all EDCA parameter changes are dictated by the AP. This means all wireless client stations are required to utilize the same set of parameters.

All of the related configurations grouped under the Centralized Approach category satisfy the primary requirement of this feature, which is that each and every EDCA parameter change is solely dictated by the AP. In [2, 24, 31] a completely centralized approach is used, while in [32] the decision making process is assisted with information frame exchanges between wireless client stations and the AP. The proposed method in [32], raises the issue of certain wireless client stations possibly falsifying information in order to gain unfair advantages over other stations.

A completely centralized approach is chosen as the design and build choice for our proposed configuration. This decision was made since our proposed configuration is expected to be implemented within the AP and thus make use of the feature allowed in the standard of initiating EDCA parameter changes via the AP itself.

• **Throughput Management Capability:**

The IEEE 802.11e EDCA mechanism provides parameterized QoS in the form of service differentiation between ACs. This service differentiation is achieved through medium access differentiation. A high priority AC is expected to access the medium with higher probability than a low priority AC. The settings for the tunable parameters for each AC are responsible for determining their respective access probabilities. However, the implicit feature of providing differentiation through Throughput Management Capability between ACs is not directly addressed in the IEEE 802.11e EDCA standard.
In [2, 24], this feature is a primary highlight of both proposed configurations. Both provide the capability of specifying exactly how much throughput proportion allocation each individual AC will obtain. However as mentioned before, this capability is only achievable under saturated conditions within the WLAN and only through the use of non-standard conforming parameter settings. In the case of [31, 32] neither of the proposed configurations is able to explicitly provide individual AC based throughput management capabilities.

Therefore just as in [2, 24], we aim to make providing Throughput Management Capability a primary feature in our proposed configuration. However, the key in our proposed configuration is to have none of the issues we have outlined with the proposed configurations in [2, 24].

- **Maximizing Throughput Performance:**

  The aspect of determining what parameter configurations are able to provide the feature of Maximizing Throughput Performance is not addressed within the standard framework for IEEE 802.11e EDCA. Thus the concept of providing this feature is left completely open to interpretation.

  The only configurations proposed in the literature, which attempt to provide this feature are [2, 24]. Both utilize an analytical model as the foundation for validating and determining the optimal EDCA parameter settings that allow Maximizing Throughput Performance to be achieved within a WLAN. However, as mentioned in previous discussions, both [2, 24] require a continually saturated WLAN in order to provide this feature and again this is only achieved through the use of non standard conforming EDCA parameter settings.

  Once again like [2, 24], we also aim for our proposed configuration to have Maximizing Throughput Performance as a key feature. As before, providing this feature should be done without requiring any explicit WLAN traffic load conditions (such as saturation). Furthermore this should all be achieved without the need for non standard settings.

- **Restriction-free Operational Requirements:**

  In [2, 24], a prerequisite requirement for the correct operation of both proposed configurations is that each station must be continually saturated for certain ACs. In the case of [2], this for all ACs within a station, while [24] requires only non real-time ACs to be saturated. Furthermore, neither of the proposed configurations are shown or depicted as being tested with TCP-based traffic associated with an applicable AC.
In the case of both [31, 32], the proposed configurations are able to operate across the broad range of possible traffic load conditions and thus neither require, the need for a continually saturated load for each station. However, just as in [2, 24], both of the proposed configurations are not shown or depicted in any test where TCP-based traffic is associated with ACs.

Therefore in order to truly achieve this feature within our proposed configuration, it must be able to operate across all possible saturated and non-saturated traffic load conditions. Additionally, this extends to the type of traffic associated with each AC. The proposed configuration should be able to cope with different traffic types, in particular TCP-based traffic.

- **Autonomous Operation Behaviour:**

The feature of *Autonomous Operation Behaviour* essentially refers to a system that is capable of operating with zero or some minimal user interaction. All of the proposed configurations in the literature we have highlighted, incorporate this feature.

In [2, 24], the only user input required is in the form of specifying the necessary properties for throughput management between ACs. Additional input information is needed in the case of [24], in the form of specifying the delay requirements for real-time traffic. Similar input requirements are also needed in [31], where both throughput for real-time traffic and delay characteristics have to be specified by the user. However, in the case of [32] a fully autonomous operation is described, with no user input requirement.

In the case of our proposed configuration, we opt for a solution similar to [2, 24, 31]. Our proposed configuration is expected to require some initial minimal user input, prior to initiating a completely autonomous operation.
Chapter 3

Throughput Control

3.1 Introduction

The IEEE 802.11e standard defines a configurable medium access control (MAC) mechanism for wireless local area networks (WLANs). The standard, as discussed in Chapter 2, defines two different access mechanisms: the Enhanced Distributed Channel Access (EDCA) and the HCF (Hybrid Coordination Function) Controlled Channel Access (HCCA). In this chapter we focus on the EDCA access mechanism, specifically the challenge in configuring the open parameters of the EDCA mechanism. The EDCA standard documentation provides a static default configuration for the open parameters, namely the parameter set for each Access Category (AC): $AIFS$, $CW_{\text{min}}$, $CW_{\text{max}}$ and $TXOP\_limit$. Not surprisingly in [16, 17, 18, 19], the default configuration is shown to be inadequate and unable to guarantee optimal performance. Therefore we specifically address the issue of developing a mechanism for setting the EDCA parameters that provides optimal performance, while allowing dynamic control of throughput allocation sharing in the network.

We first define the specific throughput allocation goal we provide within the EDCA mechanism. This is followed by a detailed description and discussion of some of the related literature. We then present our proposed allocation scheme in detail and some simulation results which demonstrate the effectiveness of this scheme.

3.2 Throughput Allocation Criterion

The EDCA mechanism defines a set of parameters that control access to the transmission channel. These parameters are allowed to be modified and dynamically configured while the network is in operation [1]. As a result, in our proposed Control Scheme-1, we dynamically configure the parameters controlling access to the transmission channel in order to achieve a specific network operating performance criteria.
3.3. RELATED WORK

We assume that Real-time traffic will be serviced by HCCA and Non Real-time traffic will be serviced by EDCA. This assumption is supported by numerous publications [19, 38, 39, 40, 41], which argue that HCCA is better suited than EDCA for servicing Real-time traffic. EDCA is unable to guarantee strict control over delay and jitter due to the fact that channel access is distributed and not deterministic as in HCCA. These are critical network performance components that need to be controlled in order to guarantee QoS for Real-time traffic. As a result, we investigate the implications of using EDCA only for delay insensitive Non Real-time traffic. Therefore we can restrict our view to just the two ACs designated for Non Real-time traffic, AC1 and AC0 [1].

We wish to control the throughput allocation to each AC so that each obtains a specified proportion of the total throughput transmitted in the network. Having achieved this, we then wish to maximize the overall throughput performance of the network. Of course, to meet an exact proportion within a random and dynamic channel access environment such as EDCA is not possible. Therefore we specify the requirement as maintaining the throughput proportions as close as possible to the required target value allocation.

This requirement should be achieved within an appropriate time scale, which we set in the order of one second time intervals. The time scale of one second intervals is chosen as it is a realistic time interval that both a network administrator and an end user can observe and quantitatively feel in terms of the performance level their specific applications are receiving from the network.

We propose a scheme that can be implemented within the standard EDCA framework that provides flexible throughput allocation management capability while the network is under saturation load. The definition of saturation load in this case is the scenario when each wireless station in the network always has at least one data frame present in its transmission queue. This definition of saturation load within an IEEE 802.11e EDCA network, is used in numerous publications [2, 20, 21, 22, 23, 24].

Essentially our proposed scheme maximizes the overall throughput performance of the network under saturation load while delivering predefined throughput proportion allocations to each AC.

3.3 Related Work

The introduction of the IEEE 802.11e standard has resulted in a remarkable amount of research, with the majority being focussed on the EDCA access mechanism. The EDCA access mechanism is a distributed probabilistic access mechanism, where performance is heavily dependent on the parameter configuration and the specific network load conditions.
Numerous publications have considered the performance of the EDCA access mechanism both through simulations [16, 17, 18, 19] and through analytical models [2, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]. The impact of different parameter configurations under specific network load conditions has been well documented in these publications. However, the important aspect of developing a configuration that achieves a specific goal in an EDCA based network is actually quite limited; we present the following [2, 24, 31, 32, 33, 34, 35, 36, 37, 42, 43] as part of our discussion.

3.3.1 Optimal Configuration of IEEE 802.11e EDCA

In [2], Banchs and Vollero derive a set of closed formulae to determine the optimal configuration of the open parameters in EDCA that maximizes the throughput performance under saturation load, subject to the weighted max-min fairness criterion for throughput allocation. The simple max-min fairness criterion for throughput allocation is to maximize the minimum throughput $r_i$ that each wireless station receives for a specific access category $AC_i$. In the case of weighted max-min fairness, a weighting value $w_i$ is given to $AC_i$, and the goal is to now maximize the minimum $(r_i/w_i)$ in the network. While trying to achieve the weighted max-min fairness solution, the following equation shown in Eq.(3.3.1) is applied, where $a$ and $b$ refer to an arbitrary $AC_a$ and $AC_b$.

$$\frac{r_a}{r_b} = \frac{w_a}{w_b}$$  \hspace{1cm} (3.3.1)

We can see from Eq.(3.3.1), that $(r_a/w_a)$ and $(r_b/w_b)$ are the same and as a result the goal is to determine a parameter configuration that maximizes both $(r_a/w_a)$ and $(r_b/w_b)$ together simultaneously.

The proposed set of closed formulae are derived from the saturation-load analytical model of the EDCA mechanism [2]. The set of closed formulae determine the optimal transmission probabilities $\tau_i$ (where $i = a, b$), such that throughput performance is maximized and the required constraint in Eq.(3.3.1) is achieved.

A critical feature used in the derivation of the closed formulae is the omission of the AIFS parameter, so that only the contention window parameters are used in the optimal parameter configuration. Using AIFS differentiation is shown to be less effective than just using the contention window parameters with minimal AIFS, when the optimization goal is to maximize throughput performance. As a result, using the chosen optimal transmission probabilities, the required contention window parameters can then be determined.

Banchs and Vollero [2], further simplify the optimal configuration by only evaluating the required $CW_{\min}$ parameter, such that $CW_{\max}$ is assigned the same value as $CW_{\min}$. This is justified in [2] on the grounds that the $CW_{\min}$ parameters are chosen to control
the collision probability, and so back-off is no longer important.

The predominant focus of Banchs and Vollero [2] is on the aspect of determining the channel access parameters required to achieve a target weighted max-min fairness criterion. They briefly discuss a possible implementation configuration in an EDCA network, where the AP evaluates the optimal parameters which are subsequently broadcast to the wireless client stations through the use of the beacon frames.

An important aspect used in the determination of the optimal parameters is that the number of active wireless client stations in the network transmitting the respective AC traffic is known. This information is critical to the optimal configuration process and how this information is acquired is not adequately addressed in [2]. However, they suggest a simple approach, utilizing an explicit admission control for wireless client stations requesting transmission access for each AC, from which the AP could obtain the number of active stations present in the network for each AC.

The proposed closed formulae is specifically tied to their analytical model, which only considers a simple homogeneous network scenario. Hence, the data frame size for each AC and the transmission rate for all wireless stations in the network are all assumed to be fixed, thus limiting the application of their mechanism.

The omission of exponential back-off through assigning $CW_{\text{max}}$ equal to $CW_{\text{min}}$ is also a critical flaw in any network implementation, since exponential back-off allows the network to be robust against any sudden increase in the number of active stations in the network. Any sudden increase results in a sudden increase in collision probability, which is gracefully reduced through the use of exponential back-off.

The implementation described by Banchs and Vollero [2] also allows the $CW_{\text{min}}$ and $CW_{\text{max}}$ parameters to be assigned any positive integer value, which is not supported in the EDCA standard. The $CW_{\text{min}}$ and $CW_{\text{max}}$ can only be assigned values that satisfy the exponential relationship shown in Chapter 2 Eq.(2.4.2). Hence, the implementation described does not fit within the EDCA standard as the optimal configuration for exponentially encoded parameters is not addressed in their proposed set of closed formulae [2]. These omissions are rectified here, by proposing and demonstrating a fully EDCA compliant, dynamic throughput allocation scheme.

### 3.3.2 Optimal Configuration of IEEE 802.11e EDCA for Real-time and Data Traffic

Serrano et al. [24] extend upon earlier work of Banchs and Vollero [2] to introduce further key developments to the configuration of IEEE 802.11e EDCA. This time by addressing the issue of handling both Real-time and data traffic. The analytical model described
previously in [2] is enhanced with key additional features that take into account specific requirements related to Real-time traffic.

The key features introduced are, the ability to evaluate the average delay for a frame that is at the head of the queue for a particular AC awaiting transmission. Furthermore the standard deviation for the delay is also evaluated analytically in the updated model. Previously the model was limited to analysis of saturated conditions, however the enhanced model in [24] is now capable of analyzing non-saturated conditions.

Therefore with these additional features Serrano et al. [24] propose the following approach to configuring an EDCA based network where both Real-time and data traffic are present. The primary goal they set out to achieve is maximizing the number of Real-time traffic flows that can be handled by the network. The requirements for Real-time traffic, as mentioned in [24], is maintaining the average delay and related standard deviation at less than or equal to maximum acceptable values ($E[delay_{max}]$, $\sigma_{max}$).

To begin with in [24], Serrano et al. specify the following static configuration for certain EDCA parameters. The AIFS value assigned for the Real-time traffic ACs is set to the lowest acceptable value, the reason simply being a means of reducing any inherent static delays. As for the contention window region, $CW_{min}$ and $CW_{max}$ are assigned the same value, this is similar to the approach presented in [2]. The explanation given is that the optimum parameters choices are selected with the known number of stations and so exponential back-off is not required; and conversely with exponential back-off in place it can cause harm to jitter performance [24]. Furthermore, the TXOP Limit for the Real-time traffic ACs is set to their maximum acceptable level. This again is chosen for the reason of minimizing delay for frame transmissions by enabling contention free bursting frame transmissions.

In the case of ACs associated with data traffic, the AIFS value is dynamically adjusted according to the network requirements. As stated before in [2], all data associated ACs are configured to use the same dynamic AIFS value, since this approach maximizes throughput. Just as with Real-time associated ACs, the contention window region is restricted to be $CW_{min}$ equal to $CW_{max}$ for all data ACs. The TXOP Limit for data based ACs is fixed to single frame transmissions only, this is done so that it reduces the delay inflicted by data traffic on Real-time traffic.

The optimal configuration therefore proposed in [24] is to determine the contention window values for Real-time ACs, a single AIFS value for all data ACs and their related contention window values. The process used assumes that information is at hand with regards to the specific number of wireless stations present generating both Real-time traffic and data traffic.

A simplified parameter search process is described in [24], that determines all required
parameter configurations, namely the contention window parameters for Real-time ACs, AIFS and related contention window parameters for data ACs. This process takes into account achieving the required delay requirements for Real-time ACs, achieving a specified weighted throughput sharing for data ACs and all the while maximizing overall throughput performance.

As noted by Serrano et al. [24], the underlying assumption is full knowledge of the number of wireless stations present requesting transmission opportunities for both Real-time and data traffic. Thus they mention that an admission control mechanism is vital in order to guarantee Real-time traffic delay performance requirements. This is due to the fact that even though it attempts to admit as many Real-time traffic requests as possible, it is expected that at some point any further admissions will result in it being unable to determine parameter configurations that satisfy the delay performance requirements.

As before there are certain flaws with the updated work presented in [24]. The choice of allocating a TXOP Limit for Real-time ACs at its maximum allowable value may also lead to severe unfairness between flows. High bit rate Real-time flows in certain wireless stations can effectively cause complete transmission lock out for other stations as they continually transmit through the contention free bursting rule. The omission of exponential back-off is also deemed to be inappropriate and problematic. Serrano et al. [24] neglect the fact that the IEEE 802.11e EDCA standard only specifies the use of exponentially encoded parameters. The optimum configuration presented all utilize integer based parameter choices for contention window values.

As mentioned before in Section 2.4, HCCA is the more appropriate mechanism to use when servicing Real-time traffic. Since it offers fine-tuned scheduling procedures that can strictly guarantee performance requirements are met. An omission in the analysis and subsequent simulation results presented is the use of CBR-based data traffic. The impact that TCP-based data traffic may have on the network configurations is not covered throughout [24].

3.3.3 HARMONICA

The HARMONICA scheme proposed by Zhang and Zeadally in [31] is designed to operate in conjunction with the EDCA access mechanism in infrastructure mode. The core of the HARMONICA architecture consists of three components which all reside within the access point (AP):

1. Link Quality Indicator Monitor (LQIM): The LQIM periodically samples the link layer quality indicator parameters, drop rate, link-layer end-to-end delay and throughput for each AC.
2. Adaptation Control Unit (ApCU): The ApCU employs two adaptation algorithms which determine the parameter set to employ for each AC based on information from the LQIM. One algorithm adjusts the relative differences between parameter sets of different ACs in order to maintain service differentiation. The other adaptation algorithm synchronously adjusts the parameter sets of all ACs in order to attain a higher channel utilization, thus leading to improved throughput.

3. Admission Controller (AdC): Admission control decisions are based on the information provided by the LQIM and the number of admission requests.

The HARMONICA architecture is driven by admission control policies. It admits all Non Real-time traffic flows, but restricts admission of Real-time traffic flows based on their throughput requirement ($\text{Req}_\text{throughput}$). A Real-time traffic flow is only admitted if it is possible to allocate its $\text{Req}_\text{throughput}$ by reducing the throughput being used for Non Real-time traffic ($\text{NRT}_\text{throughput}$), subject to the condition that a minimum bandwidth allocation for Non Real-time traffic ($\text{Bw}_{\text{NRT}_{\text{min}}}$) is still maintained. If these conditions are met, the Real-time traffic flow is admitted and allocated to an applicable AC that best matches its quality of service requirements. In detail, the HARMONICA scheme admits all Non Real-time traffic flows and only admits Real-time traffic flows if the following three requirements are satisfied:

1. The relative adaptation of the parameter sets for all ACs has achieved a stable state for the monitored network statistics determined by LQIM, so that the quality of service of current Real-time traffic flows is at a satisfactory level.

2. $\text{NRT}_\text{throughput} - \text{Req}_\text{throughput} \geq \text{Bw}_{\text{NRT}_{\text{min}}}$.

3. The bandwidth when taken out of $\text{NRT}_\text{throughput}$ can be transferred to an appropriate AC designated for Real-time traffic without loss or deterioration to existing Real-time flows.

The results for the HARMONICA scheme, demonstrate that through dynamically adjusting the parameter sets for ACs, it is possible to satisfy the quality of service requirements for Real-time traffic, while at the same time guaranteeing a minimal bandwidth allocation for Non Real-time traffic. The HARMONICA architecture is able to operate continuously regardless of network topology changes and also changing traffic conditions ranging from unsaturated to fully saturated traffic load.

HARMONICA does not require any changes to the standard EDCA framework in terms of additional signaling overhead or modified management frames. However, the architecture itself is quite complex, requiring significant modifications to both the AP
and wireless client stations. The adaptation algorithms in HARMONICA control the three parameters $AIFS$, $CW_{min}$ and $CW_{max}$ for each AC, also making the decision space complex in the presence of the four ACs defined in the EDCA mechanism. Details of the decision making scheme used in the adaptation algorithms are not given, instead only a simple skeleton outline is mentioned. This depicts an iterative process, incrementing or decrementing the parameter sets with no guidelines given as to how to increment and decrement the parameter sets.

### 3.3.4 Two-level Protection and Guarantee Protocol

The proposed Two-level Protection and Guarantee protocol by Xiao et al. [33] is based on the distributed admission control (DAC) protocol proposed in the initial draft standard for the IEEE 802.11e EDCA [44], which was subsequently omitted in the final standard [1]. The DAC protocol was proposed for use in infrastructure based wireless networks, but the ability to support deployment in wireless ad Hoc networks was later proposed and highlighted in [45]. As we are primarily concerned with infrastructure based networks, we will consider only these types of networks, as described in [33].

The basic structure of the proposed protocol in [33] is described as a Two-level Protection and Guarantee mechanism, which Protects and Guarantees the quality of service level for existing active Real-time traffic flows in the network.

The First-level Protection mechanism is based on the DAC protocol. In an infrastructure based network, the AP broadcasts an additional parameter known as the transmission budget in beacon frames for the ACs assigned for Real-time traffic ($AC_2$ and $AC_3$). The transmission budget defines the remaining transmission time for each $AC$ in each completed beacon frame period. The remaining transmission time for each $AC$ is calculated by subtracting the occupied time from a defined transmission limit of the particular $AC$ in each beacon frame period.

The wireless client stations evaluate an internal transmission limit for their own traffic for each applicable $AC$ during each beacon frame period. The local Real-time traffic transmission time for each wireless client station must not exceed the internal transmission limit for each applicable $AC$ in each beacon frame period. As a result, if the transmission budget for an $AC$ is depleted, new Real-time traffic flows will not be able to obtain any transmission time, while existing flows will not be able to increase their transmission time per beacon frame period. Therefore this mechanism appears to protect existing Real-time traffic flows. The exact implementation of the First-level Protection mechanism is quite complex, for further information please refer to [33].

An important observation about the proposed First-level Protection mechanism is that the AP and wireless client stations must store and maintain a large number of
parameters, which can be updated at every beacon frame period. However, the actual channel access parameter set \((AIFS, CW_{\text{min}}, CW_{\text{max}} \text{ and } TXOP\_limit)\) for AC2 and AC3 are not updated. Instead, the assumption is made that they are fixed to appropriate values, where the method used to determine these values is not addressed.

The First-level Protection mechanism, only controls the channel access behaviour of Real-time traffic and by itself, is unable to guarantee the quality of service level for existing Real-time traffic flows. The First-level Protection breaks down when Non Real-time traffic load increases [33], inducing an increase in collisions and interference with existing Real-time traffic. This is because priority support is not strict in EDCA, since it is a contention based access scheme. As a result, an additional control and management feature is required for Non Real-time traffic, given by the Second-level Protection mechanism [33].

The Second-level Protection mechanism controls the channel access for Non Real-time traffic in order to reduce the effective number of collisions (collision probability), caused by Non Real-time traffic transmissions. The Second-level Protection mechanism dynamically controls the parameter set, \(AIFS, CW_{\text{min}}\) and \(CW_{\text{max}}\) for the ACS classified for Non Real-time traffic use. Each wireless station (including the AP) is expected to dynamically control its parameter set usage for the ACS classified for Non Real-time traffic use (AC0 and AC1). The proposed technique of dynamic control consists of two schemes, the Intra-frame based scheme and the Inter-frame based scheme [33]. The Intra-frame based scheme utilizes an alternative back-off process as compared to the standard exponential process defined in EDCA [33]. The primary goal is to achieve a faster increase in contention window sizes for stations with Non Real-time traffic involved in collisions. While, the Inter-frame based scheme is slightly more complicated, it increases or decreases the \(AIFS\) and \(CW_{\text{min}}\) based on a given relationship between the number of successful and unsuccessful transmissions.

The combination of First-level Protection and Second-level Protection mechanism is shown through simulation results to maintain the specified quality of service level for existing active Real-time traffic flows in the network [33]. The protocol introduces an additional form of service differentiation in EDCA, in that it allows the partitioning of the transmission time available in a beacon frame period between the ACS classified for Real-time traffic use. This may also provide a form of throughput allocation under the scenario of fixed equal transmission rates for all stations in the network.

The DAC protocol assigns the responsibility for adjusting/updating parameter values (internal transmission calculation) to individual wireless client stations. The AP also determines its own individual parameter values along with broadcasting the global transmission budget information. Hence, a problem with the DAC protocol which is inherited by the proposed Two-level Protection and Guarantee protocol, is that the AP and wire-
less client stations could continually and independently adjust their parameters at each beacon frame period, leading to network performance oscillations. The issue of unfairness amongst wireless client stations also arises, as it is possible for stations to have varying parameter values causing some stations to receive better performance than others.

Another drawback with the Two-level Protection and Guarantee protocol, is the use of fixed, non-dynamic, channel access parameter set values for the AC classified for Real-time traffic, which may limit the effective capacity of the network.

The Inter-frame based scheme uses an iterative process for adjusting the channel access parameters. However, the method for determining the actual channel access parameters required in each iteration is not defined. Additionally, the alternative back-off processes proposed in the Intra-frame based scheme are not supported in the standard framework defined for EDCA and so it requires the standard to be modified.

The protocol also defines static reservation of transmission allocation times for each specific AC. When the transmission allocation time for a specific AC is not fully utilized, the remaining unused transmission allocation time can not be allocated to another AC, resulting in an under-utilization of the transmission channel.

### 3.3.5 Dynamic Two-level Protection and Guarantee Protocol

In [33], the rigid static bandwidth allocation mechanism is a critical flaw that can lead to under-utilization of wireless network resources when sufficient traffic is not present. This major flaw is addressed by Xiao et al. [34] with the introduction of dynamic budget partitioning capability.

The key feature in [34] is the ability to now partition bandwidth dynamically among Real-time traffic based on the current AC traffic load condition. Previously in [33] bandwidth partitioning was carried out in a fixed static manner, in that a specified amount was allocated to Real-time ACs independently. While in the case of [34], all Real-time ACs are combined together and allocated a bandwidth partition which is then shared proportionally.

Thus the basic concept is that, within this overall bandwidth allocation, a general proportional sharing amount is specified between the Real-time ACs. Any unused proportional bandwidth within one AC allocation can be transferred to the other AC to assist in allowing additional video (voice) flows to operate and meet their transmission requirements. This operation is carried out by the AP as it already accounts for the task of broadcasting the transmission budgets for the network.

The exact procedure of how the decision is made is not shown in detail [34]. The introduced dynamic bandwidth allocation configuration is shown to perform better than [33] in variable Real-time traffic load conditions as it is able to admit more traffic flows
in the same test case scenarios.

Therefore it is clear that the dynamic bandwidth partitioning scheme has improved the network’s ability to utilize bandwidth more efficiently, and in doing so allow more Real-time traffic flows to be admitted and serviced. However the key point remains that the majority of flaws stated in [33] are still present in the revised scheme of [34].

3.3.6 Throughput Improvement via Collision Management

The suite of papers [32, 35, 36, 37] show different approaches that can be taken to simply adjust the contention window parameter [32, 36, 37] or in addition the AIFS parameter [35] for ACs such that throughput is improved. All proposed solutions do not utilize any form of analytical model to govern what the required size of the window should be in order to achieve optimal performance or subsequently a particular proportional access rule between ACs.

The overall consensus across these proposals is the use of the collision probability as the governing network measure for making decisions on how the adjustment of parameters is carried out. The general rule used is that in an observed high collision probability environment throughput performance is being degraded and thus to overcome this, parameter adjustments should be performed to reduce collisions and thus hopefully improve throughput. Likewise the opposite rule stands in the scenario where a low collision probability is being observed.

In [35], Lenagala and Zeng describe the scenario where while using the default EDCA suggested parameter setup, at low offered traffic load the carried traffic load across all ACs is very similar. Upon increasing the offered load across all ACs, the carried load for high priority ACs stays fairly consistent. However, in the case of low priority ACs the carried load begins to drop drastically. This is simply due to the fact that the parameter configurations are such that it receives less transmission opportunities than high priority ACs.

Thus the goal in [35] is to reduce the impact this can have on overall network performance. Two measurements are carried out at some specified interval, namely the collision probability for a given AC and also its observed proportional transmission count (both are an overall network based measurement for the given AC). Using these two measurements, they are then compared to a predefined threshold value for that particular AC. Thus based on the relation to the predefined threshold parameter, changes for both AIFS and CW are then carried out.

It is hard to determine exactly how the process in [35] is performed physically, as there is no mention whether this approach is centralized with the AP performing the said tasks of measuring and distributing updated parameters, or whether it is distributed.
Additionally the actual time frame required to measure and then compare results in order to initiate parameter changes is not described in detail. The basic approach described is a simple tuning mechanism that does not attempt to identify or discover what is the required optimal network configuration. Furthermore its simplicity does not enable the ability to define a proportional allocation capability between ACs.

In [36, 37], both describe similar approaches which involve altering the value of the contention window that both the AP and wireless stations use based on the observed collision probability of the network for a given AC. The use of collision probability is common in both cases, however the two differ on the actual adjustment of the contention window values. However the same principle is still used in that in a high collision probability environment the contention window value should be increased and likewise in low collision probability environment it should be decreased.

Ojeda-Guerra and Alonso-Gonzalez [36] describe the modification of selecting a contention window value used for determining the random back-off from the range specified by $CW_{min}$ and $CW_{max}$ for a given AC. The simple premise is that in the default configuration whenever a collision is detected the contention window will double from the initial $CW_{min}$ to another value and so on, upon each additional collision, up to $CW_{max}$. However upon a successful transmission the contention window value will reset back to $CW_{min}$. Thus the proposed approach is that the initial contention window value should effectively be close to or equal to $CW_{min}$ in low congestion environments, while in high congestion it should be close to or equal to $CW_{max}$. The process of determining this is detailed in [36].

In [37], a brief and incomplete description is given of the process of adjusting the $CW_{min}$ and $CW_{max}$ values for an AC. The introduction of undefined variables makes it difficult to assertain how the configuration operates. However, the proposed approach bears the concept of a simple Multiplicative Increase/Multiplicative Decrease mechanism, where in a high observed collision environment both $CW_{min}$ and $CW_{max}$ are increased and conversely in a low observed collision environment they are both decreased.

Both [36, 37] appear to be decentralized schemes, in that each individual station is expected to carry out the said actions independently. This approach is prone to introducing unfairness with possible instances of stations operating with an unfair advantage over others due to their selected parameter configurations. Furthermore the actual measurement time required to perform specific calculations related to collision probability is not discussed in great detail in either [36, 37]. This makes it hard to determine the speed and efficiency of the schemes in updating to fluctuating network traffic loads. Finally both schemes are simplistic in nature and do not offer the ability to govern the network sharing of throughput between ACs.
In [32], Pries et al. describe a proposed mechanism, as in the case of [37], of adjusting both the $CW_{\text{min}}$ and $CW_{\text{max}}$ for each AC purely based upon the collision probability. The major difference is the centralized approach proposed in [32]. The AP is given the task of obtaining the network collision probability statistics, this is further assisted by the fact that wireless stations are required to transmit their own individually perceived collision probability information to the AP. Utilizing all the available information the AP is then able to make appropriate decisions on adjusting the $CW_{\text{min}}$ and $CW_{\text{max}}$ for the ACs.

The centralized approach described in [32] does overcome the issues with unfairness when employing a decentralized parameter adjustment scheme. However, the introduced overhead impact of transmissions of wireless stations reporting their specific collision probability is not mentioned. Furthermore this relies on wireless stations all being able to perform the said operation. Additionally, the exact process of how the information exchange occurs is not mentioned at all. As before, in [32] the time scale of how these actions are carried out is not mentioned in any great detail. Finally as with all the previous described schemes, it is relatively simplistic in that no optimal configuration is determined via the use of analytical results. Furthermore no form of proportional throughput allocation capability is possible.

### 3.3.7 Alternative Non MAC Layer Approach

In [42, 43], an alternative approach is taken where a subsequent management mechanism is over-layed on top of the IEEE 802.11e EDCA MAC layer. Thereby keeping the default configuration specified by the standard. This additional component is expected to reside in all wireless stations and the AP itself.

The primary goal described in both papers is to introduce the capability of altering the rate at which frames enter the MAC layer and effectively changing the number of wireless stations that will compete for transmission access. Thus the probability of transmission for each AC is actually being modified in a certain manner.

With this in mind both [42, 43] present similar scheduling mechanisms which produce this behavior by probabilistically altering the rate at which frames enter the MAC layer. However, the two differ in their objectives.

In [42], Hu and Hou discuss that through the use of analytical results, the reverse optimality question can be asked which is for a given EDCA parameter configuration what is the optimal number of wireless stations transmitting to maximize throughput. Likewise between ACs what number of stations competing will result in a required proportional throughput allocation between the ACs. Thus for each AC the scheduling mechanism adjusts the number of wireless stations that appear to compete for transmission access.
In [43], an analytical model is used by Wang et al. to determine what the related transmission probabilities are for the given number of wireless client stations transmitting saturated traffic load. Furthermore the related throughput and average delay can also be obtained. Wang et al. [43] then state that through the use of the scheduling mechanism within each wireless client station they can then determine what artificial transmission probability should be employed for an AC in order to now achieve different throughput and delay performances.

Note that it is implied that in both [42, 43] each wireless station is required to be aware of the total number of actual wireless stations that have specific AC traffic for transmission. The exact implementation of how this is achieved is not addressed in either [42, 43].

Both proposed scheduling mechanisms are highly dependent on the fact that all wireless stations conform to the same operations in evaluating the required artificial transmission probability. Furthermore adjusting transmission probabilities is basically achievable through altering the EDCA parameters directly, which makes the mechanisms described, using an additional overhead layer, unnecessary. Additionally this process of altering the EDCA parameters directly can be performed centrally via the AP which completely negates the need for wireless stations to carry out independent operations.

There also lies the issue of possible synchronicity when in fact all wireless stations attempt to transmit the same AC traffic together, thus leading to collisions and degraded performance. There is no mention in [42, 43] if another underlying mechanism is in place which overcomes this problem.

### 3.4 Proposed Throughput Allocation Scheme

In order to achieve the goal of maximizing throughput under saturation load subject to a defined throughput proportion allocation for each AC, we begin with a few specification requirements and assumptions regarding the implementation and deployment environment of our proposed scheme.

The proposed scheme should:

- be configurable and operate strictly within the EDCA framework.

- not introduce additional overhead or alter the basic operational process of the EDCA mechanism.

- be easily deployable into an infrastructure based IEEE 802.11e EDCA enabled network.
Our proposed throughput allocation scheme will be implemented within the AP, so that no configuration changes are required within the wireless client stations. Implementing the proposed throughput allocation scheme within the AP means that we can utilize the AP’s ability to broadcast the channel access parameters for each AC to the wireless client stations, periodically, through beacon frame transmissions. Therefore, our proposed throughput allocation scheme can dynamically reconfigure the parameters for each AC.

Another important feature that the AP offers, is the ability to observe all data frame transmissions within its service set, this is useful in allowing our proposed scheme to log the throughput performance for each AC.

The AP itself can be seen as a device that simply serves the purpose of forwarding data to and from wireless client stations. The AP is primarily a MAC layer device, which means it has no direct knowledge of what TCP based traffic interactions are occurring anywhere along the end-to-end network path. Furthermore the AP is completely unaware of the reasons that can cause TCP to reduce the transmission rate of individual flows (timeout and Triple Duplicate ACKs). This is challenging to handle in a control mechanism that is strictly localized within the AP and where the design constraints do not allow for cross layer interaction.

Therefore with these basic requirements and assumptions in mind, we now describe in detail our proposed throughput allocation scheme.

### 3.4.1 Description of Throughput Allocation Scheme

The fundamental goal of our proposed throughput allocation scheme is to maximize the throughput under saturation load subject to a defined throughput proportion allocation for each AC. Initially, we assume only saturation load conditions are present in the network. In addition, based on our important assumption that EDCA is only used for Non Real-time traffic and HCCA is for Real-time traffic, we are also able to restrict our view to only two ACs.

In order to achieve our goal of developing a throughput allocation scheme we need to control the contention access mechanism that each station uses for a given AC. We simplify and reduce the complexity of the scheme, by adjusting the contention window parameters only through the $CW_{\min}$ and $CW_{\max}$ parameters. It has been argued in [2] that $AIFS$ differentiation does not provide additional benefit in the control process and in fact reduces throughput.

Therefore, we assign the $AIFS$ parameter to be the minimal value of a $DIFS$ interval as allowed in the EDCA standard. This provides the best performance, since this is the minimum delay duration that all stations incur prior to entering the contention phase. The values for $TXOP\_limit$ do not directly affect the contention mechanism and as a result
are set to be the same across all ACs. For simplicity, the value is set to zero, allowing stations to only transmit one frame at a time. The value we assign for the individual CW\(_{max}\) for each AC is discussed later on, and is an important aspect of our proposed scheme, as this allows exponential back-off and hence maintains robustness.

**Determining the CW\(_{min}\) parameters:**

As argued earlier, we limit our focus to a two AC system, where AC0 and AC1 are assigned to *Non Real-time* traffic. In addition we assume initially that the network is operating under saturation load. We can observe the following long term relationship

\[
\left[ \frac{n_{1}}{CW_{min1}} \right] = AC1_{trx} \approx \frac{Thrpt_{1}}{Thrpt_{0}},
\]

where:

- \(n_{0}\) is the total number of stations transmitting AC0 traffic.
- \(n_{1}\) is the total number of stations transmitting AC1 traffic.
- \(CW_{min0}\) is the AC0 access category value for \(CW_{min}\).
- \(CW_{min1}\) is the AC1 access category value for \(CW_{min}\).
- \(AC0_{trx}\) is the slot-based transmission probability for AC0.
- \(AC1_{trx}\) is the slot-based transmission probability for AC1.
- \(Thrpt_{0}\) is the throughput (Bits) for AC0 (across a complete simulation period).
- \(Thrpt_{1}\) is the throughput (Bits) for AC1 (across a complete simulation period).

Given \(CW_{min0}\) and \(CW_{min1}\) along with \(n_{0}\) and \(n_{1}\), Eq.(3.4.1) determines the approximate transmission probability \(AC0_{trx}\) and \(AC1_{trx}\) that each station will achieve. The transmission probabilities \(AC0_{trx}\) and \(AC1_{trx}\) represent the proportion of slots in which transmission attempts are initiated for AC1 and AC0 respectively. Therefore, since we observe proportional transmission attempts for each AC, we can conclude that this is approximately the proportional throughput performance seen for AC0 and AC1 as in Eq.(3.4.1), when there are few collisions.

However, the relationship in Eq.(3.4.1), breaks down when the values of \(CW_{min0}\) and \(CW_{min1}\) are relatively small in comparison to the values of \(n_{0}\) and \(n_{1}\), because the collision probability for an AC is directly related to the number of stations transmitting and the contention window size used. That is, if there are many stations attempting to transmit, each with a small \(CW_{min}\), the probability of a station transmitting in a given contention slot is relatively high and hence the probability of collision is also high because of the number of stations present. Therefore, small values for \(CW_{min0}\) and \(CW_{min1}\) cause the collision probability to be significant enough to cause a degradation in throughput for
AC0 and AC1, thus leading to different throughput proportion allocations being obtained. Thus we conclude that Eq.(3.4.1) only holds when the values for $CW_{\text{min}0}$ and $CW_{\text{min}1}$ used are large enough to cause the transmission probabilities $AC0_{\text{trx}}$, $AC1_{\text{trx}} << 1$.

The relationship shown in Eq.(3.4.1) applies to a homogeneous traffic profile for all stations in AC0 and AC1 and equal physical transmission rate for all stations. In order to support heterogeneous traffic profiles and maintain the correct throughput proportion allocation between Acs, the ratio of the average data frame size transmitted for each AC has to be included as a scaling factor into the relationship depicted in Eq.(3.4.1). This is necessary in order to change the transmission probabilities $AC0_{\text{trx}}$ and $AC1_{\text{trx}}$ to achieve the correct throughput proportion allocation between ACs.

In the case of multi-rate transmissions (physical transmission rate), the transmission probability of all wireless stations remains the same if all stations utilize the same contention access parameters [46]. Therefore the effects of different physical transmission rates for wireless stations within an AC will not affect the throughput proportion allocation observed. Thus different physical transmission rates for the wireless stations do not affect the relationship shown in Eq.(3.4.1).

Hence, the correct throughput proportion allocation between ACs for all possible network scenarios is depicted in Eq.(3.4.2).

$$\frac{\left[\frac{n_1}{CW_{\text{min}1}}\right] E[P]_1}{\left[\frac{n_0}{CW_{\text{min}0}}\right] E[P]_0} = \frac{AC1_{\text{trx}} E[P]_1}{AC0_{\text{trx}} E[P]_0} \approx \frac{Thrpt_1}{Thrpt_0},$$

(3.4.2)

where:

- $E[P]_0$ is the average data frame length for AC0.
- $E[P]_1$ is the average data frame length for AC1.

The relationship depicted in Eq.(3.4.2) yields a set of candidate values for $CW_{\text{min}0}$ and $CW_{\text{min}1}$ but does not directly determine the appropriate values that maximize throughput performance. This is done by determining all possible combinations of $CW_{\text{min}1}$ and $CW_{\text{min}0}$ that satisfy Eq.(3.4.2) and then by selecting the combination with the highest throughput performance. This procedure is achieved with the aid of an analytical model of EDCA [2].

A critical omission thus far in the discussion, is the restriction imposed by the EDCA mechanism on applicable values for $CW_{\text{min}}$ and $CW_{\text{max}}$. The values for $CW_{\text{min}}$ and $CW_{\text{max}}$ are exponentially encoded as follows.

$$CW_{\text{min}} = 2^{ECW_{\text{min}}} - 1,$$

(3.4.3)

$$CW_{\text{max}} = 2^{ECW_{\text{max}}} - 1,$$

(3.4.4)
where $ECW_{\text{min}}, ECW_{\text{max}} \in \{1, \ldots, 15\}$.

This restriction is overlooked in [2], where any positive integer values are allowed in the selection process. However, we strictly adhere to the standard EDCA framework [1] and select exponentially encoded parameter values for $CW_{\text{min}}$ and $CW_{\text{max}}$.

**Determining the $CW_{\text{max}}$ parameters:**

Thus far we have only considered determining the values of $CW_{\text{min}0}$ and $CW_{\text{min}1}$ but we also need to assign values to $CW_{\text{max}0}$ and $CW_{\text{max}1}$. This defines the maximum contention window size that can be used in the exponential back-off process of the EDCA mechanism during data frame retransmissions. The exponential back-off during data frame retransmissions is a critical feature that we include in our throughput allocation scheme. It inherently adds a level of robustness to the network for situations when traffic loads dramatically increase, such as when the number of active stations increase or if the control malfunctions in any way. This critical feature is overlooked in [2, 24], where the $CW_{\text{max}}$ values are set the same as the calculated $CW_{\text{min}}$.

In general, the assumption is that during a normal operating cycle, the effective contention window used during the back-off process in a wireless station will be $CW_{\text{min}}$. In the event of a collision occurring, resulting in a data frame retransmission, the contention window used will then be a value determined by the next increment above the related $ECW_{\text{min}}$ up to the maximum defined by $ECW_{\text{max}}$.

The default IEEE 802.11 DCF, assigns $CW_{\text{min}} = 31$ and $CW_{\text{max}} = 1023$, this is equivalent to five exponential back-off increments, and the following values respectively $[ECW_{\text{min}}, ECW_{\text{max}}] = [5, 10]$. In our scheme, the $ECW_{\text{max}}$ is assigned to be five exponential back-off increments above $ECW_{\text{min}}$. However, if this exceeds the largest value allowed in the EDCA standard, we simply assign $ECW_{\text{max}}$ to be this largest allowed value, as represented in Eq.(3.4.5).

$$ECW_{\text{max}} = \text{min}([ECW_{\text{min}} + 5], 15)$$  \hspace{1cm} (3.4.5)

where:

$ECW_{\text{max}}$ is the value used in Eq.(3.4.4) to determine the required $CW_{\text{max}}$.

$ECW_{\text{min}}$ is the value used in calculating the $CW_{\text{min}}$ obtained from Eq.(3.4.2).

The use of exponentially encoded values for $CW_{\text{min}}$ and $CW_{\text{max}}$, introduces an additional problem in the form of limiting the possible throughput proportion allocations that can be directly achieved. Therefore, we propose using a simple technique involving regular switching between the allowable exponentially encoded parameters. This results
in an average throughput proportion allocation being achieved over a period of time which is close to the required target value.

For a given $n_1$ and $n_0$ configuration, we determine directly using Eq.(3.4.2) the possible AC1 throughput proportion allocations that can be achieved using combinations of values for $CW_{\text{min}1}$ and $CW_{\text{min}0}$. We can then determine using an analytical model [2], the specific combination of $CW_{\text{min}1}$ and $CW_{\text{min}0}$ that maximizes throughput performance for each possible AC1 throughput proportion allocation. This procedure effectively generates a list of parameter set combinations that will be used in Control Scheme-1.

The parameter list is configured to be in increasing AC1 throughput proportion allocation values. We highlight the specific parameter set combination that achieves the closest AC1 throughput proportion allocation to the desired value as $(x_0)$. The parameter set combinations that achieve throughput proportion allocations less than $(x_0)$ are subsequently labeled $(x_i)$: where $i = 1, 2, 3, ..., M$ ($M$: Index of the first parameter set combination in the list). Likewise the parameter set combinations that achieve throughput proportion allocations greater than $(x_0)$ are labeled $(x_j)$: where $j = 1, 2, 3, ..., N$ ($N$: Index of the last parameter set combination in the list).

This sorted and labeled parameter list is then simply referred to as the Parameter List. So far, we have provided a rough description of the process used to determine the Parameter List. The exact procedure is detailed below.

1. We use Eq.(3.4.2) to determine the approximate relationship between $CW_{\text{min}1}$ and $CW_{\text{min}0}$ that achieves the closest throughput proportion allocation to the desired value. This approximate relationship defines a set of all applicable $CW_{\text{min}1}$ and $CW_{\text{min}0}$ as shown in Figure 3.4.1 by the highlighted blue line. For simplicity, we have plotted Figure 3.4.1 using the $ECW_{\text{min}1}$ and $ECW_{\text{min}0}$ value as the point of reference instead of the actual exponential value of $CW_{\text{min}1}$ and $CW_{\text{min}0}$.

2. We then determine, using an analytical model [2], which specific parameter set combination along the blue line maximizes total throughput performance. This parameter set combination is then stored in Parameter List as $(x_0)$.

3. In Figure 3.4.1, we highlight the additional diagonal lines that define the various throughput proportion allocations that can be achieved using the specific $ECW_{\text{min}1}$ and $ECW_{\text{min}0}$ value combinations. These diagonal lines effectively represent the approximate linear relationship defined by Eq.(3.4.2). In this case we can see that the red dashed diagonal lines, effectively define AC1 throughput proportion allocations less than the blue line. Likewise the green dashed diagonal lines defined AC1 throughput proportion allocations greater than the blue line.
4. Parameter set combinations along each individual diagonal line will provide approximately the same throughput proportion allocation. The important goal in this case is to determine which specific parameter set combination maximizes throughput performance.

5. We could determine the specific parameter set combinations using the same approach that yielded \((x_0)\) from the blue line, but this is not necessary. We only need to consider the two neighboring points to \((x_0)\) that lie on each of the closest diagonal lines ‘O1’ and ‘R1’.

6. Let’s consider the case of increasing AC1 throughput proportion allocation. We determine, using an analytical model [2], which of the two neighboring parameter set combinations on ‘R1’ provides the highest throughput performance. This has been done in Figure 3.4.1, where we label the best combination as ‘1’, and the second best as ‘2’. The parameter set combination labeled as ‘1’ is then stored in
7. We then perform the same operation by looking at the two neighboring points to \((x_1)\) on the next closest diagonal line ‘R2’ to identify the parameter set combination which gives the best throughput performance on ‘R2’ which is stored as \((x_2)\). This process can then be performed recursively to determine all parameter set combinations \((x_j)\), for \(j = 1, ..., N\) \((N:\) Index of the last parameter set combination in Parameter List\).

8. We include a constraint that an accepted throughput proportion allocation must have a minimum 0.5% difference with its predecessor. This restricts the number of green and red diagonal lines used to determine an applicable parameter set combination. Essentially as parameter set combinations move outwards, the ability to provide a significant proportion allocation difference with its predecessor diminishes. Thus redundant parameter set combinations can be omitted, thereby this enforces a minor constraint on the boundary index values for ‘M’ and ‘N’ within Parameter List.

9. The same procedure is also applied for decreasing AC1 throughput proportion allocations to determine the parameter set combinations \((x_{-i})\), for \(i = 1, 2, 3, ..., M\) \((M:\) Index of the first parameter set combination in Parameter List\).

10. We then determine the required \(ECW_{max1}\) and \(ECW_{max0}\) values using Eq.(3.4.5) to obtain the final Parameter List that is used in Control Scheme-1.

Control Scheme-1:

The throughput allocation scheme we propose resides within the AP in the wireless network. A fundamental assumption of our throughput allocation scheme is that the AP is able to determine the number of active stations \(n_0\) and \(n_1\).

Using Parameter List, the AP broadcasts the appropriate parameter value \((x_0)\) in a beacon frame to the wireless client stations, whenever there is a change in \(n_0\) and \(n_1\). During the period that the number of stations \(n_0\) and \(n_1\) remains constant, a parameter set switching technique (detailed below) is employed to achieve the required throughput proportion allocation.

- If the measured throughput proportion allocation in a completed beacon frame period is above the target value, we decrement to the next corresponding parameter set values from Parameter List for the new beacon frame period.
• If the throughput proportion allocation achieved in a completed beacon frame is below the target value, we increment to the next corresponding parameter values from Parameter List for the new beacon frame period.

Hence, the parameter values may be continually switched at each beacon frame period, to achieve an average throughput proportion that is very close to the target value. The switching process is very simple with only single increments and decrements allowed within the sorted list of parameter values.

An additional benefit of the proposed switching technique is that if an incorrect starting point is used, the switching technique moves to and operates in the correct region, providing the correct throughput proportion. Therefore the proposed technique is robust and can function effectively from any arbitrary starting point ($x_0$).

There are many issues associated with maintaining throughput proportion and maximal throughput not the least of which are the stochastic fluctuations in measurements at each beacon frame. This is a key issue we have considered at great length.

We investigated the use of a sliding window averaging mechanism over a number of beacon frames as a possible solution to smooth out the stochastic fluctuations in the throughput proportion measurements. The sliding window approach calculates an average throughput proportion value based on a certain number of historic beacon frame period measurements. Therefore a large change in the measured throughput proportion in a beacon frame period is not clearly identifiable until the sliding window had sufficiently been updated with additional beacon frame measurements.

However, when we implemented the sliding window, we observed that it suffered from undesirable hysteresis effects which resulted in unsatisfactory output results. This is because Control Scheme-1 is specifically configured to update the parameter set being used in every beacon frame period.

Given that we require throughput proportion allocation performance to be on a one second time scale interval, we determined that single beacon frame period measurements in fact gave us the required level of control.

We have limited the number of ACs to two, based on the assumption discussed earlier that EDCA is utilized only to serve Non Real-time traffic. The proposed scheme and techniques discussed here can be extended to encompass a four AC system, if the same goal of throughput proportion allocation under saturation load is deemed desirable.

### 3.5 Performance Evaluations

In order to adequately evaluate our proposed Control Scheme-1, we break down the analysis into two stages. In ‘Stage - 1’, we begin with a controlled comparison between
our proposed Control Scheme-1 and a previously published scheme [2]. This is then followed by some additional simulation results that are used to highlight specific inherent features of our proposed Control Scheme-1. In ‘Stage - 2’, we specifically analyze the performance of the control in a realistic network environment, as opposed to the controlled network environment used in the first stage.

3.5.1 Performance Metrics

For all simulations, unless otherwise stated, the simulation results are averaged over one second intervals for presentation in graphical form. We present:

- The average throughput for AC1 and AC0 and combined total throughput in terms of one second intervals.

- The related achieved throughput proportion result (percentage) with respect to AC1 in terms of one second intervals.

We define two additional performance metrics referred to as the excursion performance measure, and model based throughput performance comparison. The first metric is simply the numerical value representing the absolute magnitude displacement of the throughput proportion allocation percentage against the target value. Thus for each second of simulation, an absolute magnitude displacement is recorded, and the average is taken over the total simulation run time.

The second metric, represents a relative throughput measure of the overall throughput performance against the value obtained from the mathematical model. For each second of simulation, the measured throughput value is represented as a percentage amount of the reported mathematical model throughput value for the given network condition. Once again, the average percentage value is presented for the total simulation run time. Note that the mathematical model data uses only the best case optimal exponentially encoded parameter set configurations.

The excursion performance measure provides useful information to gauge the effectiveness in maintaining the proportion allocation at the required level. Furthermore it also serves as a comparison tool between different simulation test case scenarios. The relative throughput measure metric, is primarily a tool for gauging if the control is achieving acceptable levels of throughput performance through using the mathematical model as a baseline comparison.

The following two equations Eq.(3.5.1) and Eq.(3.5.2) are the mathematical representation of the two performance metrics discussed above.
3.5. PERFORMANCE EVALUATIONS

Excursion Performance Measure:

\[ \text{Displacement}(\%) = \frac{\sum_{i=1}^{l} |P_i - \text{Target Proportion}|}{l} \]  

(3.5.1)

Relative Throughput Measure:

\[ \text{Relative}(\%) = \frac{\sum_{i=1}^{l} (T_i ÷ \text{Model Throughput}) \times 100}{l} \]  

(3.5.2)

where:

- \( l \) is the total number of per-second based measurements within the simulation duration.
- \( P_i \) is the AC1 throughput proportion value recorded at time \( i \), (%).
- \( T_i \) is the throughput value recorded at time \( i \).
- \( \text{Target Proportion} \) is the target AC1 throughput proportion allocation (%).
- \( \text{Model Throughput} \) is the throughput value obtained from the mathematical model.

3.5.2 Simulation Setup

The simulation tool used is NS-2 (version 2.29) package [47], combined with an EDCA implementation in NS-2 developed by the TKN group in Technical University of Berlin [48]. We have modified the EDCA implementation in order to simulate the actions of Control Scheme-1. We use the analytical model given in [2] for the initial part of our proposed throughput allocation scheme in order to determine the required ParameterList.

The MAC and Physical layer parameters used are shown in Table 3.5.1. In ‘Stage - 1’ of our simulation results, we assume a homogeneous traffic profile of high bit rate UDP/CBR traffic with packet size 1000 bytes for both AC0 and AC1. The use of UDP/CBR based traffic is primarily based on the fact that true saturation load conditions can be maintained at each active wireless client station. Hence, we can predict effectively the expected performance for AC1 and AC0 in the network.

In ‘Stage - 2’ of our simulation results, we introduce the use of TCP/FTP based traffic for AC0 and AC1. This forms a more realistic traffic profile for Non Real-time traffic within the network. As a result, the simulation results in ‘Stage - 2’ allow us to effectively gauge the performance in a more realistic environment.

In a wireless network, the wireless client stations are able to transmit data frames at different physical rates, the applicable rates are shown in Table 3.5.1. We assume initially, that the physical transmission rate for all wireless client stations is the same and fixed to 11Mbps. The network topology used for the range of simulations carried out is shown in Figure 3.5.1.
Table 3.5.1: MAC and PHY Parameters for IEEE 802.11b

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFS</td>
<td>10 $\mu$s</td>
</tr>
<tr>
<td>PIIFS</td>
<td>30 $\mu$s</td>
</tr>
<tr>
<td>DIFS</td>
<td>50 $\mu$s</td>
</tr>
<tr>
<td>Slot/Time</td>
<td>20 $\mu$s</td>
</tr>
<tr>
<td>AIFSN$_1$</td>
<td>2</td>
</tr>
<tr>
<td>AIFSN$_0$</td>
<td>2</td>
</tr>
<tr>
<td>AIFS$_1$</td>
<td>50 $\mu$s</td>
</tr>
<tr>
<td>AIFS$_0$</td>
<td>50 $\mu$s</td>
</tr>
<tr>
<td>PHY header (PHY)</td>
<td>192 bits</td>
</tr>
<tr>
<td>MAC header (MAC)</td>
<td>256 bits</td>
</tr>
<tr>
<td>ACK frame (ACK)</td>
<td>112 bits</td>
</tr>
<tr>
<td>Frame checksum (FCS)</td>
<td>32 bits</td>
</tr>
<tr>
<td>Data Rate ($D_{Rate}$)</td>
<td>11, 5.5, 2.2 and 1 Mbps</td>
</tr>
<tr>
<td>Control Rate ($C_{Rate}$)</td>
<td>1 Mbps</td>
</tr>
<tr>
<td>Retry Limit</td>
<td>7</td>
</tr>
<tr>
<td>TXOP Limit</td>
<td>0 (single frame transmissions only)</td>
</tr>
<tr>
<td>Beacon frame Period</td>
<td>100 ms</td>
</tr>
<tr>
<td>RTS/CTS Mode</td>
<td>OFF</td>
</tr>
<tr>
<td>Buffer Size</td>
<td>50 frames</td>
</tr>
</tbody>
</table>

3.5.3 Simulation Results

We present the throughput for AC1 and AC0, and the throughput proportion allocation for AC1 results on the same output graph. The AC1 Throughput Proportion results will be shown in the upper half of the graph using the right y-axis and common x-axis. The set of Throughput results are shown in the lower half of the graph, using the left y-axis and common x-axis. Legends for the AC1 Throughput Proportion and Throughput results are shown in the upper-right and the center-left position respectively. Looking at the graphed outputs, the Throughput results shown from top to bottom are Total Throughput, AC1 Throughput and AC0 Throughput.
3.5. PERFORMANCE EVALUATIONS

Stage - 1: Controlled UDP/CBR Traffic Environment

Comparison with an existing published scheme:

The control scheme used in [2] is not implementable without modification of the IEEE 802.11e standard for EDCA. However, it is useful to use this as a model for comparison as we show how to achieve essentially the same results, while strictly adhering to the IEEE 802.11e standard for EDCA using our EDCA compliant Control Scheme-1.

We evaluate together the following three specific configurations:

1. (A): Unchanged proposed scheme in [2].
2. (B): Closest exponentially encoded contention window values yielded from [2].
3. (C): Our proposed Control Scheme-1.

We set the required target throughput proportion allocation to be 50% for both AC0 and AC1. In addition we utilize a high bit rate UDP/CBR traffic profile for each source wireless station, as this is the form of traffic that is used by Banchs and Vollero in [2].
Therefore an accurate and realistic comparison can be performed between our proposed Control Scheme-1 and the scheme described in [2].

During the simulation period we change the number of active stations $n_1$ and $n_0$ according to Table 3.5.2. However, due to the nature of buffering in wireless client stations, this does not represent the actual number of the active wireless client stations at each time point. We account for this behavior by observing the queue size of each wireless client station in each beacon frame period, which allows us to accurately determine the true number of active wireless client station in the WLAN. However, it should be noted that this procedure is not applicable in an actual network implementation, since an AP is unable to determine this information.

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$n_0$</th>
<th>Time Duration (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td>25 - 35</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>35 - 45</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>45 - 55</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>55 - 65</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>65 - 75</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>75 - 85</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>85 - 95</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>95 - 105</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>105 - 115</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>115 - 125</td>
</tr>
</tbody>
</table>

The output graphs in Figures 3.5.2, 3.5.3 and 3.5.4, depict the throughput performance of the three configurations (A), (B) and (C) respectively.

In general, comparing the throughput section of each of the three configurations, the combined total throughput results are very similar to each other during the ten second period for each $n_1$ and $n_0$ combination. This indicates that the parameter values used in each case provide near identical total throughput performance and collision avoidance.

At certain intervals when one of the station numbers $n_1$ or $n_0$ is very small (i.e. 45 - 55s, 85 - 95s, 105 - 115s and 115 - 125s), the total combined throughput is slightly greater. This is simply due to the fact that the contention window size used becomes smaller when the number of active stations, $n_1$ or $n_0$, decreases. Since the contention window defines an additional amount of time that stations must spend not transmitting, this affects the level of throughput that can be achieved.
3.5. PERFORMANCE EVALUATIONS

The specific throughput performance for AC1 and AC0 in each configuration are quite different. In particular for configuration (B) seen in Figure 3.5.3, we are able to observe the limitation of the lack of available throughput proportion allocations due to the requirement of exponentially encoded values.

In the time intervals; 45 - 55s, 65 - 75s, 75 - 85s, 95 - 105s, 105 - 115s and 115 - 125s, the throughput for AC1 and AC0 is not near the required target of 50% throughput proportion allocation. This is due to the fact that, with these specific number of active wireless client stations $n_1$ and $n_0$ present in these particular time intervals, it is not possible to obtain a parameter combination that achieves a proportion allocation that is very near the required target 50%. Therefore this configuration is unable to effectively control the throughput proportion allocation.

In the absence of the EDCA prescribed restriction of only using exponentially encoded values. We can see in configuration (A), shown in Figure 3.5.2, the throughput performance for AC1 and AC0 is maintained such that the AC1 throughput proportion allocation of 50% is achieved and additionally it is relatively smooth. However, as we have discussed earlier such a configuration is not supported in the IEEE 802.11e standard [1].

Figure 3.5.2: Throughput Performance - (A) Integer based optimal configuration.
CHAPTER 3. THROUGHPUT CONTROL

The throughput performance for AC0 and AC1 achieved in configuration (C), is similar to that achieved in configuration (A), with the AC1 throughput proportion allocation also being very close to the required target 50%.

In configuration (C), we can see from the results in Figure 3.5.4 that we achieve a slightly less smooth and more variable level of performance compared to configuration (A). This is due to the fact that we continually switch from one parameter set to another in each beacon frame period, to remain an EDCA compliant control mechanism.

The average total throughput from the simulation results for the three configurations are (A) = 5.19Mbps, (B) = 5.15Mbps and (C) = 5.16Mbps respectively. As we can see the largest average total throughput is achieved by configuration (A), followed by our proposed solution (C) and then Configuration (B). This is attributed to the fact that in configuration (A) we utilize integer based parameter sets as opposed to exponentially encoded parameter sets in configuration (B) and (C) and so can maximize the throughput more finely.

In the case of the average AC1 throughput proportion allocation percentage, we obtain (A) = 50.2%, (B) = 49.8% and (C) = 50.1%. All three configurations register an average value close to the target 50% value. However, from the graphical simulation results
3.5. PERFORMANCE EVALUATIONS

in Figure 3.5.3, the AC1 throughput proportion allocation for configuration (B) is not consistently maintained over the whole simulation period.

The same observations can also be drawn with regards to comparing the performance between configuration (A) and (C). The graphical simulation results for configuration (A) in Figure 3.5.2, show an AC1 throughput proportion allocation that is smoother and more consistent than configuration (C).

In Table 3.5.3 we present the excursion performance metrics for each of the three test configurations (A), (B) and (C). Straight away we can see that the excursion performance of configuration is (A) is better than the other two configurations (B) and (C). All measured AC1 throughput proportion allocation points in configuration (A) are within ±5% of the target 50% value.

In the case of configuration (B), a total of 32 points are found to be at least 5% away from the target value. In configuration (C), only 7 points were found to be registering at least ±5% away from the target value.

It is clear that the performance of configuration (C) is significantly better than (B). The data presented in Table 3.5.3 shows clearly that configuration (B) registers the highest excursion measure out of the three tested configurations. As expected configuration (A) has the lowest and then followed by our Control Scheme-1. As for the throughput
performance all three cases achieve acceptable results, which are not drastically different from that reported by the analytical model results. Note that in the case of configuration (A), integer based parameter set configuration was used to evaluate the analytical model based throughput result. While for the remaining two configurations (B) and (C) exponentially encoded parameter set configurations were used.

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Excursion Performance (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>1.2</td>
<td>100.2</td>
</tr>
<tr>
<td>(B)</td>
<td>4.2</td>
<td>100.9</td>
</tr>
<tr>
<td>(C)</td>
<td>2.0</td>
<td>100.7</td>
</tr>
</tbody>
</table>

In conclusion our proposed allocation scheme configuration (C) is valid and performs comparably to the integer based optimal configuration (A). Our scheme also has the significant structural advantages of fitting strictly within the EDCA framework [1] while maintaining exponential back-off.

**Flexible throughput proportion allocation capability:**

In the following simulation results we show that our proposed throughput allocation scheme is also flexible and able to achieve various throughput proportion allocations between ACs. In this case, we optimize the network to achieve the following throughput proportion allocations with AC1 receiving 80% and 67%, respectively, of the total system throughput. During the simulation period we change the number of active stations $n_1$ and $n_0$ according to Table 3.5.2 shown earlier.

The results for the 80% case are shown in Figure 3.5.5 with the results for the 67% case shown in Figure 3.5.6. We can clearly see that our proposed allocation scheme is able to maintain, reasonably well, the throughput proportion allocations at the required target values.

The averaged simulation results for 80% and 67% test cases are present in Table 3.5.4. We can see for both test cases, the average AC1 throughput proportion allocation achieved are both very close to their respective target values.

We present in Table 3.5.5 the respective performance metrics for the 80% and 67% test case scenarios. Overall we can see that for both 80% and 67% target test case scenarios, the excursion performance values are both similar at 2.2% and 2.6% respectively. As for the throughput relative measure, both cases achieve results that are a reasonable match.
3.5. PERFORMANCE EVALUATIONS

Figure 3.5.5: Throughput Performance - 80% AC1 Throughput Allocation.

Table 3.5.4: Averaged Simulation Results.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Total (Mbps)</th>
<th>AC1 (Mbps)</th>
<th>AC0 (Mbps)</th>
<th>AC1 Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>5.29</td>
<td>4.17</td>
<td>1.12</td>
<td>78.8</td>
</tr>
<tr>
<td>67</td>
<td>5.24</td>
<td>3.47</td>
<td>1.77</td>
<td>66.3</td>
</tr>
</tbody>
</table>

to that reported by the mathematical model.

Table 3.5.5: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Excursion Performance (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>2.2</td>
<td>100.6</td>
</tr>
<tr>
<td>67</td>
<td>2.6</td>
<td>100.9</td>
</tr>
</tbody>
</table>
Incorrect starting point ($x_0$) recovery:

We verify our proposed throughput allocation scheme’s ability to recover from possible errors in the analytical model results, or more realistically from a different operational point than desired. We artificially create, an error for “each change point of $n_1$ and $n_0$” by initially setting the starting point parameter set ($x_0$) to achieve a greater throughput proportion allocation for AC1 than required.

During the simulation period we change the number of active stations $n_1$ and $n_0$ according to Table 3.5.2 shown earlier. The actual required throughput proportion allocation in this case is 50% allocation of total system throughput to AC1.

In Figure 3.5.7, we can see that the deliberate introduction of an incorrect starting point ($x_0$) at “each change point of $n_1$ and $n_0$” results in a sharp increase in the throughput proportion allocation for AC1 at each station change interval before correct operating conditions are achieved. In addition, the duration of these large increases in AC1 throughput proportion allocation can be extended, especially when the effective number of wireless client stations is reduced from the previous time interval. This is due to the fact that at each station change interval, there are active wireless client stations present from the previous interval which still have data frames present in their transmission queue.
Hence, as the buffers of these wireless client stations are being emptied, the number of active wireless client stations will change. **Control Scheme-1** uses the incorrect starting point parameter set \((x_0)\) each time the active station count changes, this resulting in an extended duration of the large increase in **AC1** throughput proportion allocation being observed. Once the number of active wireless client stations remains constant **Control Scheme-1** is then able to recover and determine the correct operating region.

Therefore, even subject to these extreme conditions, our parameter switching technique is able to recover and operate in the correct region to achieve the required throughput proportion. The time for recovery to the correct operating region is minimal, since decisions and parameter set changes occur at each beacon interval, making our proposed allocation scheme able to quickly overcome any inconsistencies that may be present.

In Table 3.5.6 we present the average simulation results for Figure 3.5.7. We can see that the large excursions have had a considerable impact on the average result of **AC1** throughput proportion allocation. The average value clearly is now more than 5% away from the target value, due to our forced increase of **AC1** throughput proportion allocation at “each change point of \(n_1\) and \(n_0\).” Thus we can also see in Table 3.5.7 that the excursion performance result reflects the impact of these excursions by reporting a
value of 7.1%. The model throughput comparison registers a slightly lower value, this is due to the incorrect starting position being used at each station number change.

Table 3.5.6: Averaged Simulation Results.

<table>
<thead>
<tr>
<th>Target Proportion</th>
<th>Total</th>
<th>AC1</th>
<th>AC0</th>
<th>AC1 Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>(%)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(%)</td>
</tr>
<tr>
<td>50</td>
<td>5.19</td>
<td>2.87</td>
<td>2.32</td>
<td>55.4</td>
</tr>
</tbody>
</table>

Table 3.5.7: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target Proportion</th>
<th>Excursion Performance (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(%)</td>
<td>7.1</td>
<td>98.9</td>
</tr>
</tbody>
</table>

Stage - 2: Realistic TCP/FTP Traffic Environment

Throughput Proportion Allocation Under TCP/FTP Traffic:

In the following test case scenario, we utilize a homogeneous traffic profile for AC1 and AC0, all wireless client stations have the same transmission rate of 11Mbps. Likewise, during the simulation period we change the number of active stations $n_1$ and $n_0$ according to Table 3.5.2 shown earlier. We assign the required throughput proportion allocation to an 80% allocation of total system throughput to AC1.

We verify that our proposed allocation scheme is able to maintain the correct throughput proportion allocation when TCP/FTP based traffic is used for AC1 and AC0, with the wireless client stations as the TCP/FTP sources. This form of TCP/FTP configuration may appear unusual since in a WLAN, the wireless client stations are usually the TCP/FTP destination nodes. We chose this configuration primarily for the reason that the wireless client stations are more likely to exhibit saturation load conditions since they are effectively responsible for continuously transmitting TCP data packets. While, if the more common configuration of wireless client stations acting as the TCP destination nodes is used, the return TCP-ACK packets are unlikely to generate saturation load conditions within each wireless client station. However, having wireless client stations operating as TCP/FTP sources is certainly not unrealistic, and we will address the other configuration in Chapter 4 onwards.
3.5. PERFORMANCE EVALUATIONS

We assume the *TCP/FTP sink* node is connected to the *AP* via a wired connection with a link speed of 1000Mbps and a one-way propagation delay of 50ms. The receive window for each *TCP* connection is set to a large value so that the *TCP* congestion window is the sole factor in determining the effective throughput performance. The *TCP* congestion control algorithm used in our test case scenario is the *Reno* algorithm with *Selective Acknowledgment (SACK)* [49, 50].

An important aspect in dealing with *TCP* traffic in our test case scenario is determining an appropriate buffer size to implement for each wireless client station such that the throughput performance is maximized. Choosing a small buffer size can artificially limit the rate of a *TCP* flow. The *TCP* congestion control will effectively be disabled as the limited buffer size does not allow the *TCP* congestion window to adjust according to link capacity and maximize throughput performance. Instead the limited buffer capacity effectively hard limits the rate at which *TCP* can send data.

In our test case scenario, we calculate that the required buffer size for each wireless client station simply using the well known *Bandwidth Delay Product (BDP)* [51] shown in Eq. (3.5.3).

\[
BDP = \text{available bandwidth (bit/sec)} \times \text{round trip time (sec)},
\]

where:
- \(\text{available bandwidth (bit/sec)}\) is the available bandwidth of the bottleneck link along the *TCP* connection path.
- \(\text{round trip time (sec)}\) is the duration of time between a *TCP* data packet being sent and its acknowledgment (*TCP-ACK*) being received.

We first determine the effective transmission capability of the *WLAN*. In Eq. (3.5.4), we evaluate the transmission time for a single data frame \(T_s\). Note that the transmission time is strictly the amount of time required to physically transmit a data frame over the air within the *WLAN*. As a result the inherent delay incurred through frame queuing and random exponential back-off is omitted.

\[
T_s = \frac{\text{PHY}}{C\_Rate} + \frac{\text{MAC} + \text{Payload} + \text{FCS}}{D\_Rate} + \text{SIFS} + \frac{\text{PHY} + \text{ACK}}{C\_Rate} + \text{AIFS}
\]

Where:
- \(\text{Payload}\) is the payload size (in bits) carried within the transmitted data frame.

In the *WLAN*, the wireless transmission channel is shared between the transmission of *TCP* data frames and *TCP-ACK* frames. As a result, we assume that in the long term
for every TCP data frame transmitted there will also be a TCP-ACK frame transmitted. In our initial calculations we also assume that the bottleneck link is the WLAN.

Using Eq. (3.5.4), we can evaluate the transmission time for a TCP data packet and a TCP-ACK packet. We assume each TCP data packet consists of the full TCP 60 byte (480 bits) header and a payload of 1000 bytes (8000 bits), while each TCP-ACK packet is a fixed size of 40 bytes (320 bits).

In Table 3.5.8, we present the transmission time \( T_s \) for a TCP data frame and TCP-ACK frame. Using this information we then evaluate the achievable throughput in the WLAN. As mentioned earlier, in our calculations we omit the transmission delays incurred through queuing and random back-off in EDCA, therefore the throughput results shown in Table 3.5.8 are an overestimated value.

Table 3.5.8: WLAN TCP Transmission time and Throughput Information.

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_s ) TCP Data Frame (seconds)</td>
<td>0.00135</td>
</tr>
<tr>
<td>( T_s ) TCP-ACK Frame (seconds)</td>
<td>0.000611</td>
</tr>
<tr>
<td>Total Time (seconds)</td>
<td>0.00197</td>
</tr>
<tr>
<td>Throughput (frames/sec)</td>
<td>509</td>
</tr>
<tr>
<td>Throughput TCP Data (Bits/sec)</td>
<td>4316442</td>
</tr>
<tr>
<td>Throughput TCP-ACK (Bits/sec)</td>
<td>162884</td>
</tr>
</tbody>
</table>

Based on the throughput (frames/sec) from Table 3.5.8 and the round trip time (100ms), we apply the BDP rule from Eq. (3.5.3), which returns a buffer size of 50 frames. We assign the wireless client stations and AP a buffer size of 50 frames, which appears to be more than adequate to allow TCP to adjust its rate of transmission such that it can as required saturate the transmission channel.

The TCP/FTP traffic throughput performance, shown in Figure 3.5.8, appears fairly inconsistent across the whole simulation period. However, in most cases the AC1 throughput proportion allocation remains close to the 80% target value, so that in terms of controlling the AC1 throughput allocation it appears our Control Scheme-1 is performing reasonably well. However there is an underlying anomaly that is present which is not visible in Figure 3.5.8.

This anomaly can be seen when we look at the TCP congestion window plot for each AC1 and AC0 flow, in Figure 3.5.9 - (A) and (B) respectively. The key observation is only a few AC1 and AC0 TCP flows are actually active and are able to increase their respective TCP congestion window values. This is clearly evident if we look at the time interval 65 - 75s. In this interval we expect to see 10 AC1 and 4 AC0 active TCP flows.
However from the congestion window results only 3 AC1 and 2 AC0 TCP flows appear to get the opportunity to transmit and effectively increment their respective congestion window values. This level of performance is clearly not satisfactory, and the primary reason for this anomaly within the WLAN is related to the transmission of TCP-ACK frames.

![Figure 3.5.8: Throughput Performance - TCP/FTP traffic sources.](image)

It is clear that TCP relies on return TCP-ACKs to control the increment of the congestion window and therefore the transmission rate of TCP data frames. In our test case scenario, the AP is solely responsible for transmitting all the return TCP-ACK frames back to each wireless client station that is a TCP/FTP source node. In addition, all return TCP-ACK frames from the TCP/FTP sink node are classified as the same access category as their respective source TCP/FTP flow. The AC1 and AC0 EDCA parameter set used by the AP is the same as every AC1 and AC0 wireless client stations. Therefore the AP competes for channel access to transmit TCP-ACKs with the TCP data frames from both AC1 and AC0 wireless client stations.

As a result TCP-ACK frame transmissions can be delayed due to collisions with TCP data frames, or the transmission channel simply being busy. In addition TCP-ACK frame losses can occur due to exceeding the MAC layer retransmission limit, or simply due
to a buffer overflow within the AP. Delayed or lost \textit{TCP-ACK} frames inevitably affect the round trip time calculation performed by \textit{TCP}. This in turn results in a reduced transmission rate of \textit{TCP} flows.

In Figure 3.5.10, we can see that the throughput characteristic of the \textit{TCP-ACKs} alone
is variable and fairly inconsistent. Therefore, due to delays in transmission of TCP-ACKs frames we observe the variable throughput performance seen in Figure 3.5.8.

Since the AP receives the same transmission opportunity as each wireless client station, when the number of active wireless client stations increases, the AP receives less opportunity to transmit data frames. Since the AP is responsible for transmitting TCP-ACK frames, large delays in transmission can occur and thus inevitably result in flows suffering timeouts. In Figure 3.5.9, we effectively see that this restriction causes certain flows to simply cease transmission and their TCP congestion window do not increase. This behavior is particularly severe for newly activated TCP flows.

This issue of unfairness and degraded performance of TCP traffic has been extensively studied and documented for wireless LANs [52, 53, 54]. In [52, 53, 54], the TCP/FTP connections are configured such that the wireless client stations are the destination nodes. Therefore this effect is less pronounced since each wireless client station competes to transmit their own TCP-ACK frames for its own TCP flow. While in our configuration the AP is transmitting all the TCP-ACK frames for each TCP flow in the network.

With the advent of IEEE 802.11e, the EDCA access mechanism allows reconfigurable channel access parameters for each AC and through the HCCA access mechanism, a means for providing controlled collision free transmissions. As a result, we believe that uplink (UL) TCP throughput performance can be improved using two alternative methods in an IEEE 802.11e based WLAN.

To begin with in the AP, we first need to identify the TCP-ACK frames. A simple approach is to identify them given their small fixed size, or through the slightly more complicated method of extracting this information directly from the TCP header.

The first method involves assigning a separate access category queue solely for TCP-ACK frames. Since we assume the EDCA mechanism is used for Non Real-time traffic, we can assign within the AP, either AC3 or AC2 access category specifically for TCP-ACK frames. We define the TCP-ACK access category to receive prioritized transmission channel access over AC1 and AC0 traffic. In order to completely negate the TCP-ACK access category competing for transmission channel access, we assign an EDCA parameter set of AIFSN value of one. In addition we also omit the random back-off process by assigning the $CW_{\text{min}}$ and $CW_{\text{max}}$ value to zero. As a result of this configuration, collision free transmission of TCP-ACK frames will occur within the WLAN.

The first method explained above only allows prioritized TCP-ACK frames transmission for UL traffic, where the TCP sources are the wireless stations. Implementing the same technique in the scenario where wireless client stations are the TCP destination node is not a viable solution. As a result, a good solution is to classify all TCP-ACK frames as high priority AC1 within each wireless client station. In addition, arriving TCP-ACK
frames should be placed at the head of the AC1 queue to reduce delays in transmission.

The key factor in the configuration described above is that all wireless client stations must support this feature. Therefore wireless client stations that perform the operation discussed above will receive a performance advantage over those that do not. As a result, in order to maintain fairness and more importantly require minimal implementation requirements within wireless client stations we do not pursue this course of action in our proposed control mechanism.

The second method is to use HCCA to transmit TCP-ACK frames in both UL and downlink (DL) direction. Therefore contention free transmission of TCP-ACK frames can take place from both the AP and wireless client stations. This can be achieved by allowing at the start of each beacon frame period through the operation of allowing the AP to poll all wireless client stations individually to allow the transmission of TCP-ACK frames in the UL direction. Note that in the case of DL direction the AP can grab the channel and transmit its own TCP-ACK frames whenever required due to the fact of its short wait period designated by a PIFS.

![Throughput Performance - TCP/FTP traffic sources - HCCA Transmission TCP-ACK.](image)

Figure 3.5.11: Throughput Performance - TCP/FTP traffic sources - HCCA Transmission TCP-ACK.

We re-simulate the same test case scenario this time using the second method of
3.5. PERFORMANCE EVALUATIONS

Figure 3.5.12: TCP Congestion Window - (A) AC1 Wireless Client Stations and (B) AC0 Wireless Client Stations.

Figure 3.5.13: Throughput Performance - TCP-ACK Transmissions - HCCA Transmission TCP-ACK.

utilizing HCCA for TCP-ACK frames transmission within the AP. We can see straight away in Figure 3.5.11 that the overall TCP throughput performance is more consistent over the whole simulation period as compared to Figure 3.5.8. Likewise, in Figure 3.5.13, the transmitted TCP-ACK throughput is also more consistent across the whole simulation
period as compared to Figure 3.5.10.

The critical and major advantage in using our proposed HCCA transmission of TCP-ACK frames can be seen in the related TCP congestion window results in Figure 3.5.12. The guaranteed transmission of TCP-ACK frames with minimal delays has allowed all TCP flows the opportunity to transmit and gain their required share of the throughput. We can effectively see that the required number of active AC1 and AC0 TCP flows are now truly active in each designated time interval across the whole simulation period.

In Table 3.5.9 we present the averaged simulation results for both Figure 3.5.8 and 3.5.11. The average total throughput achieved is slightly larger for the non HCCA TCP-ACK transmission scheme. However it must be stressed that the real advantage in our proposed HCCA TCP-ACK transmission scheme is the clearly improved TCP flow performance and fairness as seen in Figure 3.5.12. In the case of the average AC1 throughput proportion allocation, both configurations are on par with each other.

<table>
<thead>
<tr>
<th>Target Proportion</th>
<th>Total (Mbps)</th>
<th>AC1 (Mbps)</th>
<th>AC0 (Mbps)</th>
<th>AC1 Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80 (non HCCA TCP ACK)</td>
<td>3.97</td>
<td>3.11</td>
<td>0.86</td>
<td>78.3</td>
</tr>
<tr>
<td>80 (HCCA TCP ACK)</td>
<td>3.90</td>
<td>3.05</td>
<td>0.84</td>
<td>78.3</td>
</tr>
</tbody>
</table>

The configuration without the aid of HCCA delivered TCP-ACK and our proposed HCCA TCP-ACK delivery method both record displacement percentages of around 2.6%.

In terms of the model comparison for our proposed scheme using HCCA, we modify the model to simply include that for each frame transmission a corresponding TCP-ACK is also recorded. This crude alteration is reasonably sound given that TCP-ACK transmissions are effectively being guaranteed.

In the case of the Baseline scheme the model cannot be evaluated given that the analytical model is unable to capture and explain the behaviour of TCP-ACK frame transmissions competing with existing data frames within the same AC queue. In fact to the best of our knowledge, none of the currently available analytical models are able to truly capture and describe the performance characteristics when TCP traffic is present in a WLAN operating under the EDCA mechanism.

Overall we can conclude that with the aid of HCCA TCP-ACK delivery mechanism is able to maintain the required AC1 throughput proportion allocation. The proposed use of HCCA TCP-ACK delivery is necessary to allow us to retain the fair sharing principle between all users within each AC.
3.5. PERFORMANCE EVALUATIONS

Table 3.5.10: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Excursion Performance (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80 (non HCCA TCP ACK)</td>
<td>2.6</td>
<td>N/A</td>
</tr>
<tr>
<td>80 (HCCA TCP ACK)</td>
<td>2.6</td>
<td>104.1</td>
</tr>
</tbody>
</table>

Heterogeneous traffic and multi-rate wireless client stations:

In the following test we verify that our proposed throughput allocation scheme is able to achieve the correct throughput proportion allocations when a heterogeneous traffic profile is used for AC1 and AC0 and the transmission rates for wireless client stations in AC1 and AC0 are different. Note that from now on we utilize our proposed HCCA TCP-ACK delivery mechanism.

![Figure 3.5.14: Throughput Performance - Heterogeneous traffic and Multi-rate sources.](image)

We assign all AC1 and AC0 traffic to be TCP/FTP, with the exception of the packet size of AC1 traffic to be 250 bytes and the packet size of AC0 traffic to be 1000 bytes. All wireless client stations in AC1 are able to transmit at 2Mbps and the wireless client
stations in \( AC_0 \) are able to transmit at 11Mbps. During the simulation period we change the number of active stations \( n_1 \) and \( n_0 \) according to Table 3.5.2 shown earlier. The required throughput proportion allocation used in this case is a 50% allocation of total system throughput to \( AC_1 \).

The results shown in Figure 3.5.14, show that our proposed allocation scheme is still able to maintain the required throughput proportion allocation. The combined total throughput is reduced since the overall combined physical transmission capability of \( AC_1 \) wireless client stations is reduced.

The required throughput proportion allocation between \( AC_1 \) and \( AC_0 \) is set to 50%. Therefore the transmission rate of \( AC_1 \) governs the effective throughput performance. Since with a 50% allocation, \( AC_0 \) is only able to transmit effectively the same amount of data that \( AC_1 \) transmits. Thus with \( AC_1 \) traffic transmission physically limited to 2Mbps, this is roughly 20% of the physical transmission rate \( AC_0 \) is capable of at 11Mbps. Thus \( AC_1 \) transmits for approximately 80% of the available time because its transmission rate is around 20% of that of \( AC_0 \).

The average simulation results for Figure 3.5.14 are shown in Table 3.5.11. We can see that the total average throughput is as expected restricted by the physical transmission rate of \( AC_1 \) wireless client stations. The average \( AC_1 \) throughput proportion allocation is close to the target value.

Table 3.5.11: Averaged Simulation Results.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Total (Mbps)</th>
<th>AC1 (Mbps)</th>
<th>AC0 (Mbps)</th>
<th>AC1 Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1.33</td>
<td>0.69</td>
<td>0.64</td>
<td>52.1</td>
</tr>
</tbody>
</table>

We can also see from the performance metrics results in Table 3.5.12, that even with Heterogeneous traffic and Multi-rate sources we are still able to achieve an acceptable performance level.

Table 3.5.12: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Excursion Performance (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>4.5</td>
<td>103.3</td>
</tr>
</tbody>
</table>
Influence of HCCA Real-time traffic:

In the following test, we verify that our proposed throughput allocation scheme is able to maintain the correct throughput proportion allocations even if the contention period varies over time due to the presence of variable Real-time traffic serviced by HCCA. Thus far, in the previous tests, we have ignored the effects of Real-time traffic.

We vary the effective load and duration that HCCA is active in order to model the variable traffic characteristic of Real-time traffic. We assume a homogeneous traffic profile of TCP/FTP traffic for both AC1 and AC0. All wireless client stations are once again assigned the maximum physical transmission rate of 11Mbps. The required throughput proportion allocation used in this case is an 80% allocation of total system throughput to AC1.

During the simulation period we change the number of active wireless client stations $n_1$ and $n_0$ according to Table 3.5.2. In addition, we utilize within the AP our modified EDCA mechanism to provide contention free transmissions of TCP-ACK frames. We maintain the same buffer size of 50 frames for both the AP and wireless client stations.

In Figure 3.5.15, we display the throughput of the Real-time traffic being serviced by HCCA. In Figure 3.5.16, the throughput performance of our scheme is shown under the influence of the HCCA Real-time traffic with the total throughput now being for (EDCA + HCCA Real-time). The transmitted TCP-ACK throughput is shown in Figure 3.5.17 and the congestion windows in Figure 3.5.18.

The results show that our proposed scheme is able to maintain reasonably well the desired throughput proportion values even under the influence of very bursty HCCA Real-time traffic. The available transmission time for the EDCA mechanism varies and the
throughput performance reflects this fact.

The overall throughput performance is not as consistent when compared to the results shown in Figure 3.5.11. This is due to the fact that TCP/FTP traffic is being serviced in the presence of HCCA Real-time traffic. The variable HCCA Real-time traffic causes
the throughput performance of the TCP/FTP traffic serviced by EDCA to fluctuate.

The proportion results in general lie close to the target 80% value, however there are more excursions as compared to Figure 3.5.11. This is due to the presence of HCCA Real-time traffic, and in particular due to the change in the traffic load for HCCA Real-time traffic.

These excursions can be seen to be quite extreme during the time interval 105 - 115s and 115 - 125s. From Table 3.5.2, we can see that during these two time intervals, one AC1 wireless client station will be active, while four and then five AC0 wireless client stations will be active. As a result there are clearly more active AC0 wireless client stations than AC1 wireless client stations, and yet AC1 is required to achieve 80% of the available throughput.

In Figure 3.5.18 (A) and (B), we plot the TCP congestion window ($cwnd$) values for each AC1 and AC0 TCP flow respectively. As we can see in Figure 3.5.18 (A), in the time intervals 105 - 115s and 115 - 125s only a single TCP flow is active as expected. The important observation we can make is that three TCP timeouts were registered in the aforementioned time intervals. Hence the transmission rate of the single AC1 TCP flow is significantly reduced. Since there is a large number of AC0 flows in comparison, the throughput improvement in AC0 is clearly observable. The excursions observed are a result of the correct behavior of the TCP protocol and a single TCP/FTP source.

![Figure 3.5.18: TCP Congestion Window - (A) AC1 Wireless Client Stations and (B) AC0 Wireless Client Stations.](image-url)
The average simulation results also reflect similar findings. In Table 3.5.13, the total throughput achieved is much lower than seen in Table 3.5.9, this is to be expected due to the presence of Real-time traffic. In addition, we can see that the average AC1 throughput proportion allocation achieved falls below the target value by over 5%. This is primarily due to the large excursion near the end of the simulation as explained above.

Table 3.5.13: Averaged Simulation Results.

<table>
<thead>
<tr>
<th>Target Proportion</th>
<th>Total (Mbps)</th>
<th>AC1 (Mbps)</th>
<th>AC0 (Mbps)</th>
<th>AC1 Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>2.75</td>
<td>2.06</td>
<td>0.69</td>
<td>74.8</td>
</tr>
</tbody>
</table>

The effects of these excursions can also be seen in the performance metrics results in Table 3.5.14. The excursion performance registered is over 7%. However, as we discussed earlier this is a result of the TCP protocol used and single TCP/FTP source and not an impact caused by our Control Scheme-1.

Given the lack of HCCA support in the analytical model being used, we first of all evaluate approximately, the bandwidth/data rate available for servicing EDCA traffic. We do this by simply stating that a certain proportion of the total wireless bandwidth/data rate will consistently be utilized by whatever HCCA based traffic is present in the WLAN. Thus the remaining available bandwidth/data rate will then be used to towards servicing EDCA based traffic.

We use a heavily approximated approach by stating that the observed 1.6187 Mbps average total throughput of HCCA traffic in this single simulation run will be used as the scaling factor for calculating what the remaining available bandwidth/data rate is present for servicing EDCA based traffic. Given the observations are just from one simulation run alone and our approximation methods, we do expect some discrepancies in the registered model comparison value.

Table 3.5.14: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Excursion Performance (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>7.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

We conclude that our proposed scheme is still effective when TCP does not adversely affect the system with very few connections in progress. That is, it is able to maintain the
required throughput proportion even if there is variable *Real-time* traffic in the network being served by *HCCA*.

**Impact of competing DL and UL traffic:**

In the following test we simulate the scenario where network traffic consists of both UL AC1 and AC0 TCP/FTP traffic generated from wireless client stations and DL AC1 and AC0 TCP/FTP traffic being transmitted by the *AP*. We again utilize the *HCCA* mechanism as discussed earlier to serve both UL and DL *TCP-ACK* transmissions.

The network configuration simulated in this test consists of *five* wireless client stations that act as the UL traffic source nodes. Each wireless client station consists of both an AC1 and AC0 access category TCP/FTP flow. As in the previous simulated test cases, the destination node for the UL traffic is connected to the *AP* via a wired link with a propagation delay of 50ms and a link speed of 1000Mbps. In case of the DL traffic source node, we utilize the existing wired node connected to the *AP* to transmit one AC1 and AC0 categorized TCP/FTP flow to each of the wireless client stations.

All wireless client stations are assigned the maximum physical transmission rate of 11Mbps. We specify the required throughput proportion allocation for the simulation to be 80% allocation of total system throughput to AC1.

As we can see in Figure 3.5.19, the inclusion of competing DL traffic does not impact on maintaining the required throughput proportion allocation between AC1 and AC0 access categories. However, the important result that is of interest, is the throughput sharing between the DL and UL.

In Figure 3.5.20, we can see that there is a definite throughput sharing anomaly between the contending DL and UL traffic. This is due to the fact that in our current *Control Scheme-1*, both the *AP* and wireless client stations utilize the same parameter configuration for each access category.

As a result of this, we end up with the scenario of *n* wireless client stations and the *AP* contending for channel access, with each obtaining $1/(n+1)$ share of the total transmissions. Therefore, this results in $n/(n+1)$ of the transmissions to be specifically for the UL, and consequently $1/(n+1)$ of the transmissions for the DL. This is the throughput sharing anomaly that can be seen in Figure 3.5.20. As a result, in our test case scenario with *five* wireless client stations and a single *AP*, the throughput achieved for the DL path is approximately 16% as expected. The exact average value achieved is shown in Table 3.5.15.

We again conclude that current control configuration *Control Scheme-1* is able to achieve the desired throughput proportion allocation only on an overall network access category basis. In Chapter 4, we modify and extend our control to allow the additional
Table 3.5.15: Averaged Simulation Results.

<table>
<thead>
<tr>
<th>Target Proportion (%)</th>
<th>Total (Mbps)</th>
<th>DL to UL (%)</th>
<th>DL AC1 (%)</th>
<th>UL AC1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>3.79</td>
<td>17.5</td>
<td>72.6</td>
<td>76.9</td>
</tr>
</tbody>
</table>

feature of specifying throughput proportion allocation for both the DL and UL.

Figure 3.5.19: Throughput Performance - Combined UL and DL AC1 and AC0 traffic.

### 3.6 Summary

In this chapter we have developed and discussed an effective control mechanism that is able to achieve the goal of providing flexible throughput proportion allocation for ACs, provided that the network is operating under saturation load. Notably, the control fits strictly within the standard EDCA framework, and we achieve these proportion allocations while maximizing the overall throughput performance of the network.

The proposed control mechanism, Control Scheme-1 is relatively simple, and only
3.6. SUMMARY

Figure 3.5.20: Throughput Performance - Individual UL and DL traffic.

requires the input of an analytical model of the EDCA mechanism, of which there are many already in existence [2, 20, 21, 22, 23, 24, 26, 27, 28]. Additionally the procedure is not tied directly to a specific analytical model and is able to deal with inaccuracies within these models.

The selection criteria for the parameters is very simple and is based on the intuitive understanding of the behavior of the access channel when exponentially encoded parameter values are used in the EDCA mechanism. Furthermore, the selection criteria is capable of determining the required configuration to maintain a desired throughput proportion even if station transmission rates and data frame sizes are different.

Control Scheme-1 also maintains the important feature of exponential back-off for added operational robustness. However, most importantly, the use of the parameter switching technique allows Control Scheme-1 to overcome the restrictions imposed by exponential encoding of parameter values and still achieve tight throughput proportion results around the target value. It also has the inherent capability of overcoming any inconsistencies and errors in the starting point selection process ($x_0$).

We have provided various simulation results that clearly show how our proposed scheme operates and its effectiveness in achieving our specified goals in a variety of scenar-
io.s. The key importance, as mentioned, for the correct operation of our proposed Control Scheme-1, is that the network must be under saturation load conditions.

In the majority of test case scenarios discussed, we specifically observe network traffic in the UL direction alone. In these scenarios our Control Scheme-1 is able to effectively maintain the required throughput proportion allocation between access category AC1 and AC0. However, with the introduction of competing DL traffic in the last test performed in Stage - 2, we observe that the current configuration of our Control Scheme-1 is unable to provide throughput allocation between the UL and DL directions. In fact, our Control Scheme-1 (as with any default EDCA configuration [55, 56, 57, 58, 59, 60, 61, 62]) will always discriminate against the DL, which is typically regarded as the higher volume Non Real-time traffic direction. In Chapter 4, we address this issue by extending our existing Control Scheme-1 configuration to provide the ability to define UL and DL throughput proportion allocations, along with the preexisting access category based throughput proportion allocation.
Chapter 4

Differentiating Uplink and Downlink

4.1 Introduction

An important characteristic of Control Scheme-1 in Chapter 3, is that throughput proportion allocation is only viewed on an access category (AC) level. Therefore, traffic in the downlink (DL) or uplink (UL) direction that belong to the same AC are considered the same. This means, both wireless client stations and the access point (AP) utilize the same EDCA transmission parameter set for each respective AC.

In Chapter 3, we highlighted and briefly discussed the throughput sharing anomaly that exists between DL and UL traffic. The use of the same EDCA transmission parameter set configuration for both wireless client stations and the AP results in a distinct throughput sharing disadvantage for DL traffic transmitted by the AP.

In this chapter, we define additional goals which are to achieve proportion control between the DL and UL while maintaining AC1 to AC0 proportions within the DL and UL on an individual basis. This is then followed by a discussion of some of the related literature, which addresses the issue of DL and UL throughput unfairness. We then proceed to a detailed description of an updated EDCA parameter set selection process, describing the modifications required to the Control Scheme-1 and present some simulation results which demonstrate the effectiveness of these modifications.

4.2 DL and UL Throughput Anomaly

In Chapter 3, we developed Control Scheme-1 which operated within the EDCA standard framework to provide throughput proportion allocation sharing between the two ACs defined for Non Real-time traffic. Through the various simulation results shown in Chapter 3, Control Scheme-1 was shown to be effective in achieving this goal.
Note that Control Scheme-1 only specified throughput proportion allocation on an AC basis, which led to a throughput sharing asymmetry between overall DL traffic from the AP and overall UL traffic from wireless client stations. This is a direct result of both wireless client stations and the AP utilizing the same parameter set configuration for each AC. That is, when both wireless client stations and the AP continually contend for access to the wireless transmission channel, each wireless client station and the AP will obtain approximately $1/(n+1)$ (where $n =$ number of wireless client stations) share of the total transmissions and likewise network throughput. This results in an $n/(n+1)$ share of transmissions and network throughput for UL traffic, while the DL traffic through the AP only receives a $1/(n+1)$ share of transmissions and network throughput.

The asymmetry becomes more pronounced as the number of wireless client stations $n$ increases and so we modify the current Control Scheme-1 configuration in order to address the asymmetry between DL and UL traffic.

### 4.3 Extended Throughput Allocation Criterion

In Chapter 3, we defined specific goals and requirements that Control Scheme-1 aimed to achieve within an IEEE 802.11e EDCA infrastructure based wireless local area network (WLAN). These are highlighted below for convenience.

1. The ability to define specific throughput proportion allocation under saturation load between Non Real-time designated ACs AC1 and AC0.

2. Throughput performance should be maximized while maintaining the specific throughput proportion allocations.

3. Throughput proportion allocations should be effectively maintained within as close as possible to the required allocation percentages.

4. These requirements should be achieved within a time scale of one second.

In light of the throughput sharing anomaly between DL and UL traffic, additional modifications are critical to Control Scheme-1. We aim to achieve the following additional goals and requirements in our enhanced control, Control Scheme-2.

5. Ability to define and control overall throughput proportion allocations between DL and UL traffic under saturation load conditions.

6. Ability to define and independently control AC based throughput proportion allocation within DL and UL traffic under saturation load conditions.

In the next section we discuss some of the literature that is closely related.


4.4 Related Work

The throughput sharing anomaly discussed earlier between DL and UL traffic within WLANs is a well known and documented issue [55, 56, 57, 58, 59, 60, 61, 62]. The majority of related work in the literature that address the throughput sharing anomaly can be classified into two groups.

The first group [55, 60, 61] primarily aim to overcome the throughput sharing anomaly within the WLAN without altering the wireless transmission channel access mechanism. The second group [56, 57, 58, 59, 62] mainly propose altering the wireless transmission channel access between wireless client stations and the AP. In the first group, the primary focus is on a TCP based traffic environment for both the DL and UL, and the majority of techniques proposed deal with modifying the behavior of TCP traffic.

In [55, 61], they directly change the TCP advertised receiver window size value within the TCP-ACK packets for flows in both the DL and UL at the AP. The required TCP receiver window size is evaluated analytically in [55, 61] based on the number of DL and UL TCP flows within the WLAN. A hard limit can thus be enforced on the transmission rate of DL and UL TCP flows. This approach attempts to eliminate packet losses by restricting the amount of data packets the TCP sources can transmit. However, packet losses due to wireless transmission channel collisions and errors are still an issue.

Implementing the proposed approach of [55, 61] introduces more complexity within the AP, since for each individual TCP-ACK packet transmitted, the AP must edit its TCP headers in order to change the receiver advertised window size to the required value. Additionally, the AP must estimate in real-time, the number of TCP flows within the network in order to calculate the required value of the receiver window size for each TCP flow. Furthermore, another critical issue is being able to distinguish between highly active and minimally active TCP flows. The analytical results in [55, 61] assume all TCP flows are highly active when calculating the required individual TCP flow receiver window size.

In both [55, 61], fair sharing of throughput between DL and UL flows is considered as the primary aim. Therefore, defining a different overall throughput proportion allocation between DL and UL traffic is not considered. It should also be noted that the solutions proposed in [55, 61], violate the end-to-end control principles of TCP.

In [60], TCP UL traffic is artificially rate limited through the use of a Token Bucket Filter (TBF) mechanism implemented within the AP. The TBF within the AP simply restricts the rate of outgoing frames into the wired network that the AP receives from the wireless client stations. The AP is only able to forward a given number of frames based on the available tokens in the bucket at any given time, so that frames are dropped because no queuing mechanism is in place for outgoing frames. The expected result of the TBF is to indirectly control the rate of UL TCP traffic via its congestion control
mechanism.

The proposed TBF approach is quite simple to implement and is shown to overcome the throughput asymmetry between TCP UL and DL traffic in certain test case scenarios. In [60], two specific implementations were presented for the TBF, a static rate and a dynamic rate adjustment for TCP traffic in the UL direction.

The use of a static rate value enforces a hard limit on the throughput for TCP UL traffic, this can cause under utilization of the network when DL TCP traffic is minimal. In the case of the dynamic rate adjustment solution, this problem is addressed by first observing if any frames are dropped from the DL buffer within the AP at a given time interval \( T_p \). If a frame drop is observed within \( T_p \), the UL traffic rate is reduced by a specific fixed value \( R_{\text{step}} \), conversely if no frames are dropped the UL traffic rate is increased by \( R_{\text{step}} \). As a result this follows a simple additive increment/decrement process.

An important issue highlighted in [60] is that throughput oscillations can occur when particular rate adjustment values, \( R_{\text{step}} \), are used. However, methods for determining the optimal \( R_{\text{step}} \) value that reduces oscillations while providing adequate performance and determining the optimal time interval \( T_p \) are not addressed. The TBF approach is specifically aimed at controlling UL traffic only and as a result no direct control is provided for DL traffic.

The TBF mechanism can be considered an inefficient approach at addressing the throughput sharing anomaly between DL and UL because as described earlier, the TBF mechanism is designed to drop frames received from wireless client stations. The end result is an inefficient use of the WLAN transmission channel, an issue which is overlooked and not investigated in detail in [60].

The second group [56, 57, 58, 59, 62] as mentioned earlier directly deals with manipulating the wireless transmission channel access behavior of wireless client stations and the AP. In [56], the throughput sharing between UL and DL traffic is addressed solely within the AP. The proposed approach is to allow the AP prioritized contention free transmission channel access after a Point Inter-frame Space (PIFS) interval. This form of channel access occurs whenever the AP determines the throughput sharing between UL and DL to be above a specified threshold. Such an approach is trivial to implement, however the use of transmissions after a PIFS is only effective in a single AC environment such as the basic IEEE 802.11 DCF. In the case of multiple ACs such as in IEEE 802.11e EDCA the simple technique in [56] does not translate well.

The difficulty stems from the fact that in [56], no back-off process is initiated after a PIFS interval. In an EDCA environment with multiple ACs the principle of priority access is achieved through differentiated AIFS intervals and back-off processes. Applying the simple technique from [56], would mean that within the AP, internal collisions will
always occur if there are multiple AC type frames awaiting transmission. This results in the high priority AC frames being transmitted and the lower priority AC frames only being transmitted when no high priority AC frames are present. In a saturated WLAN environment, the end result can mean no transmission opportunities for the lower priority AC frames. Furthermore the key ability to maintain throughput proportion allocations between access categories is effectively negated.

In [57], Transmission Opportunity Limit (TXOP Limit) tuning is used between wireless client stations and the AP. The basic principle here relies on the result that given \( n \) wireless client stations, \( n \) total transmissions take place from all wireless client stations to one transmission from the AP. Hence an independent TXOP Limit is defined for the AP such that when it gains an opportunity to transmit, it is allowed to transmit effectively \( n \) frames each time. This is expected to equalize the throughput sharing between DL and UL traffic. The TXOP Limit can be dynamically varied such that levels of throughput sharing can be achieved between DL and UL traffic.

Such an approach appears quite simple, however the process of determining the optimal TXOP Limit when frame sizes are variable for both the AP and wireless client stations is not addressed in [57]. In addition the scenario of variable physical transmission rates for wireless client stations is also not addressed.

In Chapter 3, we described the Two-level Protection and Guarantee Protocol [33]. This scheme was shown to provide the capability of defining specific AC based throughput proportion allocations. In the description provided in Chapter 3, we mentioned that the scheme utilized a transmission reservation procedure, in that the AP would broadcast in each beacon frame period the available free transmission time for specific ACs. This mechanism therefore can be configured in such a manner that the AP is able to directly reserve its own amount of transmission time such that a specific DL and UL throughput sharing is achieved. However, as we discussed further in Chapter 3, the scheme described in [33] does not conform with the IEEE 802.11e standard framework.

In [63], we show in detail with the aid of an analytical model, and by limiting the configuration space to simple contention window parameters of \( CW_{\text{min}} \), that an optimal set of parameters can be determined for all ACs. This optimal configuration takes into account the number of stations within the network, and additionally defines the ability to state what throughput proportional allocation can be given to each AC, all the while maximizing throughput. We also highlight that through this process the AP (the source of DL traffic) could be assigned independent parameters which would give it the ability to gain its own predefined share of the overall throughput proportion allocation. Subsequent to our work in [63], this procedure was then justified and verified as a feasible course of action by Keceli et al. [62].
The procedure of operation described by Keceli et al. [62] is a centralized approach with the AP being responsible for updating its own set of $CW_{\text{min}}$ parameters and the $CW_{\text{min}}$ parameters for wireless client stations. In order to determine the number of effective wireless stations, the AP simply counts the number of unique observed source and destination MAC addresses of incoming frame transmissions.

An adaptation interval set at 5 beacon frame periods is used as the duration of time whereby the AP determines the number of unique active wireless stations. Furthermore this time frame is also used as the periodic time frame for adjusting the $CW_{\text{min}}$ parameters for the AP alone or additionally the wireless stations.

An initial arbitrary parameter setting is assumed for both the AP and wireless stations. At each adaptation interval, the AP determines via the analytical results what parameters should be set for the AP in order to achieve the required DL to UL share for a given AC. These parameters are then employed by the AP, thus in subsequent adaptation intervals a simple fine tuning process is used whereby the $CW_{\text{min}}$ parameter for each AC within the AP is either increased or decreased to maintain the required proportional DL to UL share for the AC. Note that during each adaptation interval if circumstances change, such as the number of active wireless stations, a new starting analytical parameter set is chosen for the AP.

The proposed approach in [62] is quite simple in nature and as shown through simulation results it achieves its specified goal of eliminating the unfairness that exists between the DL and UL within wireless networks. Under saturation load the scheme is able to achieve the basic fairness goal between the DL and UL for each AC. However the arbitrary nature of parameter adjustments does not identify if these parameters are optimal in allowing the best throughput performance.

In general, the work performed in the two groups address the single criterion of overcoming the throughput sharing anomaly between DL and UL traffic by controlling either the DL or UL traffic alone such that throughput sharing is equalized or set to a desired level. In all the proposed solutions barring [56], this is achieved in a compliant IEEE 802.11e EDCA environment. However, the ability to independently control the throughput proportion allocation of ACs within DL and UL is not a feature present in any of the proposed solutions and the principle of maximizing overall throughput within the WLAN has been completely overlooked.

In the next section we describe in detail our proposed solution which addresses the throughput sharing anomaly between the DL and UL traffic. We include the specific ability to define throughput proportion allocation for ACs independently within the DL and UL, while maximizing the overall throughput performance within the WLAN.
4.5 Extended Throughput Allocation Scheme

In order to achieve the additional goals (5 and 6) mentioned at the end of Section 4.3, we begin by allowing both wireless client stations and the AP to have separate EDCA parameter set configurations for AC1 and AC0. As in Chapter 3, we only need to adjust the $CW_{min}$ and $CW_{max}$ parameters in order to achieve these goals. However in this case, the complexity level is increased as we effectively have four ACs operating within the WLAN, two for the wireless client stations and two solely for the AP. The principles outlined in Chapter 3 therefore need to be generalized and extended to encompass a four AC network configuration.

In Chapter 3, we highlighted a relationship that exists in a two AC network configuration (Eq.(3.4.2)), for simplicity, we reiterate the relationship in Eq.(4.5.1). This relationship formed the basis of determining which $CW_{min}$ parameter set configurations that will achieve a desired throughput proportion allocation between AC1 and AC0.

$$\left[ \frac{n_1}{CW_{min1}} \right] E[P]_1 = \frac{AC1_{txr}E[P]_1}{AC0_{txr}E[P]_0} \approx \frac{Thrpt_1}{Thrpt_0},$$

(4.5.1)

where:

$n_0$ is the total number of stations transmitting AC0 traffic.

$n_1$ is the total number of stations transmitting AC1 traffic.

$CW_{min0}$ is the AC0 access category value for $CW_{min}$.

$CW_{min1}$ is the AC1 access category value for $CW_{min}$.

$AC0_{txr}$ is the slot-based transmission probability for AC0.

$AC1_{txr}$ is the slot-based transmission probability for AC1.

$Thrpt_0$ is the throughput (Bits) for AC0 (across complete simulation period).

$Thrpt_1$ is the throughput (Bits) for AC1 (across complete simulation period).

$E[P]_0$ is the average data frame length for AC0.

$E[P]_1$ is the average data frame length for AC1.

The WLAN now consists of essentially four ACs, but Eq.(4.5.1) can still be used to describe the internal throughput proportion allocation between two ACs within the DL and UL. We now define a second relationship that describes the specific overall throughput proportion allocation that exists between the DL and UL with four ACs, as shown in Eq.(4.5.2).

$$\left[ \frac{n_1}{CW_{min1}} \times E[P]_1 \right] + \left( \frac{n_0}{CW_{min0}} \times E[P]_0 \right) \approx \frac{Thrpt_{DL}}{Thrpt_{UL}},$$

(4.5.2)

where:
$CW_{min0}^*$ is the $AC0^*$ value for $CW_{min}$ (AP).
$CW_{min1}^*$ is the $AC1^*$ value for $CW_{min}$ (AP).
$CW_{min0}$ is the $AC0$ value for $CW_{min}$ (wireless client stations).
$CW_{min1}$ is the $AC1$ value for $CW_{min}$ (wireless client stations).
$E[P]_0^*$ is the average data frame length for $AC0^*$ (AP).
$E[P]_1^*$ is the average data frame length for $AC1^*$ (AP).
$E[P]_0$ is the average data frame length for $AC0$ (wireless client stations).
$E[P]_1$ is the average data frame length for $AC1$ (wireless client stations).
$Thrpt_{DL}$ is the throughput (Bits) for DL (across complete simulation period).
$Thrpt_{UL}$ is the throughput (Bits) for UL (across complete simulation period).

Based on these two relationships Eq.(4.5.1) and Eq.(4.5.2), we can determine the $CW_{min}$ parameter sets ($[CW_{min0}^*, CW_{min1}^*, CW_{min0}, CW_{min1}]$), which achieve a desired DL to UL throughput proportion and internal DL and UL $AC1$ to $AC0$ throughput proportions.

As mentioned in Chapter 3, the use of exponentially encoded contention window values places a limitation on the range of possible throughput proportion allocations that can be achieved. In order to overcome this limitation, we proposed a simple per-beacon-frame parameter switching technique similar to that used in Chapter 3. The parameter switching technique utilized a list of parameter sets, which achieved the limited allowable range of throughput proportion allocations when using exponentially encoded contention window values.

The procedure used in determining the list of parameter sets relied on grouping specific parameter sets together that achieved approximately the same throughput proportion. As described in Chapter 3, when viewed in the $ECW$ domain, these groupings formed a number of parallel straight lines. Each point on a line represents a specific parameter set, and all points on a line achieve approximately the same throughput proportion allocation. Hence, we employed an intuitive search mechanism aided by an analytical model [2] to select a specific point on each parallel line that provided the maximum throughput performance.

With the parameter set list finalized, the switching technique would initiate a move from one set to another at each beacon frame period. The direction of the move is dictated by a single factor, the overall network measured $AC1$ throughput proportion allocation to the desired value, which dictates whether an increase or decrease in $AC1$ throughput proportion allocation is required.

Given that similar restrictions also apply in this case, a new process is required first to determine the list of parameter sets that achieve various overall DL to UL through-
4.5. EXTENDED THROUGHPUT ALLOCATION SCHEME

put proportion allocations and certain internal $AC_1^*$ and $AC_1$ throughput proportion allocations. The important requirement for us again, is that each parameter set provides maximum throughput performance. In addition, there are now three factors dictating the direction that a parameter set change should take. These three factors are, DL to UL throughput proportion allocation, $AC_1^*$ throughput proportion allocation in the DL and $AC_1$ throughput proportion allocation in the UL.

Determining the $CW_{\text{min}}$ parameters:

In Chapter 3 we dealt with a simple two $AC$ network, however in this chapter we now have to consider a four $AC$ network and we show how the solution used to solve the two dimensional problem in Chapter 3 can be generalized to encompass a multidimensional problem.

The principle of grouping different parameter sets into multiple parallel straight lines, with each line representing a different throughput proportion allocation is an intuitive observation in the two dimensional problem described in Chapter 3 and yet this concept is still valid even if the number of dimensions increases.

In our current four dimensional problem, we can once again visualize the existence of multiple parallel lines with each providing different throughput proportion allocations. In this case, as opposed to the single overall $AC_1$ throughput proportion allocation in two dimensions, we now have three throughput proportion allocations in four dimensions, overall DL to UL, $AC_1^*$ proportion within DL and $AC_1$ proportion within UL. The point that provides the maximum throughput performance on each parallel line needs to be determined via some search mechanism.

The search mechanism used in Chapter 3 began by first determining (with the aid of an analytical model [2]) the throughput performance of each individual point on a specific line (note that a point refers to a parameter set). The specific line was chosen based on the criteria that it provided the closest throughput proportion allocation to the desired target value. The point that provided the maximum throughput performance is chosen as the starting reference point, referred to as $x_0$. With $x_0$ evaluated, each additional point on each neighboring parallel line that also provided maximum throughput performance was determined.

The relationships used in the two dimensional problem can easily be extended to the four dimensional problem. Initially we consider the case of just increasing and decreasing the individual throughput proportion allocation for an $AC$ within DL and UL. For simplicity, we present below the relationships used just for DL $AC_1^*$ ($AP$) (note that the same principles can be applied for the remaining case of UL $AC_1$).

Within the DL, we specify changing the proportion allocation for $AC_1^*$ through simple
unit moves within the ECW domain for $AC1^*$ and $AC0^*$. The simple unit moves are presented in the format of where, $[ECW_{min0}^*, ECW_{min1}^*, ECW_{min0}, ECW_{min1}]$ represents the coordinates of a given point in the four dimensional $ECW_{min}$ domain, subsequently a specific vector addition is performed to $[ECW_{min0}^*, ECW_{min1}^*, ECW_{min0}, ECW_{min1}]$. These vector addition operations effectively show the traversing process used to move from one parallel line to another in four dimensions.

In the case of increasing the DL $AC1^*$ throughput proportion allocation, the following two simple unit moves are used, Eq.(4.5.3) and Eq.(4.5.4). Eq.(4.5.3) depicts the case of just reducing the $ECW_{min1}$ value, while Eq.(4.5.4) depicts the case of increasing the $ECW_{min0}$ value. Note that both operations achieve the same objective goal of increasing the DL $AC1^*$ throughput proportion allocation. Likewise, for the case of decreasing $AC1^*$ throughput proportion allocation, there are once again two applicable simple unit moves, namely Eq.(4.5.5) and Eq.(4.5.6).

**Basic Increasing $AC1^*$ Throughput Proportion Allocation:**

\[ W_{i+1}^1 = x_i[ECW_{min0}^*, ECW_{min1}^*, ECW_{min0}, ECW_{min1}] + [0, -1, 0, 0], \]  
(4.5.3)  

\[ X_{i+1}^1 = x_i[ECW_{min0}^*, ECW_{min1}^*, ECW_{min0}, ECW_{min1}] + [1, 0, 0, 0], \]  
(4.5.4)

**Basic Decreasing $AC1^*$ Throughput Proportion Allocation:**

\[ Y_{i-1}^1 = x_i[ECW_{min0}^*, ECW_{min1}^*, ECW_{min0}, ECW_{min1}] + [-1, 0, 0, 0], \]  
(4.5.5)  

\[ Z_{i-1}^1 = x_i[ECW_{min0}^*, ECW_{min1}^*, ECW_{min0}, ECW_{min1}] + [0, 1, 0, 0], \]  
(4.5.6)

In Chapter 3, we discussed and highlighted that there were two possible vector addition moves that could be used to move from one parallel line to another. This resulted in there being two new possible starting points on a new parallel line. Since only a single point location is required, the criteria used for selecting the applicable point was based on which best achieved the goal of maximizing overall network throughput performance. The same principles also applies in this chapter. For each of the basic vector addition operations described earlier on, there exists an equivalent alternative vector addition operation that defines the second point location on each new parallel line.

The additional alternative vector addition operations used for increasing the DL $AC1^*$ throughput proportion allocation are Eq.(4.5.7) and Eq.(4.5.8). Likewise, the additional vector addition operations used for decreasing $AC1^*$ throughput proportion allocation are Eq.(4.5.9) and Eq.(4.5.10). Therefore $W_{i+1}^1$ and $W_{i+1}^2$ are effectively the two points that
lie on a particular parallel line. Similarly, $X^1_{i+1}$ and $X^2_{i+1}$, $Y^1_{i+1}$ and $Y^2_{i+1}$, $Z^1_{i+1}$ and $Z^2_{i+1}$ depict the other two points that lie on different respective parallel lines.

**Alternative Increasing $AC^1*$ Throughput Proportion Allocation:**

\[
W^2_{i+1} = x_i [ECW^*_{min0}, ECW^*_{min1}, ECW^*_{min0}, ECW^*_{min1}] + [1, 0, 1, 1],
\]

\[
X^2_{i+1} = x_i [ECW^*_{min0}, ECW^*_{min1}, ECW^*_{min0}, ECW^*_{min1}] + [0, -1, -1, -1],
\]

**Alternative Decreasing $AC^1*$ Throughput Proportion Allocation:**

\[
Y^2_{i-1} = x_i [ECW^*_{min0}, ECW^*_{min1}, ECW^*_{min0}, ECW^*_{min1}] + [0, 1, 1, 1],
\]

\[
Z^2_{i-1} = x_i [ECW^*_{min0}, ECW^*_{min1}, ECW^*_{min0}, ECW^*_{min1}] + [-1, 0, -1, -1],
\]

An important aspect not addressed thus far is, the additional effects that each specific vector addition operation has on the overall DL to UL proportion allocation. In the case of increasing the DL $AC^1*$ throughput proportion allocations we have a choice of four points labeled $W^1_{i+1}$, $W^2_{i+1}$, $X^1_{i+1}$ and $X^2_{i+1}$. Likewise for decreasing $AC^1*$ throughput proportion allocations we have the four points labeled $Y^1_{i+1}$, $Y^2_{i+1}$, $Z^1_{i+1}$ and $Z^2_{i+1}$. Now the key distinct feature is that $W^1_{i+1}$ or $W^2_{i+1}$, and also $X^1_{i+1}$ or $X^2_{i+1}$ all achieve the required goal of increasing the DL $AC^1*$ proportion allocation, however there is a vital difference present. The points $W^1_{i+1}$ or $W^2_{i+1}$ will increase the DL to UL proportion allocation, while the points $X^1_{i+1}$ or $X^2_{i+1}$ will decrease the DL to UL proportion allocation. The same behaviour is also present for reducing the DL $AC^1*$ proportion allocation. The points $Y^1_{i+1}$ or $Y^2_{i+1}$ will increase the DL to UL proportion allocation, and likewise the points $Z^1_{i+1}$ or $Z^2_{i+1}$ will decrease the DL to UL proportion allocation. Therefore based on these observations and selection criteria for acceptable points, a similar search mechanism to Chapter 3 can now be used to determine all possible parameter sets that achieve specific individual DL $AC^1*$ based throughput proportion allocations.

The initial phase of the search mechanism relies on finding the starting reference point that achieves the closest overall DL to UL throughput proportion allocations and internal DL $AC^1*$ and UL $AC^1$ throughput proportion allocations to the desired target values, while also maximizing throughput performance. In this case, we use Eq.(4.5.1) and Eq.(4.5.2) to first determine which line in four dimensions achieves the closest overall DL to UL throughput proportion allocations and internal DL $AC^1*$ and UL $AC^1$ throughput proportion allocations to the desired target values. Thus with the aid of an analytical
model [2], we define the reference point as being the point on the line that maximizes throughput performance.

With the starting reference point determined, we can then use the relationships depicted in Eq.(4.5.3) to Eq.(4.5.10) to select with the aid of an analytical model [2] the additional points that provide increasing and decreasing $AC_1^*$ throughput proportion allocations. The same process can also be applied from the starting reference point for the remaining case of UL $AC_1$ proportion allocation. Once again as in Chapter 3, we restrict the effective number of lines (combination of throughput proportion allocations between four $AC$s) we need to explore on the basis that for a reference $AC$, a minimum throughput proportion allocation change of at least 0.5% must be achieved.

Now that we have a complete list of parameter sets, we can determine for each parameter set the overall DL to UL, $AC_1^*$ proportion within the DL and $AC_1$ proportion within the UL that they provide by applying Eq.(4.5.1) and Eq.(4.5.2). In Chapter 3, the parameter sets used followed a simple and intuitive ordering scheme. Namely each parameter set progressively changed a single throughput proportion allocation factor, in this case with respect to $AC_1$. In this case, we utilize three different throughput proportion allocations, this introduces an extra level of complexity into the ordering process of each parameter set.

However we can simplify this as we can once again exploit the intuitive relationship between particular parameter sets. To begin with, a parameter set can be thought of as a set of coordinates within a four dimensional space in the $ECW_{min}$ domain. Thus, let us take for example, a reference parameter set $[ECW_{min0}, ECW_{min1}, ECW_{min0}, ECW_{min1}]$.

In Table 4.5.1, we present the vector addition values, which when applied to the reference parameter set results in a different parameter set that can change the overall DL to UL proportion, $AC_1^*$ proportion within the DL or $AC_1$ proportion within the UL in a deterministic manner as shown.

The relationships shown in Table 4.5.1 allow us to form a three dimensional lattice structure consisting of connections between each parameter set evaluated in our search process. Hence, moving from one parameter set to another based on the vector additions depicted in Table 4.5.1, we can determine the exact effect it will have on the overall DL to UL proportion, $AC_1^*$ proportion within the DL and $AC_1$ proportion within the UL. The critical part now is determining which parameter set to use and how any change is instigated in our revised control mechanism, Control Scheme-2.
Table 4.5.1: Effects on Throughput Proportion Allocation.

<table>
<thead>
<tr>
<th>Action</th>
<th>((ECW_{\text{min}}))</th>
<th>DL to UL Proportion</th>
<th>DL (AC1^<em>) to (AC0^</em>) Proportion</th>
<th>UL (AC1) to (AC0) Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>([0,0,1,1]) OR ([-1,-1,0,0])</td>
<td>Increase</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>([1,1,0,0]) OR ([0,0,-1,-1])</td>
<td>Decrease</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>([0,-1,0,0]) OR ([1,0,1,1])</td>
<td>Increase</td>
<td>Increase</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>([1,0,0,0]) OR ([0,-1,-1,-1])</td>
<td>Decrease</td>
<td>Increase</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>([-1,0,0,0]) OR ([0,1,1,1])</td>
<td>Increase</td>
<td>Decrease</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>([0,1,0,0]) OR ([-1,0,-1,-1])</td>
<td>Decrease</td>
<td>Decrease</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>([0,0,1,0]) OR ([-1,-1,0,-1])</td>
<td>Increase</td>
<td>–</td>
<td>Increase</td>
</tr>
<tr>
<td>8</td>
<td>([0,0,0,-1]) OR ([1,1,1,0])</td>
<td>Decrease</td>
<td>–</td>
<td>Increase</td>
</tr>
<tr>
<td>9</td>
<td>([0,0,0,1]) OR ([-1,-1,-1,0])</td>
<td>Increase</td>
<td>–</td>
<td>Decrease</td>
</tr>
<tr>
<td>10</td>
<td>([0,0,-1,0]) OR ([1,1,0,1])</td>
<td>Decrease</td>
<td>–</td>
<td>Decrease</td>
</tr>
</tbody>
</table>

**Control Scheme-2:**

Similar to the previous Control Scheme-1, a parameter set switching technique is also used to achieve the required proportions. In this case there are three throughput proportions of interest, namely the overall DL to UL proportion \((A)\), \(AC1^*\) proportion within the DL \((B)\) and \(AC1\) proportion within the UL \((C)\). We aim to perform parameter set changes that will effectively control all three throughput proportion allocations \([A), (B), (C)]\, at the same time in each beacon frame period.

Within the WLAN, we can easily configure the AP to measure, in each beacon frame period, the value for each of the three throughput proportion allocations \([A), (B), (C)]\, Based on these measurements, the aim is to select a parameter set that adjusts each throughput proportion allocation in the direction towards their respective desired target values.

From Table 4.5.1 we see can that independently adjusting \((B)\) and \((C)\) will also inadvertently affect \((A)\). Due to this, we choose to adjust both \((B)\) and \((C)\) sequentially while at the same time taking into account the necessary direction to adjust \((A)\).

We now consider the basic decisions made by Control Scheme-2 during the parameter set switching technique phase. When the number of stations \(n_0\) and \(n_1\) changes, Control Scheme-2 will revert to the appropriate starting parameter set for that \(n_0\) and \(n_1\). During a period when the number of stations \(n_0\) and \(n_1\) remains constant, the parameter set switching technique is employed as follows.
CHAPTER 4. DIFFERENTIATING UPLINK AND DOWNLINK

• In each completed beacon frame period, the AP has a record of the three measured throughput proportion allocations, overall DL to UL proportion (A), AC1* proportion within the DL (B) and AC1 proportion within the UL (C).

  – Step 1. Adjust DL AC1* to AC0* Proportion (B):
    * Action[3]: Increase (B), while increasing (A).
    * Action[4]: Increase (B), while decreasing (A).
    * Action[5]: Decrease (B), while increasing (A).
    * Action[6]: Decrease (B), while decreasing (A).

  – Step 2. Adjust UL AC1 to AC0 Proportion (C):
    * Action[7]: Increase (C), while increasing (A).
    * Action[8]: Increase (C), while decreasing (A).
    * Action[9]: Decrease (C), while increasing (A).
    * Action[10]: Decrease (C), while decreasing (A).

• Step 1 and 2, are performed sequentially to adjust throughput proportion allocations (B) and (C) while taking into account the necessary direction to adjust (A).

Note that Step 1 and 2 are executed only if the measured throughput proportion allocation for (B) and (C) in a beacon frame period is not at their desired target value. Thus Step 1 or 2 is simply ignored respectively if either the (B) or (C) measured proportion is equal to the desired target value.

We include a final fallback procedure (Step 3) that is used to adjust only (A). This procedure occurs if both the (B) and (C) measured proportions are equal to their desired target value, or both Step 1 and 2 failed to initiate a parameter set change, such as hitting a boundary condition.

  – Step 3. Adjust Overall DL to UL Proportion (A):
    * Action[1]: Increase (A).
    * Action[2]: Decrease (A).

4.6 Performance Evaluations

We evaluate our Control Scheme-2 through simulation using a two stage procedure that shows its main features and critical restriction.
1. In ‘Stage - 1’, we begin by showcasing the basic ability of Control Scheme-2 to overcome the throughput sharing anomaly between overall DL and UL traffic when the number of active wireless client stations is static. As a result we look at both achieving fair sharing and alternative throughput allocations between the DL and UL.

2. In ‘Stage - 2’, we showcase the additional feature of Control Scheme-2 to specify different AC throughput proportion allocations within the DL and UL. In this case, we assume the number of active wireless client stations changes through the simulation period.

### 4.6.1 Performance Metrics

For all simulations, unless otherwise stated, the simulation results are averaged over one second intervals for presentation in graphical form. In each analysis stage we utilize similar performance metrics to evaluate the performance of Control Scheme-2. The following is a description of the specific performance metric used in each stage.

- The average throughput result for DL and UL traffic and the combined total throughput.
- The achieved throughput proportion (percentage) between DL and UL traffic.
- The specific average throughput results for AC1* and AC0* in DL traffic.
- The specific average throughput results for AC1 and AC0 in UL traffic.
- The achieved throughput proportion (percentage) with respect to AC1* in the DL traffic.
- The achieved throughput proportion (percentage) with respect to AC1 in the UL traffic.

As in Chapter 3, in order to gauge the effectiveness of Control Scheme-2, the performance metrics described in Section 3.5.1 are present for each test case scenario. In this case, we determine the excursion performance for each of the three throughput proportion allocations we aim to control, namely overall DL to UL throughput proportion (A), within DL AC1* throughput proportion (B) and within UL AC1 throughput proportion (C), and likewise for the model throughput comparison.
4.6.2 Simulation Setup

As in Chapter 3, we utilize the simulation tool NS-2 (version 2.29) [47] along with an EDCA implementation [48]. In addition, we once again utilize the analytical model given [2] to determine the Parameter List. The general MAC and Physical Layer parameters used are the same as those depicted in Chapter 3, Table 3.5.1. We assume as in Chapter 3, the AP is aware of the number of active wireless client stations in the WLAN.

The basic network topology used in each stage of simulation consists of wireless client stations connected to an AP in an infrastructure mode configuration. We assume that all wireless client stations communicate with the AP at the maximum physical transmission rate of 11Mbps. An external node is connected to the AP through a wired connection with a link speed of 1000Mbps and a one-way propagation delay of 50ms. Note that there is no significance behind the choice of a one-way propagation delay of 50ms. This value was used simply based on the fact that a wired-cum-wireless test configuration in NS-2 (version 2.29) [47] prescribed a generic 50ms one-way propagation delay to individual wired links. The network topology described is represented visually in Figure 4.6.1.

![Basic Network Topology](image)

**Figure 4.6.1: Basic Network Topology.**

We utilize TCP/FTP based traffic for each AC classified for Non Real-time traffic, where the TCP congestion control algorithm used in each simulation is the Reno algorithm with Selective Acknowledgment (SACK) [49, 50]. We also set the packet size for each TCP/FTP flow to 1000 bytes and configure both the AP and wireless client stations to have a buffer size of 50 frames, based on the calculations performed in Chapter 3.

Since TCP traffic is used, we utilize the procedure described in Chapter 3 of configuring
4.6. PERFORMANCE EVALUATIONS

4.6.3 Simulation Results

We present the throughput results and related throughput proportion allocation results on the same output graph. The throughput proportion allocation results will be shown in the upper half of the graph using the right $y-axis$ and common $x-axis$. The set of throughput results are shown in the lower half of the graph, using the left $y-axis$ and common $x-axis$. Legends for the specific throughput proportion and throughput results used in each analysis stage are shown in the upper-right and the center-left position respectively.

Stage - 1: DL and UL Throughput Control

Fair Sharing Throughput Allocation:

We begin by showing Control Scheme-2’s ability to maintain fair throughput sharing between DL and UL traffic, that is, a throughput proportion allocation of 50% between DL and UL traffic. In addition to this, we specify an 80% proportion allocation with respect to $AC1^*$ and $AC1$ within the WLAN. In this test case scenario, we simulate five active wireless client stations each transmitting a single $AC1$ and $AC0$ TCP/FTP flow in the UL direction to the external node (N1). In the case of DL traffic, we configure N1 to transmit a single $AC1^*$ and $AC0^*$ TCP/FTP flow for each of the five active wireless client stations.

In Figure 4.6.2 - (A), we present the overall throughput sharing results between the DL and UL, and the throughput proportion allocation achieved between the DL and UL. Control Scheme-2 is able to maintain a reasonably consistent fair throughput sharing between DL and UL traffic.

In Figure 4.6.2 - (B), the internal DL $AC1^*$ proportion results are also shown to remain consistently close to the required target. There are numerous excursions slightly away from the 80% target value, but these are to be expected due to stochastic fluctuation and the basic nature of our parameter set switching mechanism.

In Figure 4.6.2 - (C), the internal UL $AC1$ proportion results remain more consistently close to the 80% target value as compared to the internal DL $AC1^*$ proportion results. There are fewer large excursions and all are considerably smaller in magnitude than those seen in the internal DL $AC1^*$ proportion results.

The simulation results in Figure 4.6.2 show a slight discrepancy in maintaining the required throughput proportion for internal DL $AC1^*$, which can be attributed primarily
to the effects of multiplexing/smoothing. The combined transmission from five wireless client stations produce a smoother throughput result as opposed to a single AP. Therefore the discrepancies seen in the internal DL AC1* and UL AC1 throughput proportion results in Figure 4.6.2 are not a major concern as overall performance is still very acceptable.

Thus far we have observed Control Scheme-2’s ability to maintain fair throughput sharing between the DL and UL while maintaining an internal 80% proportion allocation for AC1* in the DL and AC1 in the UL. We present two additional simulation results for 67% and 50% proportion allocation with respect to AC1* and AC1, while maintaining fair throughput sharing between the DL and UL.

In Figure 4.6.3 - (A), we can see that Control Scheme-2 is once again able to maintain fair sharing between DL and UL traffic close to the required 50% region. In Figure 4.6.3 - (B), the DL AC1* throughput proportion is predominantly maintained near the required 67% region, but we once again observe a few proportion results that are a fair degree away from the 67% target value. These excursions are again a result of the single AP versus five wireless client stations. Likewise as in the initial test case scenario, the UL proportion results in Figure 4.6.3 - (C) are again consistently maintained near the required 67% target value.

In Figure 4.6.4 - (A), fair throughput sharing between DL and UL traffic is once again maintained. In Figure 4.6.4 - (B) and (C), the vast majority of AC based throughput proportion results remain near the required 50% value.

An important observation in both Figure 4.6.3 and 4.6.4 is the visible excursions above the target in the internal DL AC1* proportion results. This behavior is due to the same observations as made in Chapter 3. We expect that changing from one parameter set to another will inevitably lead to a change in throughput proportion allocation. As noted in Chapter 3 the magnitude of expected proportion change from one parameter set to another varies. In particular as we move towards the boundaries, the magnitude of the change decreases. Thus in the first simulation test case of 80%, the parameter set changes that aim to increase the proportion have smaller impact on the proportion. In the case of the 67% and 50% proportion allocation, parameter set changes that aim to increase the proportion have a substantial impact on the proportion.

In Table 4.6.1, we record the average throughput and proportion results for each simulated throughput proportion allocation combination. The average total throughput achieved is consistent across all test case scenarios. Likewise we can see that the average throughput proportion allocations also lie close to their target value.

As in Chapter 3, we utilize the excursion performance and model throughput comparison to evaluate the effectiveness of maintaining the three throughput proportion allocations and overall throughput performance. In Table 4.6.2, we record the results for these
performance metrics for each simulation test case scenario.

Across all three simulation test case scenarios, the overall DL to UL, internal DL AC1\* to AC0\*, internal UL AC1 to AC0 proportion control all achieve a low percentage excursion value. This performance achieved is more than acceptable as we can see from the visual graphical results. A clear distinct fair sharing between DL and UL is being achieved and likewise the related internal DL and UL AC based proportion allocations are being maintained.

The overall throughput performance across all three test case scenarios for a single simulation run, achieves a comparison percentage of 100% against the analytical model results. As in the previous chapter, the information from a single simulation run is used as a simple way of observing if the control is behaving appropriately.

Therefore, overall we can conclude, regardless of the changes in overall AC based throughput proportion allocation, *Control Scheme-2* is able to achieve fair throughput sharing between DL and UL traffic.
Figure 4.6.2: Stage 1, Throughput Performance - (A): DL = 50%, (B): (AC1*) = 80% and (C): (AC1) = 80%.
Figure 4.6.3: Stage - 1, Throughput Performance - (A): $\text{DL} = 50\%$, (B): $(\text{AC1}^*) = 67\%$ and (C): $(\text{AC1}) = 67\%$. 
Figure 4.6.4: Stage - 1, Throughput Performance - (A): DL = 50%, (B): (AC1*) = 50% and (C): (AC1) = 50%.
Table 4.6.1: Average Results for Different Internal DL AC1* and UL AC1 Throughput Proportion Allocations.

<table>
<thead>
<tr>
<th>Target [DL:AC1*:AC1]</th>
<th>Total (%)</th>
<th>DL (Mbps)</th>
<th>UL (Mbps)</th>
<th>AC1* (Mbps)</th>
<th>AC0* (Mbps)</th>
<th>AC1 (Mbps)</th>
<th>AC0 (Mbps)</th>
<th>DL (%)</th>
<th>AC1* (%)</th>
<th>AC1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[50:80:80]</td>
<td>3.75</td>
<td>1.86</td>
<td>1.89</td>
<td>1.44</td>
<td>0.42</td>
<td>1.49</td>
<td>0.40</td>
<td>49.6</td>
<td>77.4</td>
<td>78.8</td>
</tr>
<tr>
<td>[50:67:67]</td>
<td>3.74</td>
<td>1.82</td>
<td>1.92</td>
<td>1.20</td>
<td>0.62</td>
<td>1.29</td>
<td>0.63</td>
<td>48.7</td>
<td>65.8</td>
<td>67.2</td>
</tr>
<tr>
<td>[50:50:50]</td>
<td>3.75</td>
<td>1.83</td>
<td>1.92</td>
<td>0.94</td>
<td>0.89</td>
<td>0.96</td>
<td>0.96</td>
<td>48.9</td>
<td>51.2</td>
<td>49.9</td>
</tr>
</tbody>
</table>

Table 4.6.2: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target [DL:AC1*:AC1]</th>
<th>Overall DL to UL (%)</th>
<th>DL AC1* to AC0* (%)</th>
<th>UL AC1 to AC0 (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[50:80:80]</td>
<td>3.2</td>
<td>4.0</td>
<td>2.5</td>
<td>100.7</td>
</tr>
<tr>
<td>[50:67:67]</td>
<td>3.3</td>
<td>4.3</td>
<td>3.2</td>
<td>100.9</td>
</tr>
<tr>
<td>[50:50:50]</td>
<td>3.1</td>
<td>3.4</td>
<td>3.5</td>
<td>101.3</td>
</tr>
</tbody>
</table>
Alternative DL to UL Throughput Allocations:

Thus far, in the earlier simulations we depicted Control Scheme-2’s ability to overcome the throughput sharing anomaly that exists between the DL and UL. In the three simulations shown Control Scheme-2 was able to maintain fair throughput sharing between the DL and UL, while also maintaining a required target internal DL and UL AC based throughput proportion allocation. However, Control Scheme-2 as it is designed to, is capable of providing different throughput proportion allocations between the DL and UL other than just fair sharing. As a result, in the next two simulations we depict the DL to UL throughput proportion allocation of 80% and 20% respectively, while maintaining an internal DL and UL AC based proportion allocation of 80%.

In Figure 4.6.5 - (A), we can see that Control Scheme-2 is able to maintain a consistent DL to UL proportion allocation of 80% as required. The black total throughput line is smooth and free of any notable fluctuations, furthermore it remains consistently just below the 4Mbps mark. In the case of the internal DL and UL AC based proportion control, we can see in Figure 4.6.5 - (B) and (C) that both red proportion lines remain close to or on the required 80% target proportion allocation mark. As expected in the case of the UL, there are more fluctuations present simply due to the fact that a low overall throughput is specified for the UL. As a result, any minor change in frame transmissions between UL AC1 and AC0 ends up causing a large impact on the measured proportion allocation. Even so, we can state that the performance obtained is still satisfactory.

In Figure 4.6.6 - (A), we can see the simulated results of the reversed test case scenario where now the UL is effectively allocated 80% of the overall throughput allocation instead of DL. As expected we obtain a slightly lower total throughput, primarily due to the fact that instead of a single AP in the DL obtaining the majority share of throughput, it is now being allocated to the five wireless client stations in the UL which all compete with each other. Even so the black total throughput line remains smooth and free of any notable fluctuations. Aside from the lower throughput performance the major and key difference in this simulation is the more notable minor fluctuations observed in the measured overall DL to UL proportion allocation line.

These additional minor fluctuations are mainly due to the proportion allocation and throughput performance response observed in the DL. Thus looking at the DL graphs depicted in Figure 4.6.6 - (B), we can see numerous pronounced fluctuations in the reported internal DL AC based proportion allocation results. The fluctuations are significantly different to those observed in the previous simulation results where the UL was subjected to the similar constraint of being allocated only 20% of the overall total network throughput. In the case of the UL, minor fluctuations were recorded due to the fact that given a lower target proportion allocation and thus overall throughput any minor change in the
number of frame transmissions inevitably impact on the recorded proportion allocation results. However, in this simulation the DL response is due to this observation and more so by the fact that there is only a single \textit{AP} present that is solely responsible for the transmission of DL traffic (even though both the DL and UL directions involve 5 streams per \textit{AC}).

The multiplexing effect of multiple wireless client stations in the UL allows for a smoother overall throughput response as opposed to a single station. Furthermore, with just a single \textit{AP} and an allocation of 20\%, this means that on average given that 50 frames can be transmitted every beacon frame, the \textit{AP} will transmit roughly 10 frames in total. Subsequently this is split in the manner where on average 8 frames are DL \textit{AC1*} and 2 frames \textit{AC0*}. With just 2 frames trying to be sent every beacon frame by a single \textit{AP}, we observed on numerous occasions that this simply does not occur even though DL \textit{AC0*} frames are present within their respective buffer. In fact there are many beacon frames that elapse with no DL \textit{AC0*} frames being transmitted. This effectively causes the internal DL proportion allocation to report a complete 100\% dominance by DL \textit{AC1*}, and in turn \textit{Control Scheme-2} reacts, as required, to assist DL \textit{AC0*}.

The numerous downward spikes observed are a direct consequence of \textit{Control Scheme-2} taking this action, by effectively allowing DL \textit{AC0*} a higher opportunity to transmit through giving it a more favorable \textit{CW} parameter configuration to that of DL \textit{AC1*}. This behaviour is not as prominent in the UL, simply due to the fact that even when the UL is allocated a 20\% share of the overall throughput and thus UL \textit{AC0} is only sending on average 2 frames a beacon frame, there are in total \textit{five} wireless client stations attempting to transmit. Therefore this observed behaviour is an inherent feature in the DL when given a low overall throughput proportion allocation.

In Figure 4.6.6 - (C) we observe that the UL is able to maintain, as required, an internal \textit{AC} based throughput allocation sharing of 80\%. The measured proportion allocation line (in \textit{red}) is seen to clearly lie on the 80\% mark for the whole simulation duration. With a higher total throughput allocation prescribed to UL, we can see that there are only minimal fluctuations present and thus the achieved performance is highly acceptable.

In Table 4.6.3, we record once again the average throughput and proportion results for the two simulated throughput proportion allocation combination. The average total throughput achieved is slightly higher in the case of the DL being allocated an 80\% proportion allocation, as expected, since the single \textit{AP} accounts for the majority of the \textit{WLAN} traffic.

For the average throughput proportion allocations, the results obtained for when the DL is allocated an 80\% proportion allocation, are such that all the related average proportion allocations lie close to their \textit{target} value. However, in the case when the DL is
allocated a 20% proportion of the total throughput, we observe that the average internal DL $AC$ based proportion allocation is over 5% away from the required target value of 80%, and registers a value of 73.6%. This is a direct result of the numerous downward spikes observed in the simulation plots of 4.6.6 - (B). Apart from this the remaining two proportion allocations are both very close to the required target value of 20% for the overall DL to UL proportion and likewise the 80% mark for the internal UL $AC$ based proportion.

In Table 4.6.4, we record the results for the excursion performance metrics and model throughput comparisons for each of the two simulated test case scenarios. For the first simulation, the overall DL to UL, internal DL $AC_1^*$ to $AC_0^*$, internal UL $AC_1$ to $AC_0$ proportion control all achieve a low percentage excursion value. However as expected in the second simulation, a larger excursion percentage amount is registered for the internal DL $AC_1^*$ to $AC_0^*$ proportion control. Even so, we deem the achieved performance as being acceptable, given the inherent difficulties in managing a low throughput channel a contention-based system and only one station representing that channel the contention.

The overall throughput performance across the two test case scenarios for a single simulation run, achieves a comparison percentage that is acceptable in our view. Once again, the information from a single simulation run is used as a simple way of observing if the control is behaving appropriately and no major loss in throughput performance is occurring.

Therefore, overall we can conclude, that Control Scheme-2 is able to maintain alternative DL to UL proportion allocations aside from the previously tested fair sharing arrangement.
4.6. PERFORMANCE EVALUATIONS

Figure 4.6.5: Stage 1, Throughput Performance - (A): $DL = 80\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$. 

Throughput (Bits/sec)
Figure 4.6.6: Stage - 1, Throughput Performance - (A): DL = 20%, (B): (AC1*) = 80% and (C): (AC1) = 80%.
### Table 4.6.3: Average Results for Different DL to UL Throughput Proportion Allocations.

<table>
<thead>
<tr>
<th>Target [DL;AC1*;AC1]</th>
<th>Total</th>
<th>DL</th>
<th>UL</th>
<th>AC1*</th>
<th>AC0*</th>
<th>AC1</th>
<th>AC0</th>
<th>DL</th>
<th>AC1*</th>
<th>AC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(%)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>[80;80;80]</td>
<td>3.85</td>
<td>2.98</td>
<td>0.87</td>
<td>2.33</td>
<td>0.65</td>
<td>0.68</td>
<td>0.19</td>
<td>77.4</td>
<td>78.2</td>
<td>78.0</td>
</tr>
<tr>
<td>[20;80;80]</td>
<td>3.62</td>
<td>0.74</td>
<td>2.88</td>
<td>0.54</td>
<td>0.20</td>
<td>2.27</td>
<td>0.61</td>
<td>20.4</td>
<td>73.6</td>
<td>78.7</td>
</tr>
</tbody>
</table>

### Table 4.6.4: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target [DL;AC1*;AC1]</th>
<th>Overall DL to UL</th>
<th>DL AC1* to AC0*</th>
<th>UL AC1 to AC0</th>
<th>Model Throughput Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>[80;80;80]</td>
<td>3.6</td>
<td>3.0</td>
<td>5.0</td>
<td>104.0</td>
</tr>
<tr>
<td>[20;80;80]</td>
<td>4.7</td>
<td>11.0</td>
<td>3.0</td>
<td>98.1</td>
</tr>
</tbody>
</table>


**Stage - 2: Performance in Changing WLAN Topology**

In **Stage - 1**, we strictly considered a static WLAN topology of only five active wireless client stations. As a result, we only observed five UL AC1 and AC0 TCP/FTP flows respectively. In the following test case scenarios we simulate a changing WLAN topology with variable number of active wireless client stations contributing to the overall UL traffic. Therefore the number of UL AC1 and AC0 TCP/FTP flows is changed over the course of the simulation duration. Note that as in the previous analysis, the DL traffic is still static in the sense that the external node continues to transmit data over the whole simulation period for both AC1* and AC0*. However, the number of TCP/FTP flows for DL AC1* and AC0* is increased to ten, each, given that the maximum number of wireless client stations is now ten.

In Table 4.6.5, we define the specific number of UL AC1 and AC0 TCP/FTP flows present in the WLAN during the course of simulation. An important thing to notice from Table 4.6.5, is that when the number of flows for AC1 and AC0 are not the same, this simply means that some wireless client stations are exclusively transmitting one AC alone.

<table>
<thead>
<tr>
<th>AC1</th>
<th>AC0</th>
<th>Time Duration (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td>25 - 45</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>45 - 65</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>65 - 85</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>85 - 105</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>105 - 125</td>
</tr>
</tbody>
</table>

We begin with two basic simulation test case scenarios, in the first case we specify a DL to UL proportion of 80% and likewise for both the internal DL and UL a proportion of 80%. In the second test case scenario we change the DL and UL proportion to 50% and introduce different internal DL and UL AC based proportions of 70% and 60%, respectively.

In Figure 4.6.7, The DL to UL proportion in (A) is maintained reasonably well against the required 80% target value. In the case of the DL and UL AC based proportion, in (B) and (C) we can see that overall the proportion results also remain reasonably close to the 80% target value. In the second test case scenario, results shown in Figure 4.6.8, we can see that in (A), the DL and UL proportions are once again maintained near the required 50% target value. While in (B) and (C) the AC based proportion results within
the DL and UL are also maintained reasonably well.

As in the previous Stage - 1 of our analysis, we present a table of average results for all measured quantities. We can see in Table 4.6.6, the average total throughput achieved for the first test case scenario is consistent with previous simulated results. However a slight throughput reduction is achieved in the second test case scenario. The average throughput proportion results for both test case scenarios are close to their respective target values.

In Table 4.6.7 we present the set of performance metrics for both test case scenarios. Both test case scenarios achieve an acceptable excursion percentage value for all three proportion allocations. However in the case of the UL AC1 to AC0 excursion percentage, we observe a slightly higher result for the first test case scenario. This can be attributed to the low throughput allocated to the UL, thus slight changes in frame transmissions for the UL result in a distinctly larger variation in proportion results.

Here AC1 in the UL is allocated 80% of 20% of the total WLAN bandwidth, which is approximately 8 transmissions per beacon frame. Hence, just a single change of 1 transmission reflects a 12% change in throughput. In contrast, with fair sharing approximately 20 transmissions per beacon frame are expected and so any single change represents just a 5% change. This, together with the proven difficulty in maintaining consistent TCP throughput on slow links explains the greater excursions in this case. As for the second test case scenario a lower value is obtained at just over 3%. This time round with fair sharing being prescribed between the DL and UL, no such effect is seen.

As expected the model throughput comparison reflects the results seen in the average throughput results. The first test case scenario reports a clear one-to-one comparison with the analytical model results. However in the second test case scenario we can see that a 98% comparison percentage is reported against the model. This value is still reasonably high, and in conjunction with the effective proportion allocation being achieved, we view this as an acceptable result.

Overall we can conclude that changing the number of AC1 and AC0 flows over the simulation period has not greatly affected the ability of Control Scheme-2 to maintain the required proportion values. This is reflected in the presented graphical simulation results and also in Tables 4.6.6 and 4.6.7.
CHAPTER 4. DIFFERENTIATING UPLINK AND DOWNLINK

Figure 4.6.7: Stage - 2, Throughput Performance - (A): $DL = 80\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$. 
4.6. PERFORMANCE EVALUATIONS

Figure 4.6.8: Stage 2, Throughput Performance - (A): DL = 50%, (B): (AC1∗) = 70% and (C): (AC1) = 60%.
Table 4.6.6: Average Results for Various Throughput Proportion Allocations.

<table>
<thead>
<tr>
<th>Target</th>
<th>Total</th>
<th>DL</th>
<th>UL</th>
<th>AC1*</th>
<th>AC0*</th>
<th>AC1</th>
<th>AC0</th>
<th>DL</th>
<th>AC1*</th>
<th>AC1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(%)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>[80;80;80]</td>
<td>3.75</td>
<td>2.90</td>
<td>0.843</td>
<td>2.27</td>
<td>0.63</td>
<td>0.66</td>
<td>0.19</td>
<td>77.5</td>
<td>78.2</td>
<td>78.0</td>
</tr>
<tr>
<td>[50;70;60]</td>
<td>3.67</td>
<td>1.78</td>
<td>1.89</td>
<td>1.22</td>
<td>0.56</td>
<td>1.13</td>
<td>0.75</td>
<td>48.5</td>
<td>68.7</td>
<td>60.1</td>
</tr>
</tbody>
</table>

Table 4.6.7: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target</th>
<th>Overall DL to UL</th>
<th>DL AC1* to AC0*</th>
<th>UL AC1 to AC0</th>
<th>Model Throughput Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>[80;80;80]</td>
<td>3.5</td>
<td>3.0</td>
<td>5.2</td>
<td>101.2</td>
</tr>
<tr>
<td>[50;70;60]</td>
<td>3.4</td>
<td>3.9</td>
<td>3.4</td>
<td>98.8</td>
</tr>
</tbody>
</table>
Impact of Non-Saturated Downlink Transmissions:

We have only considered Control Scheme-2 in a saturated environment and now show how it performs under non-saturation. In the previous two stages, the network topology used was such that the WLAN was clearly the bottleneck in the network. All TCP/FTP flows use this bottleneck and effectively create saturation load conditions within the WLAN for all wireless client stations and the AP.

In this case, we change our network topology such that in the DL direction the WLAN is no longer the bottleneck link, by changing the external node’s (N1) link speed connection to the AP to 0.5Mbps. In the case of the UL traffic, we include an additional external node (N2) connected to the AP that acts as the destination node for all UL TCP/FTP flows. The additional external node is connected to the AP and retains the previous link speed of 1000Mbps and the generic propagation delay of 50ms. The updated network topology used is shown in Figure 4.6.9.

We simulate the test case scenario of a DL and UL target proportion of 50% and both internal DL and UL AC based proportions set to 50%, likewise we assume the same number of active wireless client stations as in Stage - 2, see Table 4.6.5.

We can see from Figure 4.6.10 - (A), the throughput performance for the DL is consistently at the 0.5Mbps mark, coinciding with the bottleneck link speed of N1 for all incoming DL traffic into the WLAN. Based on the bottleneck link speed of 0.5Mbps, and the approximate WLAN capacity of 3.7Mbps, the DL is only generating approximately 15% of the total throughput capacity. This mark is highlighted by the green dashed line.
in Figure 4.6.10 - (A). Unfortunately, the UL does not use all of the remaining capacity in the WLAN because our control tries to ensure that the overall DL to UL proportion allocation is kept at the required target value of 50%. This effectively throttles the UL traffic to a similar overall rate.

As a result, we can see the total throughput falls well below the values seen in previous results for the entire simulation period. We can see slight evidence of the throughput value increasing at the station count changeover points but these still fall well below the overall mark of 3.7Mbps. The outgoing UL TCP/FTP traffic in this particular test case scenario has not been able to utilize the available throughput that is present in the WLAN. Likewise the overall average throughput results shown in Table 4.6.8 are clearly less than those seen in Stage - 2, Table 4.6.6.

This result exposes a shortcoming of Control Scheme-2, which stems from the fact that Control Scheme-2 is built on an assumption of both the wireless client stations and the AP being continually saturated. In the current simulated network topology, the AP was effectively non-saturated for both AC1\* and AC0\* traffic. The strict observance of maintaining the required target throughput proportion allocations is an effective strategy in WLAN environment when fully saturated conditions are present. However from these results it does not work well when non-saturated traffic is introduced.
Figure 4.6.10: Throughput Performance - (A): 50% - DL and UL Traffic (Highlighted 14%), (B): \((AC1^*) = 50\%\) Throughput Proportion Allocation and (C): \((AC1) = 50\%\) Throughput Proportion Allocation.
Table 4.6.8: Average Result for [50;50;50] Throughput Proportion Allocation.

<table>
<thead>
<tr>
<th>Target [DL;AC1*;AC1]</th>
<th>Total (%)</th>
<th>DL (Mbps)</th>
<th>UL (Mbps)</th>
<th>AC1* (Mbps)</th>
<th>AC0* (Mbps)</th>
<th>AC1 (Mbps)</th>
<th>AC0 (Mbps)</th>
<th>DL (%)</th>
<th>AC1* (%)</th>
<th>AC1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[50;50;50]</td>
<td>1.14</td>
<td>0.50</td>
<td>0.64</td>
<td>0.25</td>
<td>0.25</td>
<td>0.33</td>
<td>0.31</td>
<td>46.2</td>
<td>50.6</td>
<td>50.8</td>
</tr>
</tbody>
</table>

Table 4.6.9: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Target [DL;AC1*;AC1]</th>
<th>Overall DL to UL (%)</th>
<th>DL AC1* to AC0* (%)</th>
<th>UL AC1 to AC0 (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[50;50;50]</td>
<td>5.4</td>
<td>3.4</td>
<td>5.0</td>
<td>30.3</td>
</tr>
</tbody>
</table>
4.7 Summary

In this chapter our main goal was to address the issue of the throughput sharing anomaly that exists between the DL and UL traffic within a WLAN, which we highlighted in the previous chapter. In this chapter we successfully described and defined Control Scheme-2 that overcomes this throughput sharing anomaly in a saturated environment.

Control Scheme-2 is unique in that we can define a specific throughput sharing scheme between the DL and UL, and also at the same time define separate AC based throughput proportions within the DL and UL. We analyzed Control Scheme-2 through different test case scenarios in Stage - 1 and Stage - 2. In each case we are able to see that Control Scheme-2 is able to maintain the required throughput proportion allocation between DL and UL, while also maintaining the required internal DL and UL AC based throughput proportion allocation.

However, an important shortcoming of Control Scheme-2 is that it is built on an unrealistic assumption that all wireless client stations and the AP are always saturated. We configured a network topology where the DL traffic is unable to produce saturation load conditions within the AP, and observed that Control Scheme-2e is able to maintain the required throughput proportion allocations when non-saturated load is present within the WLAN but at the expense of overall throughput. In Chapter 5 and 6, we address this issue in detail and propose additional mechanisms to aid Control Scheme-2 such that adequate performance is maintained in changing load conditions.
Chapter 5

Non-saturated Downlink Traffic

5.1 Introduction

In Chapter 4, Control Scheme-2 was shown to meet the required goals of providing dynamic service differentiation in the form of overall throughput proportion allocation between the DL and UL and the related internal ACs in a WLAN operating under the EDCA mechanism. However, a critical requirement in Control Scheme-2 is that each AC in all wireless client stations and the AP itself must be saturated. An AC within a client wireless station or AP is classified as being saturated when it always has a data frame awaiting transmission at the end of its back-off process.

The requirement for saturated ACs within each client wireless station and the AP itself is clearly unrealistic. As briefly shown and discussed in Chapter 4, with the introduction of non-saturated DL traffic Control Scheme-2 was unable to achieve the goal of maximizing throughput performance within the WLAN. Therefore, in this chapter we assume that the DL traffic may be non-saturated within the WLAN, while UL traffic is deemed to be saturated.

We consider this scenario, and delay the UL scenario until a later chapter, because DL traffic enters the WLAN through the AP. With the control mechanism residing within the AP, the logical approach is therefore to first address control issues regarding non-saturated DL traffic alone, prior to considering the more complex scenario of non-saturated UL traffic.

We begin by first defining the requirements when non-saturated DL traffic is present within the WLAN. This is followed by a description of the critical components that must be included as modifications to Control Scheme-2 in order to allow it to operate effectively in the presence of non-saturated DL traffic. Through the use of simple simulation results, we provide an analysis of the importance of each of the critical components. In particular, we highlight the fact that all of the critical components must be present together in order
to achieve the required performance goals.

## 5.2 Performance Criteria

The following performance criteria are the requirements we have set on the modified Control Scheme-2 in the scenario where, within the WLAN, non-saturated DL traffic is present for either of its ACs, while the UL traffic is fully saturated for both its related ACs. In this scenario, we can immediately see that two network proportions, namely the overall DL to UL allocation and the internal DL, AC1* to AC0* allocation, will need to be specially managed. The internal UL AC1 to AC0 allocation will just be managed to its existing target. We define the following requirements, in priority order, as:

- Maximize overall throughput performance,
- Maintain DL to UL target proportion whenever possible, and
- Maintain the internal target DL AC1* to AC0* proportion allocation whenever possible.

The choice of priority order described above is consistent with previous chapters. Maximizing throughput performance was the key goal in the process used in generating the set of optimal contention window parameters. Furthermore the obvious ingredient to maximize throughput performance is that there must be sufficient traffic generated within the WLAN for the DL and UL. Thus, it is natural that the control of the DL to UL target proportion is of higher priority compared to the internal control of DL AC1* to AC0* proportion.

## 5.3 Components Required in the Control Scheme

In the design and development of Control Scheme-2, we have always specified the control to be an autonomous system that solely operates within the AP. We now present the three necessary components in order to achieve the stated requirements.

1. DL Load Classification Mechanism
2. Dynamic Proportion Control
3. Buffer Occupancy Monitor
5.3.1 DL Load Classification Mechanism

The ability to classify the traffic load characteristics within the WLAN, as saturated or not, is the first necessary component.

In order to determine this information we require a means to observe directly the data frame transmission characteristics of the DL. This is made easier given that the control resides within the AP. As a result, we observe directly the nature of the back-off process initiated within the AP.

For each elapsed beacon frame period, the transmission status for each DL AC can be classified as being Inactive, Non-saturated or Saturated.

- Inactive: during the beacon frame period the back-off process of the AC does not initiate any slot-based countdown operation, therefore no data frame transmissions have occurred.

- Non-saturated: during the beacon frame period, on at least one occasion, the back-off process of the AC countdown operation elapsed without a data frame being transmitted, however some data frame transmissions have taken place.

- Saturated: during the beacon frame period every time the back-off process of the AC countdown operation elapsed, a data frame was transmitted, or an internal collision occurred.

As a result, we can now define the following five overall traffic regimes for DL.

1. Both DL AC1* and AC0* are inactive.
2. Both DL AC1* and AC0* are non-saturated or one DL AC is non-saturated, while the other is inactive.
3. Only DL AC0* is saturated, while AC1* is either inactive or non-saturated.
4. Only DL AC1* is saturated, while AC0* is either inactive or non-saturated.
5. Both DL AC1* and AC0* are saturated.

5.3.2 Dynamic Proportion Control

Control Scheme-2 was shown to have the ability to maintain proportion allocations when given required target values. However, the major shortcoming is that this strict adherence to specified target values would only produce acceptable results when the WLAN is fully saturated, that is the DL must be operating in regime ‘5’.
In the ideal situation where the AP knew all the transmission rates, the AP could determine target values that would ensure saturation and meet the specified performance criteria. Therefore, in the realistic scenario where the AP does not have all this information, if the target values can be manipulated to reflect the current network loading conditions, we should expect that acceptable performance results could be obtained. Therefore, a dynamic target value for both DL to UL and DL $AC_1^*$ to $AC_0^*$ proportion allocations are seen as necessary components of the control in a dynamically changing network environment. We now address the problem of how to set these dynamic targets for the respective proportion allocations.

5.3.3 Buffer Occupancy Monitor

Within the AP, buffering of data frames is a necessary component for allowing TCP based traffic to operate efficiently. Aside from this fact, a buffer which always contains data frames provides the additional feature of guaranteeing saturation load conditions. Since, upon transmission of a data frame there will always be an additional data frame awaiting transmission that is stored in the buffer.

The benefit of maintaining saturation load conditions is obvious, as it guarantees the correct operation of Control Scheme-2, which in turn means achieving the desired performance requirements (see Chapter 4). So, we should try and maintain a consistent level of data frame buffering with the AP.

Now taking a simplistic view, we can assume that $DL_B u f f e r = Max_{Trx} \times dDLULprop$ data frames will be transmitted by the DL per beacon frame, and $DL_B u f f e r \times dDLAC1prop$ data frames will be transmitted for $AC_1^*$ and $DL_B u f f e r \times (100 - dDLAC1prop)$ data frames will be transmitted for $AC_0^*$. Here, $Max_{Trx}$ is the maximum number of data frames that can be transmitted in a beacon frame period and is approximately 50 data frames, see Table 3.5.8. Recall that $dDLULprop$ is the target value proportion allocation for DL, and $dDLAC1prop$ is the target proportion allocation for $AC_1^*$.

We can then use appropriately scaled versions of these numbers to determine if the current target proportion allocations are likely to be sustainable or not. Therefore, at the start of each beacon frame period, the observed buffer occupancy level can be directly used to infer this information. The ability to observe buffer occupancy level is easy for the DL since the AP can directly obtain buffer occupancy level information for the entire DL.
5.4 Importance of the Three Necessary Components

In the previous section, we briefly described three necessary components, and argued why they were necessary. In this section, we discuss in detail, the importance of each of the three components, and why each is vital. In addition to this, we describe the procedure used to implement each of the components within Control Scheme-2. Throughout the discussions, we utilize simulation results of simple network test case scenarios to further illustrate both the importance and implementation considerations of each necessary component within modified Control Scheme-2.

5.4.1 Utilizing DL Load Classification Information

The DL load classification mechanism described in the previous section, identified five specific operating regimes for the WLAN with respect to the DL. We define the following actions that should be performed in each of the five operation regimes. Note that, DL load classification data is obtained for a completed beacon frame period, thus actions are taken based on this classification in the next beacon frame.

1. The DL is observed as being inactive and all WLAN traffic is comprised of UL alone. Therefore, the target DL to UL proportion allocation and the internal DL AC1* to AC0* proportion allocation are irrelevant. Hence, we assist the UL, while respecting its internal target AC1 to AC0 proportion allocation, using direct contention window (CW) changes. However, we restrict acceptable CW changes to those that do not alter the pre-existing CW parameters for the DL. In Table 4.5.1 from Chapter 4, we can determine what are the acceptable CW changes. This restriction is required so that when the DL becomes active, it is not then unfairly disadvantaged by having large initial CW parameters due to previous UL assistance actions.

2. In this operating regime, similar actions are used to those presented in operating regime ‘1’. The only difference is that since the DL is deemed active, the restriction to acceptable CW changes is no longer relevant and is omitted. Therefore, CW parameter changes that assist the UL and also alter the preexisting DL CW parameters are allowed.

3. In this operating regime, only DL AC0* is identified as being saturated within the DL. We assume that DL AC0* itself can provide sufficient load to meet the target DL to UL proportion allocation. Therefore the target internal DL AC1* to AC0* proportion allocation is ignored in order to maximize overall throughput. Thus, CW parameter changes performed for the DL are chosen in such a manner that they always assist DL AC0*, while taking into account the target DL to UL proportion
allocation. In a similar vein to that of regime ‘1’, CW parameter changes that assist DL AC0* are acceptable only if they do not alter the preexisting CW parameters for DL AC1*. In the case of the UL, CW parameter changes occur based upon the target DL to UL proportion allocation and its internal AC1 to AC0 proportion allocation.

4. This regime is symmetric with regime ‘3’, where AC1* is now the only saturated AC within DL. Therefore DL CW parameter changes assist DL AC1* while maintaining the target DL to UL proportion allocation. Once again these actions are only acceptable if they do not alter the preexisting CW parameters for DL AC0*. UL operation is the same as in operating regime ‘3’.

5. Both DL AC1* and AC0* are saturated, thus both the target DL to UL and target DL AC1* to AC0* proportion allocations are respected. In this regime, all DL CW parameter changes are then based on these two target proportion allocations. The UL CW parameters changes are also based upon the target DL to UL and its target internal AC1 to AC0 proportion allocations. Therefore, the actions taken in this regime are exactly the same as those described in Chapter 4.

The implementation of this DL load classification mechanism along with the related actions, is a relatively straightforward procedure. In the following set of simple simulation results, we illustrate the importance of this component, but we also show that DL load classification, alone, is unable to meet the necessary performance requirements.

Scenario 1: DL AC1* and AC0* Saturated Load

The first basic simulation test case, involves the simple scenario of a fully saturated WLAN under the basic network configuration depicted in Figure 5.4.1.

Traffic for the UL is generated from the five active source wireless client station WS1 to WS5 to the destination external wired node N1, where a single TCP/FTP flow is used both for AC1 and AC0 in the UL. Likewise, in the case of the DL, the reverse configuration holds true, with N1 acting as the source and the destination wireless client station is WS-DL. Note that WS-DL only generates return TCP-ACK for the DL traffic, thus it is not part of the UL active wireless client station count. Once again, a single TCP/FTP flow is used each, for DL AC1* and AC0*. The packet size used for all TCP/FTP flows is set to 1000 bytes. As in all previous network configurations, we assign the buffer size for both the AP and wireless client stations to be 50 frames. Finally, we define the three target proportion allocations to be, DL to UL: 80%, DL AC1* to AC0*: 80% and UL AC1 to AC0: 80%.
In the first simulation run, we revert to the Control Scheme-2 used in Chapter 4, which does not use the DL load classification mechanism. We can see in Figure 5.4.2 that under fully saturated load the control is able to maintain the required three proportion allocations, all the while providing maximum throughput performance. Therefore this result serves as a benchmark comparison to the simulation run, where the DL load classification mechanism and related actions are executed.

In Figure 5.4.3, with the introduction of the DL load classification mechanism, the overall performance can be compared with that in Figure 5.4.2. All three proportion allocations are maintained, and the primary goal of maximizing throughput is also achieved. As seen in Figure 5.4.4: (A), the majority of classifications observed across the simulation period are ‘5’s. This is to be expected given the fully saturated test case scenario. However, there are a few observed classifications that are not ‘5’; these occur early in the simulation run specifically 25s - 30s and afterwards the classification of ‘3’ is registered at relatively regular intervals..

Using Figure 5.4.4 (B), we can see that the slow-start phase is taking place during the 25s - 30s time period, thus full saturation transmission load has not been achieved. The next feature we can see is the timeout for AC1* where the slow-start phase has effectively over shot the network capacity, likewise for AC0* we can see a similar result but with two adjacent halvings of the congestion window instead. As a result of these occurrences we observe the varied classification readings.
The near periodic classification readings of ‘3’ are again due to the behavior of the TCP/FTP flows, where the loss of data packets results in the affected flows reducing their transmission rate and thus for a short period of time non-saturation conditions are present. As we can see from Figure 5.4.4 (B), they all roughly occur just after the halving of the AC1* congestion window value.

The introduction of the DL load classification mechanism has therefore not adversely affected the resulting performance when the WLAN is fully saturated.
Figure 5.4.2: Scenario 1, Throughput Performance - (A): DL = 80%, (B): (AC1*) = 80% and (C): (AC1) = 80%
5.4. IMPORTANCE OF THE THREE NECESSARY COMPONENTS

Figure 5.4.3: Scenario 1, Throughput Performance - (A): DL = 80%, (B): (AC1*) = 80% and (C): (AC1) = 80%
Figure 5.4.4: Scenario 1, DL Trace Log - (A): DL Network Load Classification Regime and (B): TCP Congestion Window
5.4. IMPORTANCE OF THE THREE NECESSARY COMPONENTS

Scenario 2: DL AC1* Non-saturated Load and AC0* Saturated Load

In this simulation test case scenario, we configure the network such that AC1* in the DL is unable to maintain saturation load within the WLAN at its target proportion. In order to achieve this test case scenario, we alter the network topology to that shown in Figure 5.4.5.

![Network Configuration - Scenario 2: DL AC1* Non-saturated Load and AC0* Saturated Load](image)

In this network configuration, the TCP/FTP source node for AC1* in the DL is now N2. As we can see, N2 is connected to the AP via a restricted link with bandwidth of only 0.6Mbps. The bandwidth of 0.6Mbps for the restricted link is chosen because, it coincides approximately to the throughput AC1* would receive in the fully saturated WLAN scenario, were the target proportion allocations to be set at DL to UL: 80% and DL AC1* to AC0*: 20%. However, we still impose the usual target of DL AC1* to AC0*: 80%.

Once again, we begin by simulating this test case scenario, first with the unmodified Control Scheme-2. As we can see in Figure 5.4.6, without the DL load classification mechanism, the control maintains all three target proportion allocations irrespective of the actual load conditions of the WLAN. Although the three target proportions allocations are maintained, it is at the expense of disastrously low throughput performance. In Figure 5.4.6 (A), we can see that the total throughput is around 1Mbps, well below that achieved...
under fully saturated conditions, which is approximately 4Mbps.

In the case of the modified Control Scheme-2 operating with the DL load classification mechanism, the simulation results are shown in Figure 5.4.7, where we can see that the overall throughput performance is far better than that seen in Figure 5.4.6 and is consistent with that of the fully saturated WLAN scenario at about 4Mbps. In addition, the DL to UL proportion allocation is still maintained at the desired level of 80%.

The internal DL AC1* to AC0* proportion allocation was specified to be 80%. However, as we mentioned before, the achievable allocation value should actually be around 20%, which is clearly being maintained as seen in Figure 5.4.7 (B). Through the use of the DL load classification mechanism and its related actions, we are able to allow DL AC0* to gain the required 80% proportion allocation internally and thereby allow the DL throughput to also achieve the specified DL to UL 80% proportion allocation all at maximum throughput.

The DL load classification mechanism itself as shown in Figure 5.4.8 (A), is registering the majority of readings as the WLAN operating in regime ‘3’. This is to be expected since only DL AC1* alone is non-saturated. However, there are numerous brief occasions when operating regime ‘5’ is observed. This simply coincides with the times when the CW parameter changes caused by assisting DL AC0* have effectively caused AC1* to experience momentary levels of saturation. On a few occasions we also observe a few regime ‘2’ classifications, such as those seen at the start of the simulation period, briefly after the 40s mark and a group of classifications nearing the end of the simulation, right after the 80s mark. These are due to the behavior of the DL AC0* TCP/FTP flow mechanism.

In Figure 5.4.8 (B), we can see the start-up behavior of the congestion windows for DL AC1* and DL AC0*. As before, the initial small congestion window values lead to the rate of transmission being low and thus saturation level is not achieved. Looking at the congestion window for DL AC0*, the regime ‘2’ classification registered after the 40s occurs right after the congestion window reduces by half. The prolonged regime ‘2’ classifications just after the 80s mark, follows the timeout suffered by DL AC0*.
Figure 5.4.6: Scenario 2, Throughput Performance - (A): $DL = 80\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$
Figure 5.4.7: Scenario 2, Throughput Performance - (A): DL = 80%, (B): (AC1*) = 80% and (C): (AC1) = 80%
Figure 5.4.8: Scenario 2, DL Trace Log - (A): DL Network Load Classification Regime and (B): TCP Congestion Window
Scenario 3: DL AC1* and AC0* Non-saturated Load

In the previous test case, we depicted the scenario where only one AC within the DL was non-saturated. We now introduce, the scenario where both AC1* and AC0* within the DL are non-saturated. This is achieved by once again altering the network topology, this time to that depicted in Figure 5.4.9.

![Network Configuration - Scenario 3: DL AC1* and AC0* Non-saturated Load](image)

In this network configuration the source node of both AC1* and AC0* TCP/FTP flows is N2. N2 is connected to the AP via a restricted link with bandwidth of only 1.8Mbps. This amount falls well below the 3Mbps that the DL is aiming to sustain in a fully saturated network configuration, with a DL to UL proportion allocation of 80%. The bandwidth of 1.8Mbps for the restricted link is chosen simply for the fact that it coincides approximately to a proportion allocation of 45% for the DL to UL proportion. As a result, the achievable DL to UL proportion allocation is around this mark. Furthermore with both DL ACs sharing the bandwidth of 1.8Mbps, there are no further restrictions on DL AC1* or DL AC0*, so we again set the internal DL AC1* to AC0* proportion allocation to 80%.

The simulation results produced using only the DL load classification mechanism and its related actions are depicted in Figure 5.4.10, showing that we have not adequately achieved the goal of maximizing throughput performance. The total throughput in Figure
5.4. IMPORTANCE OF THE THREE NECESSARY COMPONENTS

5.4.10 (A), is consistently less than that achieved in the fully saturated load scenario. Likewise the DL to UL proportion allocation achieved is consistently above the correct achievable proportion of 45%. Finally, and most significantly, the internal DL AC1* to AC0* proportion allocation falls well below the target value of 80%. The cause of this is simply that the prioritization within the restricted link of 1.8Mbps is not being respected. Instead the TCP/FTP flow belonging to DL AC0* has managed to gain equal share with the TCP/FTP flow belonging to DL AC1*.

The DL load classification regime results shown in Figure 5.4.11 (A), are observed to be inconsistent and predominantly register an operating regime of ‘2’. The desired operating regime should predominantly be ‘5’.

This highlights a major shortcoming of utilizing only the DL load classification mechanism. Since the DL to UL proportion allocation is static and remains always at 80%, each time we register a classification of ‘3’, ‘4’ or ‘5’ we aim to assist the DL. This is due to the DL to UL proportion allocation remaining at 80%, which stops us from establishing a fully-saturated regime. Furthermore, the goal of allowing DL AC1* the required 80% share of bandwidth fails as the internal proportion allocation can only be adjusted based on the relative frequency of ‘3’, ‘4’ and ‘5’ classifications. Thus without consistent ‘5’ classifications in this scenario, the required 80% for DL AC1* can not be met. Creating saturation conditions whereby every back-off process elapses with a frame transmission attempt in such a network environment is necessary, but not possible by just using the DL load classification mechanism alone.

Therefore we expect that if the DL to UL proportion allocation could be altered and thus somehow set to the correct achievable proportion of 45%, then saturation conditions would be achieved. As a simple experiment, we change the actual target DL to UL proportion allocation to 45%. We can confidently expect that full saturated conditions will be created for both DL ACs. Therefore, by doing this, the target internal DL AC1* to AC0* proportion allocation of 80% should also then be achieved.

The throughput simulation results for this simple experiment are shown in Figure 5.4.12 and Figure 5.4.13. In Figure 5.4.12 (A), we can see that the overall throughput is similar to that achieved in the fully saturated DL ACs test case scenario. Therefore, we are achieving the goal of maximizing throughput. In the case of the internal DL AC1* to AC0* proportion allocation in Figure 5.4.12 (B), at first glance the initial 25s - 40s results are somewhat unsatisfactory. However, afterwards we can see for the remaining duration of the simulation, the target internal DL AC1* to AC0* begins to stabilize at the desired level of 80%. Upon further investigation, we determined that the initial unsatisfactory behavior is due to the actions of the TCP/FTP flows for AC1* and AC0*. Both ACs suffer a TCP timeout at given points during the initial 25s - 40s simulation period. The
first being $AC0^* \text{ just before the 30s mark, as a result we can see straight after this, its throughput has dropped considerably, and thus } AC1^* \text{ effectively gains a large share of the throughput proportion allocation. Conversely the opposite then occurs just prior to the 40s mark, now } AC1^* \text{ TCP/FTP flow suffers a timeout and thus } AC0^* \text{ now gains a considerable increase in throughput.}

The DL load classifications results are shown in Figure 5.4.13 (B). As expected, the initial 25s - 40s simulation period worth of regime classification are not stable. However, after this initial period of instability, we can see that for the remaining duration of the simulation, the majority of classification regimes recorded are ‘5’. There are a few short periods of regime ‘3’ classifications, these simply correspond to the $TCP$ flow of $AC1^*$ suffering a packet loss and thus reducing its transmission rate.

Overall, we can now grasp the importance of being able to set the DL to UL proportion allocation to an adequate value according to the traffic load present in the $WLAN$. This simple experiment of artificially applying a static fixed value for the DL to UL proportion allocation, highlighted this directly. As a result, we now consider the means of introducing a dynamic DL to UL proportion allocation control.
Figure 5.4.10: Scenario 3, Throughput Performance - (A): $DL = 80\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$
CHAPTER 5. NON-SATURATED DOWNLINK TRAFFIC

Figure 5.4.11: Scenario 3, DL Trace Log - (A): DL Network Load Classification Regime and (B): TCP Congestion Window
Figure 5.4.12: Scenario 3, Throughput Performance - (A): $DL = 45\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$
Figure 5.4.13: Scenario 3, DL Trace Log - (A): DL Network Load Classification Regime and (B): TCP Congestion Window
5.4.2 Introducing Dynamic Proportion Control

As we observed, when both DL $AC_1^*$ and $AC_0^*$ were made to be non-saturated through sharing a small bandwidth capacity link, the DL load classification mechanism alone was unable to produce acceptable results. The major issue lies in the fact that we are restricted to controlling to a static target DL to UL proportion allocation, which may be unreasonable. As a result, we aim to introduce the ability to adjust the relevant proportion allocation values such that they reflect the given traffic load within the network.

The two proportion allocation values that can be changed are DL to UL, and DL $AC_1^*$ to $AC_0^*$. It is clear from the results of our last experiment that we need to adjust the DL to UL proportion allocation.

We argue that it is not necessary to adjust the target internal DL $AC_1^*$ to $AC_0^*$. In the test case scenario of a single non-saturated $AC$, we observed a clear difference between the target and achieved DL $AC_1^*$ to $AC_0^*$ proportion allocation. In fact, the achieved proportion allocation was in fact at the desired level. Therefore the actions instigated by the DL load classification mechanism alone was sufficient in achieving the desired DL $AC_1^*$ to $AC_0^*$ proportion allocation. As a result, we retain a static target internal DL $AC_1^*$ to $AC_0^*$ proportion allocation.

The process of adjusting the DL to UL proportion allocation relies on the DL load classification mechanism. We define simple rules to alter the DL to UL proportion allocation based on the operating regime reported by the DL load classification mechanism. We define a fixed target DL to UL proportion allocation value that serves as the ultimate goal when the DL is fully saturated. A secondary dynamic target DL to UL proportion will be used to control the network, and it is this value that will be adjusted accordingly.

The following rules are used in each beacon frame period for each classification regime.

1. When the DL is classified as inactive, no adjustment of the dynamic DL to UL proportion allocation is instigated.

2. When the DL is deemed as being active, but both $ACs$ are non-saturated, we assume that the current dynamic target DL to UL proportion allocation does not allow saturation to be achieved. Therefore we reduce the value by one percentage point.

3, 4 and 5. Since at least one access category is saturated, we assume that the dynamic target DL to UL proportion allocation allows saturation conditions to be achieved. If the dynamic target DL to UL proportion allocation is below the fixed target DL to UL proportion allocation in a fully saturated network, we simply increase the value by one percentage point. If not, we do not adjust the dynamic target DL to UL proportion allocation.
It is clear that with the inclusion of dynamic proportion control, we can still expect acceptable results for the Scenarios 1 and 2. In Scenario 1, where the DL $AC1^*$ and $AC0^*$ are both saturated, we expect consistent regime ‘5’ classifications, thus the dynamic target DL to UL proportion allocation will consistently remain at 80%. Likewise in Scenario 2, where the DL $AC1^*$ is non-saturated and DL $AC0^*$ is saturated, we expect the majority of regime classifications to be ‘3’ with the occasional ‘5’ classification, these all result in the dynamic target DL to UL proportion allocation consistently remaining at 80%. Therefore we only need to consider Scenario 3, where both the DL $AC1^*$ and $AC0^*$ are non-saturated.

**Scenario 3: DL $AC1^*$ and $AC0^*$ Non-saturated Load, revisited.**

Thus far, we have seen that whenever both or at least one of the DL ACs is saturated, the Control Scheme-2 along with the DL load classification mechanism alone, is sufficient to achieve acceptable performance results. Further, the introduction of a dynamic DL to UL proportion allocation, will not affect the observed performance.

However, based on the configuration depicted in Figure 5.4.9, and with a target DL to UL proportion allocation of 80%, both DL ACs are non-saturated since they share the 1.8Mbps wired link connection to the AP. As we observed from the simulation results for the DL load classification mechanism, we were able to achieve acceptable overall throughput performance, at the expense of failing to maintain the target internal DL $AC1^*$ to $AC0^*$ proportion allocation. We highlighted the fact that the use of the fixed DL to UL proportion allocation causes this behavior. As a simple experiment, we simulated the case where a fixed target DL to UL proportion allocation of 45% was used instead of the usual 80%. From these new results we observed that with the DL to UL proportion allocation set at an appropriate level the desired performance level could be achieved. Therefore the use of a dynamic proportion allocation is particularly important in this network scenario.

Accordingly, we introduced the dynamic DL to UL proportion allocation mechanism, and the related results are shown in Figures 5.4.14 and 5.4.15. Unfortunately we observe that the internal DL $AC1^*$ to $AC0^*$ proportion allocation fails to achieve the target 80% mark. The dynamic DL to UL proportion allocation values used during the simulation are shown as red crosses in Figure 5.4.15: (A), are for the most part situated above the desired level of 45%. Further the one second average DL proportion remains below the dynamic target and hovers around the 45% mark. As a result, we simply obtain a performance similar to that of the Control Scheme-2 utilizing the DL load classification mechanism alone, where the DL to UL proportion allocation was fixed at 80%.

Therefore, only relying on the DL load classification mechanism as a means to adjust...
the DL to UL proportion allocation is not sufficient. The primary issue is that the DL load classification mechanism is unable to provide information regarding the sustainability of a given DL to UL proportion allocation. The instant we obtain a classification of ‘3’, ‘4’ or ‘5’, we immediately increase the DL to UL proportion allocation, without knowing if the current proportion allocation is likely to be sustainable or not.

Thus, we have seen, from Figures 5.4.12 and 5.4.13, that being able to adjust the DL to UL proportion allocation is vital. If the right target value is set then the required performance is achieved. However in the case of Figures 5.4.14 and 5.4.15, we observed that adjusting the DL to UL proportion allocation solely based on the DL load classification mechanism is unable to reach the desired level of control. The dynamic target is below 80%, but still remains well above the known 45% level.
Figure 5.4.14: *Scenario 3*, Throughput Performance - (A): $DL = 80\%$, (B): $(AC^1) = 80\%$ and (C): $(AC1) = 80\%$
Figure 5.4.15: Scenario 3, DL Trace Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime and (C): TCP Congestion Window
5.4.3 Utilizing Buffer Occupancy Level Information

So far we have observed that the modified Control Scheme-2 solely based on the DL load classification mechanism is able to achieve the required performance goals only in the scenarios where at least one DL AC is saturated. In the case of both ACs being non-saturated we observed that such a scheme is able to achieve the goal of maximizing throughput, however it fails to maintain the desired target internal DL AC1* to AC0* proportion allocation.

We identified the cause of this failure to be the fixed DL to UL proportion allocation used in the control. Initially, we verified through a simple experiment that by setting the DL to UL proportion allocation to an appropriate value, we can achieve both the goals of maximizing throughput and maintaining the required target internal DL AC1* to AC0* proportion allocation. Therefore, we introduced the ability to dynamically adjust the DL to UL proportion allocation based on the recorded DL load classification regimes. However, upon using the actual DL load classification mechanism to directly adjust the DL to UL proportion allocation, we observed its failure to maintain the required target internal DL AC1* to AC0* proportion allocation. We determined that purely relying on the DL load classification mechanism is insufficient, due to its inability to provide information regarding the saturation ‘sustainability’ of the DL.

Since the DL load classification mechanism is unable to provide information regarding the saturation ‘sustainability’ of the DL, we chose to observe the buffer occupancy level within the AP. Using the relationships discussed earlier in Section 5.3.3, we can calculate the required buffer occupancy level needed to sustain a given DL to UL proportion allocation for a beacon frame period. Therefore, by simply comparing the required buffer occupancy level to that observed, we can determine if the DL is able to sustain the specified DL to UL and DL AC1* to AC0* proportion allocations.

As a result, we define the following additional controls to adjust the DL to UL proportion allocation in each of the DL classification regimes. We denote the required and observed total buffer occupancy level to be $DL_{Buffer}$ and $Ob_{DLBuffer}$.

1. DL is inactive, expected zero buffer occupancy level. Thus as before, we do not alter the dynamic DL to UL proportion allocation.

2. We retain the clause used in the DL load classification based scheme, thus the dynamic target DL to UL proportion allocation is reduced by one percentage point.

3, 4 and 5.  
- If $Ob_{DLBuffer} > DL_{Buffer}$: we increase the dynamic target DL to UL proportion allocation by one percentage point, only if the current dynamic target is less than the fixed target DL to UL proportion allocation (note that the
dynamic target proportion allocation can not exceed the fixed target proportion allocation).

- If ‘Ob\_DL\_Buffer < DL\_Buffer’: we reduce the dynamic target DL to UL proportion allocation by one percentage point.
- If ‘Ob\_DL\_Buffer = DL\_Buffer’: we do not alter the dynamic target DL to UL proportion allocation.

With the inclusion of a buffer occupancy monitor into the control, we expect there to be more cases of the dynamic proportion allocation being adjusted even though the network is saturated from a regime classification point of view. Therefore, we now revert to showcasing all three test case scenarios.

**Scenario 1: DL AC1* and AC0* Saturated Load, revisited.**

With the introduction of the buffer occupancy monitor information also dictating the adjustment process of the DL to UL proportion allocation, we can see in Figure 5.4.16 that the throughput performance under full saturation load is still at an acceptable level. All three proportion allocations are being controlled to their respective target values.

In the case of the dynamic DL to UL proportion allocation adjustment, we can see in Figure 5.4.17: (A) that the majority of the time it resides on the desired value of 80%. However, there are a few occasions where it drops for a short period of time. These correspond to the periods where the observed total buffer occupancy falls below the required value, as seen in Figure 5.4.17: (D). The drops in buffer occupancy correspond to the TCP flow transmission rates reducing on the occasion of dropped packets. With a reduction in transmission rate, it is to be expected that the buffer occupancy will reduce during the period of time that the TCP flow recovers. This is due to the fact that the rate of incoming frames is expected to be less than that of the outgoing frames during this period.

Therefore, we can see that with the introduction of buffer occupancy information into the control, we still obtain very satisfactory results in a saturated environment.
Figure 5.4.16: Scenario 1, Throughput Performance - (A): $DL = 80\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$
Figure 5.4.17: Scenario 1, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information
Scenario 2: DL AC1* Non-saturated Load and AC0* Saturated Load, revisited.

As we saw earlier, the DL load classification mechanism alone, is able to achieve very satisfactory performance in the network environment where only one DL AC is saturated.

With the introduction of the buffer occupancy information, this does not quite hold true, as we can see in Figure 5.4.18 and Figure 5.4.19. The overall throughput performance is still comparable to the previous simulation results in the same network environment. However, the required target DL to UL proportion allocation is not being consistently maintained.

We observe in Figure 5.4.19: (A) a near periodic nature of increase and decrease in the DL to UL proportion allocation. This behavior can be attributed to the buffer occupancy level requirement being set at $DL_{Buffer} = Max_{Trx} \times dDLULprop$. With only a single DL AC TCP flow this requirement is too strict. As a result, we simply employ a scaling factor of ‘0.5’ on the required $DL_{Buffer}$ amount, effectively reducing the requirement in the buffer occupancy level to $DL_{Buffer} = 0.5 \times Max_{Trx} \times dDLULprop$. The choice of 0.5 simply corresponds to allowing more opportunity for frames to be emptied from the buffer prior to initiating a decrease in proportion allocation. This allows us to overcome the saw tooth nature seen in Figure 5.4.19: (A), due to it being a conservative approach to reacting to buffer depletion.

Therefore, with these changes included we now obtain the following results in Figure 5.4.20 and Figure 5.4.21. As we can see from Figure 5.4.21: (A), the DL to UL proportion allocation is essentially steady apart from short minor reductions on the occasion of TCP packet losses. The results presented in Figure 5.4.20: (A) are now in line with those seen previously, and are easily acceptable.
Figure 5.4.18: Scenario 2, Throughput Performance - (A): DL = 80%, (B): (AC1*) = 80% and (C): (AC1) = 80%
Figure 5.4.19: Scenario 2, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information
Figure 5.4.20: *Scenario 2, Throughput Performance* - *(A): \( DL = 80\% \), *(B): \( (AC1^*) = 80\% \) and *(C): \( (AC1) = 80\% \)
Figure 5.4.21: Scenario 2, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information
5.4. IMPORTANCE OF THE THREE NECESSARY COMPONENTS

Scenario 3: DL AC1* and AC0* Non-saturated Load, revisited.

Thus far, with both DL ACs effectively non-saturated through sharing a restricted bandwidth link it has proven to be highly difficult to achieve the desired control over the internal DL AC1* to AC0* proportion allocation. Both of the previous configurations, namely the DL load classification mechanism alone and related dynamic DL to UL proportion allocation, failed to maintain the desired internal DL AC1* to AC0* proportion allocation. However, as shown in the simple test experiment in Figures 5.4.12 and 5.4.13, with the DL to UL proportion allocation artificially fixed to a value of 45%, we were able to achieve the desired internal DL AC1* to AC0* proportion allocation of 80%.

In both previous configurations, we observed that the internal DL AC1* to AC0* proportion allocation achieved was effectively 50%. Therefore, no bias towards AC1* was registered and both DL ACs achieved similar performance. This is not acceptable according to our requirement as clearly, from the simple test experiment results seen in Figure 5.4.12, it is feasible to ensure that AC1* received greater throughput than AC0*.

The important observations made were that the difficulty lies in the DL to UL proportion allocation value being too large in this network environment. In the case of the DL load classification mechanism alone, it was continually at the target value of 80%. Then with the introduced dynamic DL to UL proportion allocation, the values used during the simulation were less than 80%, but still well above the desired level of roughly 45%.

The simulation results produced after introducing the additional buffer occupancy monitor information (with a scaling factor of ‘0.5’ as in the last set of data), are shown in Figures 5.4.22 and 5.4.23. The overall throughput performance in Figure 5.4.22: (A) is still at an acceptable level, and the DL to UL proportion allocation also lies around the correct achievable level. The internal DL AC1* to AC0* proportion achieved, as shown in Figure 5.4.22: (B), this time is markedly different from the previous attempts. We can see that it is well above the 50% mark, and thus AC1* is able to receive a greater throughput performance than AC0*. However, the proportion allocation is still well below the target level of 80%.

In Figure 5.4.23: (A), we can see that the dynamic DL to UL proportion allocation is in general much lower, now, with the introduced buffer occupancy level information. However, it is highly inconsistent, and it is due to this inconsistency that we fail to maintain the internal DL AC1* to AC0* at quite the desired level. The problem lies in the simplicity of the control, where we simply increase the dynamic target DL to UL proportion allocation value always upon registering a buffer occupancy level above the required value. As a result of this, we observe long consecutive increases in proportion allocation, followed by long decreases in the proportion allocation. This creates a sawtooth behaviour in the dynamic target DL to UL proportion allocation. However it should be noted, that
the sawtooth lies roughly at the right desired level. It is clear that if the severity of the sawtooth behavior were dampened, we could expect acceptable performance.

Therefore, introducing buffer level occupancy information has improved the ability to separate the throughput performance achieved between DL $AC1^*$ and $AC0^*$ under the given network environment. However, it is still insufficient in terms of being able to allow the degree of separation to the required target level.

An important thing to note from the results is that, during the short periods of time where consistent DL buffering of frames were evident (for example 50s - 55s, 65s - 70s and 95s - 100s), the respective proportion allocations were controlled to the target level. Therefore, it appears that a crucial factor in being able to maintain proper control is the consistent buffering of data frames within the DL.
Figure 5.4.22: 

Scenario 3, Throughput Performance - 

(A): $DL = 80\%$, 
(B): $(AC1^*) = 80\%$ and 
(C): $(AC1) = 80\%$
Figure 5.4.23: *Scenario 3*, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information.
5.5 Summary

In this chapter we have described the performance requirements that we set under the assumption that non-saturated DL traffic load is present within the WLAN. We then introduced and described three necessary components included into Control Scheme-2 that is able to meet the performance requirements.

With the aid of simple simulation results, we highlighted the importance of each of the three components. In particular the fact that without all three being present together, an unmodified Control Scheme-2 is unable to function adequately and achieve the desired performance targets in a range of non-saturated DL traffic load test case scenarios.

Even so, as we observed in the last set of simulation results, the simplest configuration of all three components is still unable to achieve our performance requirements. This highlights the distinct and surprising difficulty in being able to achieve the performance requirements across the full set of test case scenarios.

In all simulation configurations used in this chapter, we used only a single TCP/FTP flow for each AC. This appears to be a very simple configuration, but in fact, in terms of the behavior of TCP we expect better performance in a multiplexed environment with numerous TCP flows interacting with one another. Furthermore, the presence of multiple interacting TCP flows is a more realistic scenario.

As described briefly in the last set of simulation results where both ACs within the DL are non-saturated, during periods of consistent DL buffering of data frames, it was possible to achieve acceptable performance levels.

In Chapter 6, we describe in full detail the operational structure of the completed DL related modifications to Control Scheme-2, which is able to exploit this crucial observation. Furthermore, we show that the completed DL modifications to Control Scheme-2 is able achieve all the performance requirements across the full range of test case scenarios presented in this chapter.
Chapter 6

Control of the Downlink

6.1 Introduction

In Chapter 5, we described a set of performance requirement criteria that should be met under the conditions where DL traffic flowing into the WLAN is variable in nature and not continuously creating saturation load conditions. In the presence of non-saturated DL traffic flowing into the WLAN, we highlighted three necessary components that are required in the construction of a control scheme that will achieve the set of performance requirements. These three components are once again mentioned below.

- DL Load Classification Mechanism
- Dynamic Proportion Control
- Buffer Occupancy Level

We systemically described step-by-step how each component played a vital role within the control scheme. The simple solution is however, still far from adequate in achieving the performance requirements.

A useful piece of information that enables us to make significant improvement can be gathered from the last set of simulation results shown in Chapter 5. During periods of consistent DL frame buffering within the AP, it was possible for the simple control scheme described in Chapter 5 to achieve acceptable performance levels. Therefore in this chapter, we describe some important changes that will produce a more effective DL control by exploiting this crucial observation.

We begin by defining and describing the modifications and enhancements which we found necessary to provide the complete DL control features to Control Scheme-2. Using this new DL control, we present and analyze a suite of simulation results based on the same simple network configurations of Chapter 5. In particular, we highlight the fact that the complete DL control is now capable of achieving the performance requirements.
6.2 Important Modifications and Enhancements

The following modifications and enhancements are necessary additions to the existing three components in order to introduce complete DL control capabilities to Control Scheme-2 such that it is able to achieve the prescribed performance requirements.

6.2.1 Modifications

From the discussions presented in Chapter 5, we highlighted three components that are necessary in a control scheme capable of meeting the performance requirements. However, in order to achieve acceptable levels of performance, we introduce the following two modifications.

Multiple Buffer Occupancy Level Crossing

In Chapter 5 we described the simple process of adjusting the DL proportion allocation based on the buffer occupancy level within the AP. We used the relationship given in Eq.(6.2.1) to define a single buffer occupancy level threshold for the current DL proportion allocation in operation. We increased or decreased the amount of DL proportion allocation in each beacon frame period, purely based on whether the observed buffer occupancy level was above or below the single defined threshold level:

\[ DL\_Buffer = 0.5 \times Max\_Trx \times dDLULprop, \]  

(6.2.1)

where:

- \( Max\_Trx \), is the maximum number of data frames that can be transmitted in a beacon frame period. With TCP data packet size of 1000 bytes, \( Max\_Trx \) is equal to 50 data frames.

- \( dDLULprop \), is the given target value for the DL to UL proportion allocation, with respect to DL.

- \( DL\_Buffer \), is the minimum amount of data frames that needs to be buffered in the AP in order to sustain the given \( dDLULprop \) proportion allocation.

As shown through the simulation results in Chapter 5, the use of a single buffer occupancy level threshold allowed a clear proportion allocation difference to form between the DL AC1* and AC0*. However, it was still far from being controlled at the desired proportion.
A vital piece of information is that during periods of consistent buffering of data frames, we were able to obtain acceptable levels of performance. These periods of consistent buffering are in general short lived as in the original simple DL control we increased the DL proportion allocation each time we observed the buffer occupancy level to be above the single defined threshold level. Thus an over allocation of the proportion amount to DL was made and which in turn caused the rapid depletion of the buffers within the AP. In order to overcome this behaviour we propose the following modifications to the existing simple DL control:

- The introduction of multiple buffer occupancy level thresholds.
- A modified approach to adjusting the DL proportion allocation.

The first aspect of introducing multiple buffer occupancy level thresholds is to expand upon the concept of level thresholds determined by Eq.(6.2.1). The approach taken is to specify additional threshold levels as simple linear multiples of Eq.(6.2.1) without the scaling factor of 0.5, thus the following buffer occupancy level thresholds are specified for the given current DL proportion allocation by Eq.(6.2.2).

\[ DL_{Buffer_T}[i] = Max_{Trx} \times dDLU_{prop} \times i; \quad "i = 1, 2, 3, \ldots N", \quad (6.2.2) \]

where:

- \(Max_{Trx}\), is the maximum number of data frames that can be transmission in a beacon frame period. With TCP data packet size of 1000 bytes, \(Max_{Trx}\) is equal to 50 data frames.

- \(dDLU_{prop}\), is the given target value for the DL to UL proportion allocation, with respect to DL.

- \(DL_{Buffer_T}[i]\), is the buffer occupancy level threshold for a given value of ‘\(i\)’.

- ‘\(i\)’, is the simple scalar integer multiple factor used to calculate each threshold band, and also serves as an index value for a particular threshold band. Note that the initial primary threshold band dictated by Eq.(6.2.1) has ‘\(index = 0\)’.

- ‘\(N\)’ represents the maximum value for ‘\(i\)’ and thus the largest threshold band. The value of ‘\(N\)’ is governed by \(dDLU_{prop}\) and the combined total AP buffer size across \(AC1^*\) and \(AC0^*\).
With the introduction of multiple level threshold, we propose level-threshold crossing as the method of adjusting the DL proportion allocation. This is in contrast to the simplistic approach used in Chapter 5, where being above or below the single level threshold justified adjusting the DL proportion allocation in each beacon frame period. In order to track the occurrence of crossing a level threshold, we record two readings of the buffer occupancy amount within the AP and classify them in terms of the highest buffer threshold band index that it lies above.

- $Ob_{DL\_Buffer\_Prev}$, the second to last completed beacon frame period record of the buffer threshold band index.
- $Ob_{DL\_Buffer\_Curr}$, the last just completed beacon frame period record of the buffer threshold band index.

In each beacon frame period, we specify a level crossing whenever there is a change in the index value reported by $Ob_{DL\_Buffer\_Prev}$ and $Ob_{DL\_Buffer\_Curr}$. Therefore the following relationship is used, $\text{Change} = Ob_{DL\_Buffer\_Curr} - Ob_{DL\_Buffer\_Prev}$. Based on a positive or negative result reported by $\text{Change}$, we specify an increase or decrease in DL proportion allocation respectively. Additionally the value of $\text{Change}$ dictates the number of scalar multiple of one percentage point changes to the current DL proportion allocation. An example of this procedure is depicted in Figure 6.2.1.

![Buffer Occupancy Level Thresholds](image)

Figure 6.2.1: Transitioning between Positive (A) and Negative Level (B) Crossings

Figure 6.2.1: (A), depicts a positive crossing occurring in the region defined by the two recorded buffer occupancy values ‘B1’ ($Ob_{DL\_Buffer\_Prev} = 3$) and ‘B2’ ($Ob_{DL\_Buffer\_Curr} = 4$). Furthermore, we can see that $\text{Change} = 1$, so we increase the DL proportion allocation from ‘$x\%$’ to ‘$(x + 1)\%$’. In Figure 6.2.1: (B),
we depict the case of a negative crossing being initiated as a result of the values ‘B1’ \((Ob\_DL\_Buffer\_Prev = 4)\) and ‘B2’ \((Ob\_DL\_Buffer\_Curr = 3)\). This produces the value \(Change = -1\), and therefore a decrease in the DL proportion allocation from ‘\(x\)%’ to ‘\((x - 1)\)%’.

In addition during a positive level threshold crossing of only the initial buffer occupancy level, \(Ob\_DL\_Buffer\_Prev = N/A\) and \(Ob\_DL\_Buffer\_Curr = 0\) where \(Change = N/A\), we do not initiate an increase in proportion due to the fact that crossing this threshold alone is not necessarily sufficient for the DL to sustain a higher proportion allocation.

We have described the means of adjusting the DL proportion allocation purely based on level-threshold crossings. However we need to specify two other cases where a crossing is not required in order to force a DL proportion allocation change.

- A one percentage point reduction of the DL proportion allocation is applied every beacon frame period whenever we observe the that the recorded buffer occupancy value is non-zero and below the initial buffer occupancy level threshold, which means \(Ob\_DL\_Buffer\_Curr = N/A\). This is the same procedure described in Chapter 5.

- A one percentage point increase of the DL proportion allocation is applied every beacon frame period whenever \(Ob\_DL\_Buffer\_Curr\) is equal to the index of the largest buffer occupancy level threshold possible for the current \(dDLULprop\). This feature is an intuitive one, given that we have exceeded the last possible buffer occupancy level threshold and no further positive crossings can occur. Note that when both DL ACs are saturated the combined total buffer occupancy is used to determine the largest valid threshold band. Conversely when a single DL AC is saturated, its individual total buffer occupancy is only used to determine the largest valid threshold band.

**Executing Correct Downward Crossing Behaviour**

We have described the inclusion of multiple buffer occupancy level thresholds, and the principle behind each positive and negative level-threshold crossing initiating a change in the DL proportion allocation. However, such an approach of allowing each positive and negative level threshold crossings to initiate a change in the DL proportion allocation is in itself inadequate. The inadequacy can be shown by way of example in Figure 6.2.2.

Figure 6.2.2: (A) depicts the current percentage of ‘\(x\)%’ from which the level threshold bands with index ‘0, 1, 2, 3, 4, 5’ are calculated. The buffer occupancy level has increased from ‘B1’ to ‘B2’ in the most recent completed beacon frame period, resulting in a positive crossing across level threshold band into ‘\(index = 4\)’. This initiates a one percentage point increase in DL proportion allocation from ‘\(x\)%’ to ‘\(y = x + 1\)%’ for the subsequent
6.2. IMPORTANT MODIFICATIONS AND ENHANCEMENTS

beacon frame period. Based on ‘$y\%$’, a new set of level threshold bands with index ‘$0^*, 1^*, 2^*, 3^*, 4^*$’ are calculated.

Now in Figure 6.2.2: (B), we depict the result that occurs in the beacon frame period right after the initiated increase in DL proportion from ‘$x\%$ to ‘$y\%$’. The buffer occupancy level has decreased from the previous ‘B2’ to the now current ‘B3’ position. According to the current active threshold bands ‘$0^*, 1^*, 2^*, 3^*, 4^*$’, it appears that no negative crossing has occurred since both ‘B2’ and ‘B3’ report an index value of 3* . However, as we can see the buffer level has in fact crossed below the previous level threshold band ‘4’ based on $x\%$.

Crossing below the previous ‘4’ band is a significant issue since this was the registered positive crossing required for the increase in DL proportion ‘$x\%$ to ‘$y\%$’. In the current scheme with just the level thresholds ‘$0^*, 1^*, 2^*, 3^*, 4^*$’ present, no negative crossing will be registered and thus no decrease action is taken. A decrease from ‘$y\%$’ would only be registered had a larger buffer depletion amount occurred resulting in the negative crossing from level threshold ‘3’ to ‘2’. But the fact that the buffer occupancy level has dropped below ‘4’, means clearly the increase from ‘$x\%$’ to ‘$y\%$’ has not been sustained.

Therefore in order to exploit this earlier observation of unsustainable DL proportion, we can include the ‘4’ level threshold as part of the list of level threshold bands of ‘$y\%$’. As a result, the crossing of 4’ will thereby initiate a one percentage point decrease in the DL proportion allocation. This process will therefore further reduce the chances of catastrophic buffer depletion occurring.

![Figure 6.2.2: Updated Transitioning between Positive (A) and Negative Level (B) Crossings](image)

Implementing the basic concepts described earlier into the DL control is fairly straight-
forward. The basic procedure involves recording a list of all positive threshold band crossings that have occurred over the course of the network operation. Each positive crossing is recorded in the format where the current DL proportion is stored along with the index of the related threshold band being crossed, ‘\([dDLULprop, index]\)’.

In order to explain this process we once again refer to Figure 6.2.2. During the positive level crossing in Figure 6.2.2: (A), the \(dDLULprop\) was ‘\(x\)%’ and the level threshold band crossed is ‘4’. Therefore we record the following information regarding the positive crossing, ‘\([x,4]\)’. In the case where multiple threshold bands are crossed in one go in a positive manner, each individual crossed band is stored.

Now in Figure 6.2.2: (B), the current DL proportion allocation is at ‘\(y\)%’. As usual we generate the required set of related threshold levels \(index\) related to ‘\(y\)%’. Given that a decrease in buffer occupancy is observed since ‘B3’ less than ‘B2’, we need to identify if a negative threshold band crossing has occurred. Therefore we first determine if a negative crossing has occurred based on the threshold levels \(index\) related to ‘\(y\)%’. Followed by also identifying whether a negative crossing has occurred based on any of the stored list of previous positive crossings for the range of stored DL proportion allocations ‘\(Stored_dDLULprop < y\)%\)’.

In this case, we can see that only a single crossing has occurred, namely across ‘\([x,4]\)’. This results in a reduction from the current DL proportion ‘\(y\)%’ to ‘\(y - 1 = x\)%’. With this operation executed, the entry ‘\([x,4]\)’ is cleared from the positive crossing storage list. In generalized form, this operation is described as clearing all entries within the positive crossing storage list that falls into the range governed by ‘\(Stored_dDLULprop \geq x\)%\)’.

In the situation where no valid level threshold is present in the positive crossing storage list, operation defaults back to the approach described before, the basic principles of ‘\(Multiple Buffer Occupancy Level Crossing\)’.

In the case where the dynamic DL proportion allocation is equal to the target DL proportion allocation, all entries within the positive crossing storage list are cleared. Similarly, the same actions are also taken when through the process of adjusting the dynamic DL proportion allocation it becomes equal to zero and in conjunction, we observe that the AP buffer occupancy is also zero. This decision is taken as an insurance measure for when the DL traffic does come back online, and given the fact that currently, the UL is not being controlled to the target percentage.

Note that the procedures just described are only used for negative level-threshold crossing. For positive crossings the valid bands are those directly evaluated using the current dynamic DL to UL proportion allocation. Note also that any positive level-threshold crossing will always initiate a one percentage point increase in the DL to UL proportion allocation.
6.2.2 Enhancements

In the previous subsection entitled ‘Modifications’ we described two configuration modifications that are necessary inclusions in the DL control. Without them, it is not possible to achieve acceptable levels of performance. Therefore strictly speaking, it can be said that the DL control is basically complete with the addition of the two configuration modifications described earlier. However, we now propose two configuration enhancements which improve the response and speed of the DL control converging to the appropriate operating point.

- Enforced Contention Window Parameter Change
- Direct Increase of DL Proportion Allocation

**Enforced Contention Window Parameter Change**

This enhancement involves directly related contention window parameter changes to be initiated each time the DL proportion allocation is changed. As it stands, contention window parameters are changed in direct relation to the current target DL proportion allocation and the current actual network recorded DL proportion allocation.

We propose applying direct contention window parameter changes whenever a reduction in the DL proportion allocation is initiated. Only negative level threshold crossings will result in a direct contention window change. Furthermore, a single direct contention window parameter change occurs regardless of whether there is a single or multiple negative level threshold crossing. Note that we do not apply direct contention window changes if the buffer occupancy level is below the initial level threshold. Instead we allow the normal DL proportion allocation based operation to occur.

A direct contention window parameter change is one that favours the UL at all times. This is logical given that we have effectively reduced the proportion allocation to DL. Also, given the current network loading, with the UL assumed to be continually saturated, there is always sufficient traffic present and thus overall network throughput performance is maintained more effectively.

Applying this approach results in faster convergence to the appropriate operating point for the network. Additionally it reduces the opportunity for a catastrophic empty buffer occurring, by once again reacting more quickly to early signs of buffer depletion.

**Direct Increase of DL Proportion Allocation**

Thus far the only two ways that the DL proportion allocation can be increased are:

- Crossing a level threshold (other than the minimal threshold) in a positive direction.
The buffer occupancy level is situated above the highest applicable level threshold.

These two methods of increasing the DL proportion allocation are sufficient to achieve the required performance goals, however the response times can be improved. We propose a simple rule of applying a forced increase in the DL proportion allocation whenever the buffer occupancy level remains within a fixed level threshold band for a specified number of consecutive beacon frames, ‘n’.

In the scenario depicted in Figure 6.2.3, we observe a number of buffer occupancy level values ‘B1’ right through to ‘Bn’ over ‘n’ consecutive beacon frames, which have been recorded as lying consistently in the level threshold band of index ‘2’. The observation of ‘n’ consecutive buffer occupancy level values in the same band with no decrease action in the DL proportion allocation being triggered, suggests we can allow a direct increase in the DL proportion allocation to take place. An additional requirement to this observation, is that each ‘n’ consecutive values of the DL saturation load classification must all be the same, to ensure we are in a stable regime.

The final part of this enhancement basically now involves determining the number of ‘n’ required observed beacon frame period results before a direct DL proportion allocation increase can occur. Through extensive testing and evaluation via simulations we arrived at the number $n = 5$ beacon frame periods as providing the best compromise between allowing faster response and the least detrimental effect on achieving the overall performance goals.

![Figure 6.2.3: Direct Increase of DL Proportion Allocation](image-url)
6.3 Performance Evaluation

In Chapter 5 we produced simulation results for a set of simple network configurations. We now evaluate the complete DL control through simulation of the same set of simple network configurations used in Chapter 5.

6.3.1 Performance Metrics

In this chapter, we present simulation results for throughput and proportion allocation averaged over one second intervals. In terms of analysis of the simulation results we use the same performance metrics as those described in Chapter 3 and Chapter 4.

- The average throughput result for DL and UL traffic and the combined total throughput.
- The achieved throughput proportion result (percentage) between DL and UL traffic.
- The specific average throughput results for AC1* and AC1* in DL traffic.
- The specific average throughput results for AC1 and AC0 in UL traffic.
- The achieved throughput proportion result (percentage) with respect to AC1* in DL traffic.
- The achieved throughput proportion result (percentage) with respect to AC1 in UL traffic.

The excursion performance metric and model throughput comparison described in Section 3.5.1 is also reported for each completed simulation. In this case, three excursion performance values are presented for each of the three throughput proportion allocations we aim to control, namely the overall DL to UL throughput proportion (A), the within DL AC1* throughput proportion (B) and the within UL AC1 throughput proportion (C).

6.3.2 Simulation Setup

The simulation tool NS-2 (version 2.29) [47] is used along with an EDCA implementation [48], with the same general MAC and Physical Layer parameters used as those given in Table 3.5.1. Furthermore we assume that the AP is fully aware of the number of active wireless client stations in the WLAN.

We utilize the three network topologies described in Chapter 5 where each network topology is configured in such a manner that we force the network to operate in the following specific scenarios.
• Scenario 1: DL AC1* and AC0* Saturated Load

• Scenario 2: DL AC1* Non-saturated Load and AC0* Saturated Load

• Scenario 3: DL AC1* and AC0* Non-saturated Load

Note that we assume that the UL is continually saturated in all of the three network topologies.

An important configuration modification employed is the need for an increase in the buffer capacity within the DL. While performing the initial testing of the DL control, we determined that adequate performance could not be achieved using the baseline 50 data frame buffer capacity for each individual AC. Through these tests we tried buffering capacities of 100, 150 and 200 frames respectively for each AC within DL. We arrived at an acceptable buffer capacity of 150 data frames. The results about to be presented, will show that there are no negative impacts from employing such a larger buffer capacity. It should be noted that the choice of large buffers can lead to high latencies being experienced by users. As a result, a further avenue of research can be to investigate other means to reduce the need for large buffers if reducing latency becomes an important issue. In the current iteration of the control, we assume that latency issues are not critical given that AC traffic is TCP based.

Scenario 1: DL AC1* and AC0* Saturated Load

The network configuration used to create a fully saturated load environment for DL AC1* and AC0* is depicted in Figure 6.3.1.

Traffic for the UL is generated from the five active source wireless client stations WS1 through to WS5 to the destination external wired node N1, where a single TCP/FTP flow is used each in each wireless client station, for UL AC1 and AC0. In the case of the DL, we have N1 acting as the source and the destination being WS-DL. A single TCP/FTP flow is used each, for DL AC1* and AC0*. The packet size used for all TCP/FTP flows is set to 1000 bytes. The three target saturation proportion allocations are set to, (DL to UL: 80%), (DL AC1* to AC0*: 80%) and (UL AC1 to AC0: 80%). Since this network configuration creates a fully saturated load environment the desired proportion allocations are the same as the three target saturation proportion allocations.

Scenario 2: DL AC1* Non-saturated Load and AC0* Saturated Load

In order to achieve this test case scenario, the network topology used is that shown in Figure 6.3.2.
In this network configuration the DL AC1* TCP/FTP flow source node is N2, where N2 is connected to the AP via a restricted link with bandwidth of only 0.6Mbps. The bandwidth of 0.6Mbps is approximately the throughput AC0* receives in the fully saturated WLAN scenario, with the current given target proportion allocations set to (DL to UL: 80%), (DL AC1* to AC0*: 80%) and (UL AC1 to AC0: 80%). Therefore, AC1* is only able to sustain a 20% allocation of the DL, and so the overall desired proportion allocations are (DL to UL: 80%), (DL AC1* to AC0*: 20%) and (UL AC1 to AC0: 80%).

Scenario 3: DL AC1* and AC0* Non-saturated Load

This scenario requires both DL AC1* and AC0* to be non-saturated, and is achieved by using the network configuration depicted in Figure 6.3.3.

In this network configuration both the DL AC1* and AC0* TCP/FTP flows source node is N2. Where N2 is connected to the AP via a restricted link with bandwidth of only 1.8Mbps. This amount falls well below the 3Mbps that the DL can sustain in a fully saturated network configuration, with a given DL to UL proportion allocation of 80%. The bandwidth of 1.8Mbps for the restricted link coincides approximately to a proportion allocation of just over 45% for the DL to UL (assuming a total bandwidth of 3.75Mbps).
As a result, the expected DL to UL proportion allocation achieved should lie around the 45% mark, also with both DL ACs sharing the bandwidth of 1.8Mbps, the achievable internal DL AC1* to AC0* proportion allocation is 80%. As in Scenario 1 and 2, we retain the same three target saturation proportion allocations. Thus the overall desired proportion allocations are (DL to UL: 45%), (DL AC1* to AC0*: 80%) and (UL AC1 to AC0: 80%).

### 6.3.3 Simulation Results

We present the simulation results in the same manner to the output figures seen in the ‘Buffer Occupancy Level’ section of Chapter 5. The first figure presents the throughput and related throughput proportion allocation results on the same output graph. The second figure shows the trace of the dynamic DL to UL proportion allocation value, the recorded DL saturation indicator results, the TCP congestion window trace and the buffer occupancy level.
TARGET
DL to UL: 80%
DL $AC1^*$ to $AC0^*$: 80%
UL $AC1$ to $AC0$: 80%

DESIRED GOAL
DL to UL: ~45%
DL $AC1^*$ to $AC0^*$: 80%
UL $AC1$ to $AC0$: 80%

Figure 6.3.3: Network Configuration - Scenario 3: DL $AC1^*$ and $AC0^*$ Non-saturated Load

Scenario 1: DL $AC1^*$ and $AC0^*$ Saturated Load

As we can see from Figure 6.3.4: (A), in a fully saturated network the complete DL control scheme maximizes the overall throughput performance (the ‘black’ line in (A)). This result is similar to the earlier results seen in Chapter 5. Additionally it is able to maintain, reasonably well, the throughput proportion allocations, DL to UL, DL $AC1^*$ to $AC0^*$ and UL $AC1$ to $AC0$. In each of the three plots in Figure 6.3.4: (A), (B) and (C) there are no major throughput excursions observed.

The graphs presented in Figure 6.3.5 depict the complete DL control operating as required. The dynamic DL proportion allocation shown in Figure 6.3.5: (A) remains as required consistently on the 80% mark. The DL saturation indicator results in Figure 6.3.5: (B) are also as expected consistently registering a fully saturated reading namely indicator ‘5’. The overall buffer occupancy level in Figure 6.3.5: (D), as expected in a full saturated network scenario, is consistently above the required level. Both DL $AC1^*$ and $AC0^*$ maintain strong buffer occupancy levels.

The average throughput and proportion allocation results for this test case scenario are shown in Table 6.3.1. The average overall throughput and proportion allocation values
are consistent with the earlier basic control configuration described in Chapter 4. All three averaged throughput proportion allocation results lie close to their required target values.

The excursion performance and model comparison results for this test case scenario are presented in Table 6.3.2. Once again the complete DL control’s ability to maintain the throughput proportion allocation is clearly highlighted by the tabulated results. Each of the three proportion allocations record an excursion percentage of under 5%, with the highest being the internal UL proportion allocation control. This is to be expected given the low throughput proportion allocated to UL. The model throughput percentage reported for this single simulation run also shows that adequate performance is being achieved, and the introduced control modifications have not for this single run created any unacceptable performance penalties.

Overall we can conclude that the complete DL control is able to perform to a level that easily achieves the required performance goals in a fully saturated network scenario. The additional modifications in place now have not had any detrimental effects as the results are comparable to the basic saturation-only control scheme described in Chapter 4 (Control Scheme-2).
6.3. PERFORMANCE EVALUATION

Figure 6.3.4: Scenario 1, Throughput Performance - (A): $DL = 80\%$, (B): $(AC1^*) = 80\%$ and (C): $(AC1) = 80\%$
Figure 6.3.5: Scenario 1, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information.
Scenario 2: DL $AC1^*$ Non-saturated Load and $AC0^*$ Saturated Load

Figure 6.3.6 clearly shows the non-saturation constraint placed on DL $AC1^*$ has been adequately handled by the complete DL control. The overall throughput performance is similar to that seen in the previous fully saturated network scenario. Furthermore, the throughput proportion allocations, DL to UL, DL $AC1^*$ to $AC0^*$ and UL $AC1$ to $AC0$ in general all lie close to their desired values. Once again, in each of the three plots in Figure 6.3.6: (A), (B) and (C) there are only no major visible excursions observed.

In results shown in Figure 6.3.7, we can see the complete DL control is achieving the desired results. The dynamic target DL to UL proportion allocation in graph Figure 6.3.7: (A) is consistently on the required 80% mark. The DL saturation indicator results in Figure 6.3.7: (B) are correctly reporting the fact that DL $AC1^*$ is non-saturated and $AC0^*$ is saturated by registering consistent blocks of ‘3’. Also, as expected, the buffer occupancy level results in Figure 6.3.7: (D) show that the buffering is practically all due to $AC0^*$. The average throughput and proportion allocation results for this test case scenario (B) are shown in Table 6.3.1.

Once again the complete DL control’s ability to maintain the throughput proportion allocation is clearly highlighted by the excursion performance results tabulated in Table 6.3.2. The excursion percentage for the DL to UL allocation is just under 3%, while the internal DL and UL excursion percentages registered is under 1% and just below 5% respectively. The relatively static minimal fluctuating throughput of DL $AC1^*$ allows for such a low excursion performance to be achieved. We can see again that the control modifications introduced has not damaged throughput performance, as for this single simulation run, we achieve a throughput percentage well on the mark reported by the analytical model.

We can once again conclude that the complete DL control achieves the required performance goals in the scenario where one DL $AC$ is non-saturated.
Figure 6.3.6: Scenario 2, Throughput Performance - (A): DL = 80%, (B): (AC1*) = 80% (20%) and (C): (AC1) = 80%
Figure 6.3.7: Scenario 2, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information
Scenario 3: DL AC1* and AC0* Non-saturated Load

This particular test case scenario, has proven to be the most difficult to achieve the required performance goals. Referring back to Chapter 5, with just the three necessary components as part of the DL control we described in detail the two issues that remained. These were, the inability to maintain a consistent value for the dynamic DL to UL proportion allocation and the subsequent buffer depletion. These factors all contributed to the inability to maintain the required target DL internal proportion allocation at 80%. The primary issue was with the inconsistent dynamic DL to UL proportion allocation. Once we artificially fixed the value to 45%, the correct performance was then obtained.

With the introduction of our complete DL control, we see from Figure 6.3.8 that we have made a vast improvement. We achieve the goal of maximizing throughput performance. Additionally we can see that the DL to UL proportion allocation settles around the desired level. Therefore the UL has been able to take up the left over bandwidth, that the DL is unable to use.

The real benefit of the complete DL control modifications can be seen in Figure 6.3.8: (B), where the goal of being able to maintain the required internal DL proportion allocation of 80% is met. During the initial short period of 25s - 35s, the required performance is not attained, however as will be explained later this is not a critical issue.

As for the UL internal control, we can see once again, it is achieving the required proportion allocation between AC1 and AC0 of 80%. The issues relating to the performance during the initial 25s - 35s are a result of the complete DL control adjusting the dynamic DL to UL proportion allocation. As we can see from Figure 6.3.9: (A), the dynamic DL to UL proportion allocation is on a continuous downward path from the initial value of 80%. Afterwards we can see that the dynamic DL to UL proportion allocation settles into a fairly consistent value.

Even with these issues present, the important fact remains that consistent control is maintained over the dynamic DL to UL proportion allocation. The complete DL control is able to allow consistent buffering of data frames. Thus with the buffer occupancy levels being maintained, this allows the network to operate in a ‘pseudo’ saturated load environment. This is clearly highlighted by the results in Figure 6.3.9: (B), where we register for the majority of the simulation period saturation indicator results of ‘5’.

The average throughput and proportion allocation results for this test case scenario (C) are shown in Table 6.3.1. The average total throughput result are once again on the mark with what is reported by the analytical model. In the case of the averaged throughput proportion allocation results, all are within an acceptable range of their respective target or desired values.

The complete DL control’s ability to maintain the throughput proportion allocation
is highlighted by the excursion performance metric results tabulated in Table 6.3.2. The initial issues faced with the DL to UL proportion allocation during the time period 25s - 35s have not had a large impact as we only registered a percentage of just over 4%.

The internal DL and UL percentage in this case is highest at just over 6%, this may appear quite large but from the output results this still more than acceptable, given a clear throughput advantage is always being provided to DL AC1'. In terms of the UL internal control, the total impact percentage is marginal, at below 3%.

We conclude that the introduced complete DL control modifications achieves the required performance goals in the scenario even when both DL ACs are non-saturated.
Figure 6.3.8: Scenario 3, Throughput Performance - (A): DL = 80% (45%), (B): (AC1*) = 80% and (C): (AC1) = 80%
**Figure 6.3.9:** Scenario 3, DL Trace Log and Buffer Occupancy Log - (A): DL to UL Proportion, (B): DL Network Load Classification Regime, (C): TCP Congestion Window and (D): Buffer Occupancy Information
Table 6.3.1: Average Results for Various Throughput Proportion Allocations.

<table>
<thead>
<tr>
<th>Test Case Scenario</th>
<th>Total</th>
<th>DL</th>
<th>UL</th>
<th>AC1*</th>
<th>AC0*</th>
<th>AC1</th>
<th>AC0</th>
<th>DL</th>
<th>AC1*</th>
<th>AC1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(Mbps)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>1</td>
<td>3.85</td>
<td>2.98</td>
<td>0.87</td>
<td>2.34</td>
<td>0.64</td>
<td>0.19</td>
<td>77.5</td>
<td>78.4</td>
<td>78.3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.71</td>
<td>2.91</td>
<td>0.80</td>
<td>0.60</td>
<td>2.31</td>
<td>0.62</td>
<td>0.18</td>
<td>78.5</td>
<td>20.6</td>
<td>78.0</td>
</tr>
<tr>
<td>3</td>
<td>3.75</td>
<td>1.72</td>
<td>2.03</td>
<td>1.30</td>
<td>0.42</td>
<td>1.60</td>
<td>0.43</td>
<td>45.9</td>
<td>75.5</td>
<td>78.8</td>
</tr>
</tbody>
</table>

Table 6.3.2: Excursion Performance Metrics and Model Throughput Comparison.

<table>
<thead>
<tr>
<th>Test Case Scenario</th>
<th>DL to UL (%)</th>
<th>DL AC1* (%)</th>
<th>UL AC1 (%)</th>
<th>Model Throughput Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.2</td>
<td>2.7</td>
<td>4.5</td>
<td>104.6</td>
</tr>
<tr>
<td>2</td>
<td>2.7</td>
<td>0.8</td>
<td>4.6</td>
<td>100.8</td>
</tr>
<tr>
<td>3</td>
<td>4.2</td>
<td>6.4</td>
<td>2.8</td>
<td>100.8</td>
</tr>
</tbody>
</table>
6.4 Summary

In this chapter, we have provided the full description of each component that make up the complete DL control modifications for *Control Scheme-2*. The complete DL control was developed in order to fully achieve the performance requirements described in Chapter 5. Furthermore, we highlighted the crucial benefits of the modifications and enhancements employed, along with the basic three necessary components from Chapter 5.

Using the network configurations described in Chapter 5, through simulation results we have shown that the complete DL control is able to achieve the required performance goals across the various DL traffic load conditions. The complete DL control is able to adequately control the proportions in the three scenarios listed below.

- Scenario 1: DL AC1* and AC0* Saturated Load
- Scenario 2: DL AC1* Non-saturated Load and AC0* Saturated Load
- Scenario 3: DL AC1* and AC0* Non-saturated Load

Overall the aspect of dealing with non-saturated DL traffic loading scenarios, is well managed by the complete DL control, where we have assumed that the UL is fully saturated. The next objective will also be to deal with non-saturated traffic load in the UL and thus complete all the required modifications to *Control Scheme-2*. 
Chapter 7

Uplink Traffic Saturation Detection

7.1 Introduction

In Chapter 6, the complete DL side modifications to Control Scheme-2 was shown to be able to meet the required performance goals when non-saturated DL traffic load was present in the network. The throughput was maximized and the allocated proportion between the DL and UL and each AC was maintained at their expected target or desired levels. However, an important requirement in the control scheme was that the UL was assumed to be saturated. As a result each AC within a client wireless station always has a data frame awaiting transmission. The scenario of continuously saturated ACs within each wireless client station is unrealistic.

We begin this chapter by defining some important requirements of an acceptable UL load detection scheme for use in the complete control scheme where the UL is not assumed to be saturated. This is followed by an investigation of various schemes that can be used to detect the presence of saturation and non-saturation traffic load within the UL. This is followed by a detailed description and discussion of some of the detection schemes that can be used based on the related literature. We then present our proposed UL detection scheme in detail and the subsequent variations that can be used.

7.2 Detection Scheme Requirements

The primary goal of a UL detection scheme, is to distinguish between when the UL is operating under saturation and non-saturation load, so that the information it provides can be used in the complete control scheme.

The UL detection scheme should reside within the AP as this reduces the overall complexity of the system, so that no modifications are required to the wireless client stations within the network. Additionally, with both the UL detection scheme and complete
control scheme residing within the AP, information can be directly integrated.

The UL detection scheme itself, should be simple and not require complex computational procedures to evaluate the load conditions in the network, in keeping with the simplicity of our proposed control scheme.

An important and primary requirement for the detection scheme is that it must work passively without interfering with the normal operation of the network. This means it should not introduce any additional transmission overhead, in the form of management frame transmissions or alter the existing MAC frame header structures, strictly adhering to the IEEE 802.11e EDCA standard.

A direct copy of the DL based solution used in the previous chapters is not possible, as we can not determine directly at the AP the following information from each wireless client station:

- Contention window countdown and back-off process.
- Exact buffer occupancy level within each wireless client station.

Thus, with the requirement of a passive UL detection scheme, only network statistics that the AP is capable of measuring can be used. Further, the UL detection scheme must determine the load conditions in the wireless LAN for the UL in a short period of time, given that the control scheme updates the parameter sets for each AC on a per beacon frame basis. Finally, a critical assumption is that the AP is aware of the number of active wireless client stations within the wireless LAN.

7.3 Available Network Statistics in the AP

The primary objective is to develop a UL detection scheme that only uses the information available from the network statistics measured by the AP. The following are the primary network statistics that can be measured and recorded in the AP.

Throughput Statistics:

The AP is able to observe all transmissions within its service set, so that for each successful data frame transmission, the AP can obtain the following information:

- which wireless client station is responsible for initiating each data frame transmission.
- the AC classification information for each UL data frame transmission in the wireless LAN.
• the size of the data frames transmitted by each wireless client station.

From the above information recorded by the AP, the throughput statistics for each AC from an individual wireless client station or an overall global network perspective can be obtained.

**Idle Time Statistics:**

An idle time statistic from the point of view of the AP is, the duration of time when the network is idle (no transmission activity is occurring on the channel). Therefore a record of the idle times between data frame transmissions in the network can be obtained.

**Successful Transmission and Collision Statistics:**

The number of successful data frame transmissions for each AC and hence the number of collisions observed can be recorded by the AP. However, collisions observed by the AP, do not provide information on the actual number of data frame transmissions or AC type involved in the collision.

Hence we restrict our view to detection schemes, which are based on analysis of the limited information described above.

### 7.3.1 Throughput Load Analysis in the Literature

In [28], Engelstad and Osterbo propose several extensions to the analytical model of the IEEE 802.11e EDCA mechanism by Xiao [64]. They incorporate the additional features of the EDCA mechanism, namely the behavior of virtual collisions and AIFS differentiation. However, their most important extension is the ability to describe the performance of the EDCA mechanism not only in the saturated case as in [64], but the whole range from non saturated up to saturated conditions.

This result is achieved by introducing an additional probability parameter that also governs the Markov chain for the transmission process of an AC in [64]. The parameter is referred to as the utilization factor, $\rho_i$ (where $i = 0, 1, 2, 3$), which represents the probability that there is a data frame waiting in the transmission queue of AC$_i$ (where $i = 0, 1, 2, 3$) of a wireless client station. As a result with $\rho_i = 1$, the Markov chain models the saturated AC case shown in [64]. On the other hand, when $\rho_i < 1$, the Markov chain models a non saturated AC. Therefore, their critical issue lies in the evaluation and estimation of $\rho_i$.

A $G/G/1$ queue is used to describe the data frame queuing behavior for each AC$_i$ (where $i = 0, 1, 2, 3$) of each wireless client station in order to evaluate the values for $\rho_i$. 
The method used to obtain the value of $\rho_i$ in both a non-saturated and saturated scenario is described in detail in [28]. Then using the $\rho_i$, a set of non linear equations governing the analytical model in [28] can be solved.

In [28], their proposed detection scheme determines when the offered load of each $AC$ for a given number of wireless client stations will begin to cause a throughput starvation for lower priority $AC$s. Throughput starvation refers to the point when the serviced load for lower priority $AC$s does not match its offered load. This is effectively the point of onset of overall saturation of the channel.

When the offered load is sufficiently small in each $AC$ for a given number of wireless client stations, the related $EDCA$ parameters have no impact in the form of providing service differentiation between low and high priority $AC$s. The service differentiation phenomena only begins to take effect when offered load increases.

Thus the point at which the offered load for each $AC$ causes there to be a discernible service differentiation between low and high priority $AC$s is referred to as the throughput starvation point. Effectively, the low priority $AC$s serviced load is no longer the same as that of its offered load. This explanation can be taken further to basically conclude that the transition from non-saturation to saturation load conditions is taking place within the network.

The proposed detection scheme resides in the $AP$ in the network since, the $AP$ is aware of the parameter set of each $AC$ and it can also easily record the throughput usage in the network.

In [28], they monitor the overall probability that the channel is busy ‘$p_b$’, and whenever $p_b \geq \frac{1}{1-AIFSN_i}$, they record starvation of $AC_i$ and hence saturation of the entire channel. This therefore depicts the point at which we can conclude the transition between non-saturation and saturation traffic load is about to occur.

As the case with most of the highlighted related work, the number of wireless client stations present in the network must be known beforehand by the detection scheme proposed in [28]. The proposed detection scheme in [28] requires the use of an analytical model, and any errors in the analytical model will result in inaccurate predictions of the starvation points for lower priority $AC$s. The complexity of their detection scheme is increased significantly due to the need for complex numerical techniques to solve the analytical model.

Their detection scheme relies on recording the throughput statistics in the network in order to compare to the predicted starvation points from the analytical model. Therefore an appropriate time scale must be used to record the throughput statistics in order to perform accurate and fast predictions of throughput starvation.

A critical assumption in the detection scheme proposed in [28] is that the parameters
set for each AC are constant. This is necessary in order to correctly compare the predicted starvation points to the recorded statistical results. Further, the relationship mentioned earlier, is only applicable when AIFS differentiation is used in the network for each AC.

Thus given these requirements of the detection scheme proposed in [28], we highlight the following major issues.

- The detection scheme is based on a complex analytical model, as a result the computational load of the AP is increased.

- Simplified analytical relationship based detection scheme in [28] relies on AIFS differentiation being present. This is not the case in our proposed control scheme where AIFS differentiation is omitted.

- Re-evaluation of the predicted starvation points from the analytical model at each beacon frame period is required since the parameter set may be changed in each beacon frame period by our proposed control scheme.

- Accuracy of comparison between measured throughput statistics requires the correct time scale. This may conflict with our proposed control scheme, which changes parameter set for each AC at each beacon frame period.

As a result, the detection scheme proposed in [28], is inadequate to use in conjunction with our proposed control scheme.

7.3.2 Idle Time Analysis in the Literature

In [65], Coyle proposed an alternative and novel approach to analyze the behavior of networks based on a CSMA/CA protocol. The analysis begins with the CSMA/CA protocol used in the IEEE 802.11 DCF standard, and is later extended to include the IEEE 802.11e EDCA standard.

Coyle [65] investigates the distribution of idle time durations in an IEEE 802.11 DCF and IEEE 802.11e EDCA network under the critical assumption that wireless stations are saturated when attempting data frame transmissions. The idle time in this case, is the time duration between consecutive successfully transmitted data frames. Coyle [65] represents the idle times as an integer number of slots (a slot is the duration of a single back-off counter).

From the investigations of actual idle time distributions in a network, it is shown and described in detail, that an analytical formula can be derived which can calculate the idle time distributions based on a function of the collision probability in the network. The idle time distribution analytical formula can also be used to evaluate additional system
parameters (as described below), as is the case in other analytical models [20, 21, 22, 23, 26, 27, 28].

The analytical result for the idle time distribution evaluated in [65], can also be used in designing a possible load detection scheme, where the AP can be configured to record the idle time durations in the network. In such a detection scheme the number of wireless client stations must be known a priori and given the parameter set for each AC which is known by the AP, a histogram of the actual idle time distributions can be generated.

The analysis in [65] can then be used to generate analytic idle time curves for an assumed saturated network configuration, which can be compared to the measured idle time histogram. If an accurate fit is obtained between the analytical curve and the measured idle time histogram, then the detection scheme can assume that saturation conditions are present in the network, and conversely if a fit is not achieved, then the network is non-saturated.

A detection scheme based on [65] requires an analytic idle time curve to be generated each time the number of wireless client stations changes or the parameter set for an AC changes. This increases the complexity and computational load of the detection scheme, since parameter set updates for each AC occur at each beacon frame period in our proposed control scheme.

This detection scheme relies on a histogram of the measured idle times, which raises the issue of selecting an appropriate time scale such that an accurate and appropriate speed of saturation load detection is achieved. Furthermore, with parameter set updates occurring at each beacon frame period, this could lead to inaccurate saturation load detection.

Given the issues discussed above, a detection scheme based on [65], would be unable to meet the requirements needed in order to operate in conjunction with our proposed control scheme.

### 7.3.3 Transmission and Collision Probability Analysis in the Literature

A detection scheme based on transmission and collision probability analysis, could be constructed by comparing the value for the transmission and collision probabilities obtained from an analytical model [20, 21, 22, 23, 26, 27, 28], to the values measured in the network from the AP.

A detection scheme that relies on this analysis requires calculation of model based expected probabilities each time the network changes. Furthermore, measurement of probabilities taken from the network have to be established with minimal information in minimal time leading to poor estimates which are then compared to long term model
based expected probabilities.

In a dynamic environment long term probabilities are not realistic and calculating probabilities from measurements are extremely unreliable from minimal data. Even if sufficient data could be gathered, a dynamic environment implies that such data would be useful only in retrospect. Thus in a beacon frame by beacon frame control scheme this is not sensible and therefore inappropriate for our use.

### 7.4 Proposed UL Detection Scheme

The possible detection schemes based on the related literature are inadequate for our requirements, and so we propose an alternative UL specific detection scheme. The UL detection scheme we propose overcomes the issues highlighted in each of the other possible detection schemes while achieving the key requirements discussed earlier.

#### 7.4.1 Description

The primary goal of our UL detection scheme is to integrate it into the current control scheme as defined in Chapter 6. It should notify the control scheme at a given time interval, if the UL is non-saturated or saturated. Therefore, using this information, we can then define how the control scheme can act accordingly and apply the the required parameter set for each AC in the network.

The control scheme described in Chapter 6, updates parameter sets as necessary, at every beacon frame interval. The UL detection scheme should therefore also operate on a beacon frame interval resolution, as we require a relatively fast detection scheme. Thus, in order to explain the operation of the proposed UL detection scheme, we need to understand the specific actions that take place during the process of transmitting a data frame.

In Figure 7.4.1, the wireless client station SS1 is shown to have multiple data frames awaiting transmission, each data frame is of type AC1. To begin the transmission process of a data frame, SS1 has to first observe that the transmission channel is idle for the given AIFS1 period. If the transmission channel is detected as being idle for an AIFS1 period, SS1 then initiates its back-off phase. During the back-off phase, SS1 must observe the transmission channel being idle for an additional period of time defined by the random number of slotted time intervals chosen from the range governed by the CW value for AC1. In this example the CW value is seven and the first random number of slots chosen is equal to four. The slot number count is then decremented each time the transmission channel is detected as being idle for a slotted time interval, and at the end of which data frame ‘Frame 1’ is transmitted.
Each time a data frame is successfully transmitted by a wireless client station or the AP, an additional procedure referred to as post back-off is initiated, which simply involves the wireless client station initiating another attempt to obtain access to the transmission channel. Thus from Figure 7.4.1, after SS1 successfully transmits ‘Frame 1’, post back-off is initiated with a new random number of slots chosen in this case, to be five.

During the back-off or post back-off phase if the transmission channel becomes busy as shown in Figure 7.4.1, the count down process of the random number is paused. SS1 is then required to wait for the transmission channel to be idle for another AIFS1 period, before it recommences the countdown procedure. Once the random number count is successfully decremented, SS1 can again initiate a transmission, ‘Frame 2’.

For reference, note that in the case of a post back-off phase being completed, the wireless client station or the AP are allowed to begin transmission immediately if a data frame was queued awaiting transmission prior to the start of the post back-off phase, or a data frame arrived in the transmission queue before the completion of the post back-off phase.

Figure 7.4.1: Transmission Characteristics - Saturation Load

In Figure 7.4.2, the example of post back-off being completed with no data frames present in the transmission queue is shown. We can see that SS1 will simply remain inactive and not attempt to access the transmission channel. If a data frame arrives after the completion of post back-off, then SS1 is required to first observe the transmission
channel being idle for the $AIFS_1$ period alone before it can immediately commence transmission. However, if the transmission channel is detected as being busy, the $SS1$ will initiate a new back-off process.

![Figure 7.4.2: Transmission Characteristics - Non Saturation Load](image)

From the described transmission procedure of a wireless client station, we refer to an idle time as basically the time duration between a pair of data frame transmissions for a wireless client station. This idle time is directly related to the random number of slotted intervals chosen in the back-off process. Therefore, saturation of a wireless client station can be inferred from observing the idle time duration between each of its data frame transmissions if the back-off interval is known.

As a result, Figure 7.4.1 depicts the data frame transmission characteristics of a wireless client station under saturation load, as effectively there is always a data frame awaiting transmission at the completion of each back-off or post back-off phase. The effective idle times observed between each data frame transmission for a wireless client station exhibits the relation given in Eq.(7.4.1). Note that as discussed in Chapter 2, an elapsed $AIFS$ interval will also account for a count down in the slot value.

$$\text{Idle\_time} \leq \text{Idle\_slots} \times (AIFS_i - \sigma) + (\sigma \times CW_{min}), \quad (7.4.1)$$

where:
7.4. PROPOSED UL DETECTION SCHEME

- **Idle time** is the time duration between consecutive successfully transmitted data frames, for a $AC_i$ on the given wireless client station.

- **Idle slots** is the number of times the transmission channel is idle between consecutive successfully transmitted data frames, for $AC_i$ on the given wireless client station.

- **AIFS$_i$** is the time duration a wireless station must observe an idle transmission channel before commencing the back-off interval for a data frame of type $AC_i$.

- ‘$\sigma$’ is the time duration of a single slotted time interval, this is directly related to the physical layer implementation being used.

- $CW_{\text{min},i}$ is the contention window size (under no collision scenario) that the random number of slotted time intervals is selected from and used in the back-off interval for data frame type $AC_i$.

Conversely Figure 7.4.2, depicts the transmission characteristics of a wireless client station under non-saturation load, since we can see that a post back-off phase elapsed without a data frame transmission occurring. In this case, the idle time observed between two data frames will not necessarily obey relationship, Eq.(7.4.1). An important assumption used in the relationship given in Eq.(7.4.1), is that we are unable to determine the actual random back-off interval selected by the wireless client station. Hence, we can begin by approximating this problem by assuming an upper limit on the selected back-off intervals by using $CW_{\text{min},i}$ in the relationship.

The use of the upper limit $CW_{\text{min},i}$, effectively means that a wireless client station that exhibits an idle time larger than that described in the relationship Eq.(7.4.1) is guaranteed to have been non-saturated for this transmission.

Therefore, in our proposed UL detection scheme, we can test Eq.(7.4.1) to identify guaranteed non-saturation idle time periods of a wireless client station. To do this, we need to configure the AP in a network to record the idle time durations for each wireless client station and each $AC$ traffic type.

With the detection scheme implemented in the AP, the $AIFS_i$ period and $CW_{\text{min},i}$ configuration is known in each beacon frame period. Given that post back-off is initiated after each successful data frame transmission, the $CW$ value used in the post back-off process is chosen directly from the $CW_{\text{min},i}$ value applied in the beacon frame period of the observed successfully transmitted data frame. Since this information is readily available, it can be used in the classification process of each total observed idle time duration for each wireless client station.

Since we have used the upper limit of $CW_{\text{min},i}$, we can guarantee there will be no “no false positive”, since any wireless client station detected as being non-saturated is
truly non-saturated. However, we may observe potentially some “false negatives”, where a non-saturated wireless client station may be classified as being saturated.

A critical factor omitted thus far, is the impact of internal and physical collisions. From the AP perspective, internal collisions within a wireless client station between $AC_1$ and $AC_0$ data frames cannot be resolved and we can no longer explicitly guarantee no false positives in the case of $AC_0$ transmissions. However, no false positive results are still guaranteed in the case of $AC_1$ transmissions when internal collisions occur, as they are transmitted regardless.

In the case where we observe physical collisions, there are three different approaches that we consider in classifying the related idle times.

(A) An idle time duration that records a physical collision will be discarded and not be classified.

(B) Physical collisions result in the doubling of the $CW$ value each time from the initial $CW_{min}$ value.

(C) Physical collisions do not result in the doubling of the initial contention window value.

(A) Discarding Physical Collisions

This approach is the simplest because for any idle time period during which collisions take place, we discard it and record no classification. Hence, only idle time periods with no collisions will be classified, as being saturated or non-saturated, based upon the relationship Eq.(7.4.1). Note that as described before we still use the contention window value $CW_{min}$. In this case, the principle of maintaining no false positive classifications for $AC_1$ transmissions will remain true.

(B) Doubling the Contention Window Value

The basic principle behind this approach is to try and mimic the real actions taken by a wireless station upon suffering a collision. The first step is to see that the relationship in Eq.(7.4.1) is no longer exactly valid and requires modification. Previously we expressed an idle time duration as simply the total period of time that no channel activity is detected between a pair of $AC_i$ data frame transmissions for a specific wireless station. In the case of collisions an idle time duration can now be seen as the total period of time between a data frame transmission and a collision, and subsequently in between collision to collision and collision to data frame transmission. Therefore the initial case of a data frame to collision, the idle time classification still refers back to the relationship used in Eq.(7.4.1).
However for the subsequent idle time durations after this initial collision, and all remaining
detected collisions, each time the contention window will be doubled as per the protocol.
Thus the new idle time relationships used after the first detected event of a collision is
shown in Eq.(7.4.2).

\[
\text{Idle time}_C \leq (T) + \text{Idle slots}_k \times (EIFS_i - \sigma) + (\sigma \times CW_{i_k}),
\]

(7.4.2)

where:

- \( \text{Idle time}_C \) is each of the idle time durations observed after the first collision and
  subsequent others leading up to the final data frame transmission of specific pair in
  question.

- ‘\( k \)’ indicates the specific pair of collisions or the last collision and the last data frame,
  that is used to calculate the related idle time.

- ‘\( T \)’ is the time-out duration used in the IEEE 802.11e EDCA standard to identify
  a collision has occurred.

- \( \text{Idle slots}_k \) is the number of times the transmission channel is idle between a pair of
  collisions or a collision and the last data frame, specific pair in question, as indicated
  by ‘\( k \)’.

- \( EIFS_i \) is the extended inter-frame spacing interval used as the defer period for \( AC_i \)
  by wireless client stations involved in collisions.

- \( CW_{i_k} \) is the contention window size used after each collision for \( AC_i \). The contention
  window size used is based on the first collision in the pairing indicated by ‘\( k \)’.

The new relationships shown assume that each collision involves each and every wireless
client station. This assumption is clearly not realistic. However, given the fact that the
\( AP \) is unable to determine the actual wireless client stations involved in each collision,
we have to apply this assumption if we still wish to maintain the situation of no false
positive classifications for \( AC1 \) transmissions. Note that the post back-off contention
window value is doubled each time for each observed collision.

In addition to the assumption we make of ‘\( all \ wireless \ client \ stations \ are \ involved \ in \ collisions \)’,
we diverge slightly from the rule related to the short retry limit of seven. The
IEEE 802.11e EDCA standard [1] describes that in the advent that seven collisions ( or
retransmissions) have occurred, the current frame is discarded and the back-off process is
reset back to using \( CW_{\text{min}} \) for the next frame awaiting transmission. In our case, we do
not simply reset back the countdown process to $CW_{\text{min}}$ and move to the next transmission pair. Instead we continue the count down process using whatever the existing contention window value is from the \textit{seventh} collision.

\textbf{(C) Retain $CW_{\text{min}}$ Contention Window Value}

This approach is similar to \textbf{(B)}’, but with the key difference being that the contention window value is not doubled upon observing collisions. That is, we retain the initial $CW_{\text{min}}$ throughout the classification process of each idle time period. This simpler approach no longer guarantees that there will be no false positives.

\textbf{SUMMARY}

Each of the described approaches is designed to operate on a per beacon frame period, so that at the end of each beacon frame period, a simple record of classification results is obtained for each active wireless client station. This record is the total number of saturation and non-saturation idle time durations for that beacon frame period for each AC related to a particular active wireless client station. An idle time is essentially the period of time between two data frame transmissions (case (B) includes collisions). However, we include the exception that an idle time duration can also be classified as the time duration between the end of a data frame transmission and the end of a beacon frame period. This allows a classification reading to always be recorded for each elapsed beacon frame period.

In order to obtain an overall representation of the level of UL saturation load for AC1 and AC0, we first determine the total number of saturation and non-saturation idle time periods for UL AC1 and UL AC0 across all wireless client stations. Once these values are obtained, we can finally calculate for each beacon frame the overall UL AC1 and UL AC0 saturation percentage amounts.

\section{Comparison with Ground Truth}

We have defined three approaches that can be used in a real-time UL detection scheme, (A), (B) or (C). In order to determine which is best suited and therefore to be used in the final implementable UL detection scheme, we compare each approach to ground truth. The ground truth can be determined in a manner analogous to the DL Classification Mechanism, in that we are able to determine the exact back-off process of each individual wireless client station.

Note that as mentioned before, obtaining the exact back-off process of each wireless client station is not feasible in a real world implementation of a UL Detection Scheme, but
serves the role of gauging which of the three feasible approaches (A), (B) or (C) should be used in the final implementable UL detection scheme.

### 7.5.1 Description of Ground Truth Evaluation

In an ideal situation, we are able to observe the back-off process for AC1 and AC0 running in each wireless client station and record in each beacon frame for each AC in a wireless client station the total number of saturated and non-saturated transmissions. A saturated transmission is simply when the back-off process countdown elapses and a data frame transmission, attempted or an internal collision occurs. A non-saturated transmission occurs sometime after the countdown elapses.

Using this recorded information, we can then calculate the total number of saturation and non-saturation transmissions for UL AC1 and UL AC0 across all wireless client stations. Once these values are obtained, we finally calculate for each beacon frame, the overall UL AC1 and UL AC0 saturation percentage amounts.

### 7.5.2 Method for Comparison

In order to compare each of the three approaches we simulate the following three different forms of traffic load for the UL.

1. Fully saturated load for both UL AC1 and UL AC0.

2. Near saturation load for both UL AC1 and UL AC0.

3. High degree of non-saturation load for both UL AC1 and UL AC0.

Note that in each of these three scenarios, the DL is continually saturated for both AC1* and AC0*, and we retain fixed CW parameters for the duration of the simulations. This ensures that we are only comparing the detection process in each scheme as each detection scheme sees the same simulation.

For each of the three test case scenarios described above, we report the per beacon frame results of the overall saturation percentage for UL AC1 and UL AC0. The ground truth is able to identify and distinguish between the three test case scenarios by reporting markedly different saturation percentage result plots for UL AC1 and UL AC0. Thus the goal is to compare, which of the three approaches (A), (B) or (C), is able to report saturation percentage result plots for UL AC1 and UL AC0 that can also identify and distinguish between the three test case scenarios. Thus there is no need to perform a direct comparison to the ground truth method.
7.5.3 Network Simulation Configuration

The three test case scenarios described before are created using the network configuration depicted in Figure 7.5.1. Since the DL traffic is specified to be continually saturated, we configure N2 to be the source node and wireless client station WS DL to be the destination node. Two TCP flows originate from N2, which are categorized as AC1* and AC0* within the DL. It is clear that continually saturated DL traffic is generated given that the bandwidth of the link between N2 and the AP is 1000Mbps.

In the case of the UL traffic, each wireless client station WS 1 through to WS 5 acts as a source node for an AC1 and an AC0 TCP flow. The destination node for the UL traffic is configured to be N1, and as we can see, by adjusting the bandwidth of ‘X’ Mbps we can create the required test case scenario.

For comparison to each of the required test case scenarios, we utilized fixed CW parameters for DL and UL. The fixed CW parameters used are presented in Table 7.5.1. Note that the chosen CW parameters are optimal for the scenario of five wireless client stations and the AP each transmitting both AC type traffic.

Finally, the simulation duration used is the same as in all previous cases, 60 seconds (600 beacon frame periods).

Table 7.5.1: Fixed CW Parameter Set Configuration.

<table>
<thead>
<tr>
<th></th>
<th>DL AC1*</th>
<th>DL AC0*</th>
<th>UL AC1</th>
<th>UL AC0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW min</td>
<td>63</td>
<td>63</td>
<td>255</td>
<td>255</td>
</tr>
<tr>
<td>CW max</td>
<td>2047</td>
<td>2047</td>
<td>8191</td>
<td>8191</td>
</tr>
</tbody>
</table>

7.5.4 Comparison of Results

We compare the reported saturation load percentages for AC1 and AC0 within the UL from a ground truth basis and for each of the three feasible approaches ((A), (B) and (C)). We present results in the form of graphs with the x-axis representing the UL AC1 saturation percentage, and the y-axis representing the UL AC0 saturation percentage, as measured by the ground truth and the three approaches, all presented together on the same page in separate graphs.

1. Fully saturated load for both UL AC1 and UL AC0

In order to create the test case scenario, we configure the link bandwidth between the AP and N1 to be ‘X’ = 1000Mbps. Thus allowing fully saturated load for both AC1 and
7.5. **COMPARISON WITH GROUND TRUTH**

With the correct network conditions attained we compare the recorded UL $AC_1$ and UL $AC_0$ saturation percentages under ground truth, and for the three feasible approaches (A), (B) and (C) in Figure 7.5.2.

There are 600 recorded saturation percentages plotted for each beacon frame period on each graph with many reporting the same percentage values, thus they are not all individually visible since they lie on top of each other. The saturation percentages shown for the ground truth are as expected, very consistent and reflect the current saturation traffic load scenario in UL. We can see that there are only five distinct saturation percentages registered for both $AC_1$ and $AC_0$ within the UL. Four out of the five registered values fall below the 100% mark, and these are simply a result of the start-up behavior of the individual UL $TCP$ flows from each wireless client station. We expect a brief period of non-saturation to occur and this reflected in these four values. The majority of saturation percentages recorded (total 594 beacon frame period recordings) are all as expected at the 100% mark for both $AC_1$ and $AC_0$.

The saturation percentages shown for Approach (A) are not too dissimilar for this test
case scenario to that of the ground truth. Three saturation percentage values registered for both $AC_1$ and $AC_0$ fall well below the 100%, these three are again related to the start-up behavior of TCP flows. Aside from these three values, the majority lie at 100% with a few exceptions where $AC_1$ registers 100% saturation but $AC_0$ falls between 80% and 100%. This behavior is due to the fact that internal collisions can not be detected for $AC_0$, and thus we incorrectly report a saturation percentage below 100%.

Approach (B), is also quite similar to Approach (A) in terms of the reported saturation percentages for $AC_1$ and $AC_0$ within the UL. We can see the characteristic few saturation percentages that fall below the 100% mark, and likewise the effects of internal collisions causing a few of the reported saturation percentages for $AC_0$ to be below 100%, while the related $AC_1$ saturation percentage is right on the 100% mark.

In the case of Approach (C), we observe a completely different saturation percentage record for UL $AC_1$ and UL $AC_0$ as compared to the previous three detection configurations. We observe a large cluster of saturation percentage results grouped together within the square defined by the percentages of 75% to 100%. This is a result of Approach (C)’s method of dealing with physical collisions, where a collision is simply treated as occupying transmission time (channel busy), thus we assume that $CW_{min}$ is always used by the wireless client stations.

We know that wireless client stations, which are truly involved in a collision will double their contention window value and begin a new back-off process, while stations not involved in the collision will effectively continue their current back-off based on $CW_{min}$. Therefore this behavior is due to the stations being involved in collisions that Approach (C) incorrectly registers as being non-saturated and thus reports a lower overall saturation percentage than a 100%.

Collisions are correctly identified by the ground truth. In the case of Approach (A), anytime a collision is detected the idle time period between two successful transmissions surrounding a collision is simply ignored and remains unclassified. Therefore a clean idle time free of collisions, is expected be saturated given the current test case scenario. In Approach (B) we account for collisions in a realistic (but conservative) way by doubling the contention window value and restarting the countdown process. However, the unrealistic aspect of this is applying the doubling of contention window value to all wireless client stations. In this particular test case scenario applying this unrealistic rule just causes everything to appear saturated.

Overall, the three feasible approaches provide a distinct clear and identifiable pattern with regards to the recorded saturation percentages in this test case scenario. For Approaches (A) and (B), we can see a large majority of recordings lying on the 100% mark. As for Approach (C) we can still see that a majority of recordings cluster within the area
defined between 75% and 100% on each axis. In fact the main count is towards the upper right of this area.
Figure 7.5.2: UL Saturation Percentage - Ground Truth, Approach (A), Approach (B) and Approach (C)
7.5. COMPARISON WITH GROUND TRUTH

2. Near saturation load for both UL AC1 and UL AC0

In this test case scenario we require near saturation load conditions to be created for both UL AC1 and UL AC0 and thus we adjust the link between the AP and N1 to be ‘X’ = 1.8Mbps. This is just below the average total UL throughput of 2.0346Mbps recorded within the WLAN in the previous experiment.

The saturation percentages recorded under ground truth and for Approach (A), (B) and (C) are shown in Figure 7.5.3. In this test case scenario we begin to see some clear distinctive results amongst all four configurations.

Ground truth has a large number of saturation percentages for both AC1 and AC0 below the 100% mark, so that saturation load within the UL is clearly not being maintained. We specified the outside link bandwidth between the AP and N1 to be 1.8Mbps, and the resulting average throughput for the UL within the WLAN was 1.838Mbps. This value is approximately 90% of the within WLAN UL throughput achieved in the fully saturated test case scenario 2.0346Mbps, and we can see a major clustering of saturation percentage values just below and above this mark. As expected, ground truth demonstrates a distinctively different saturation percentage record from the saturated load test case scenario.

In Approach (A), we also obtain a different saturation percentage spread as compared to the fully saturated load example for Approach (A) shown earlier, but there are some major shortcomings in Approach (A). Comparing visually the results from the ground truth with Approach (A), we observe Approach (A) does not have the distinct grouping of a large number of percentage records below the 100% mark. There are a few present, however the majority still tend to lie at 100% mark. This is a result of the non-classification and discarding of idle time records involving collisions, this reduces the available information from which saturation percentage values are calculated.

Approach (B), also produces a different recorded saturation percentage spread as compared to its fully saturated load results. However, Approach (B) still maintains a strong clustering of saturation percentages for AC1 and AC0 on the 100% mark. Only a few clear saturation percentages can be seen that do not report a value of 100%, this is due to the unrealistic reporting of collisions involving all wireless client stations, which inherently introduces a strong bias towards repeatedly classifying idle time periods as being saturated.

In the case of Approach (C), the first impression is that there is no real difference between its fully saturated load test case scenario results. However, upon further inspection we can see subtle differences that highlight the benefits of this approach as opposed to Approach (A) and (B). The first key thing is visually the saturation percentages recorded resemble those of the ground truth. Aside from this the actual saturation percentages are
slightly more wide spread as compared to the fully saturated test case scenario. There is still the majority grouping of saturation percentages clustered within the square defined by the percentage marks 75% to 100%. However, we do not observe as many values situated on the 100% mark and generally the clustering appears uniform within the defined square region. Therefore we can actually say there is a detectable difference between fully saturated test case scenario and the near saturation test case scenario.

Therefore, we can begin to now see a clear benefit for Approach (C) being a viable choice for a feasible UL detection scheme compared to Approach (A) and (B). The ability to distinguish the difference between saturation and non-saturation traffic load scenarios is critical. From experience with the DL, being able to distinguish the difference was vital in developing a control scheme capable of achieving acceptable results in all conditions. Thus we expect when trying to achieve the same goals with UL this will also prove to be a critical factor.
Figure 7.5.3: UL Saturation Percentage - Ground Truth, Approach (A), Approach (B) and Approach (C)
3. High degree of non-saturation load for both UL AC1 and UL AC0

In this test case scenario we adjust the link between the AP and NI to be ‘X’ = 0.6Mbps. This is well below the average total UL throughput recorded for the previous two test case scenarios.

This test case scenario begins to truly capture the real benefit for selecting Approach (C) as a viable UL detection scheme. The immediate visual observation of results of the recorded saturation percentages show that Approach (C) is the closest representation to that of ground truth. As we can see ground truth sees no saturation percentage at the 100% mark. While both Approach (A) and (B) still continue to show numerous records of 100% saturation being achieved by UL AC1 and UL AC0. In the case of Approach (C), this is clearly not the true as, like ground truth, we obtain a consistent grouping of saturation percentages in the lower left quadrant. In Approach (A) and Approach (B), they have a higher reporting of saturation percentages, leading to a very serious of lack of reporting the very situation which any control scheme needs to be made aware of and address appropriately.

Therefore with this particular test case scenario, we can now conclusively see the real benefit of selecting Approach (C) in the feasible UL detection scheme. Approach (C) achieves the vital goal of being able to distinguish between saturation and non-saturation load conditions within UL. As mentioned earlier, for a control scheme this is absolutely paramount, since dealing with saturation conditions is not an issue, the issue that does arise is knowing exactly when to behave differently in the advent of non-saturation load conditions.
Figure 7.5.4: UL Saturation Percentage - Ground Truth, Approach (A), Approach (B) and Approach (C)
CHAPTER 7. UPLINK TRAFFIC SATURATION DETECTION

7.6 Summary

In this chapter, we have discussed the need for a UL detection scheme that is capable of reporting non-saturation traffic load presence within the UL for AC1 and AC0 respectively. We defined a specific set of requirements that a suitable UL detection scheme should exhibit and then highlighted possible solutions that are present in the literature. However, as we discovered, none of the approaches detailed proved to be adequate in meeting the set of requirements, thus we then designed and developed our own UL detection scheme.

As part of our UL detection scheme, we proposed three possible implementation approaches that dealt with physical collisions differently, namely Approach (A), (B) or (C).

(A) Any idle time period that records a physical collision will be discarded and not be classified.

(B) Physical collisions result in the doubling of the contention window value each time from the initial $CW_{min}$ value.

(C) Physical collisions do not result in the doubling of the initial contention window value, $CW_{min}$.

We utilized three test case scenarios using fixed CW parameters to determine which of the three approaches was most suitable in the final UL detection scheme.

1. Fully saturated load for both UL AC1 and UL AC0.
2. Near saturation load for both UL AC1 and UL AC0.
3. High degree of non-saturation load for both UL AC1 and UL AC0.

For comparison purposes we determined the correct saturation of status for each transmission in order to gauge the suitability of each approach, and selected Approach (C) as the only suitable choice to implement within the final UL detection scheme. This decision was made based on the ability to distinguish non-saturation being clearly the most important feature that needs to be detected.

With the UL detection scheme finalized, we now begin in Chapter 8 the process of integrating the detection scheme into the complete DL and UL control scheme. As in the case of the DL, we isolate our analysis in the next chapter just to observing the effectiveness of the final control scheme in dealing with non-saturation UL traffic load. This is demonstrated via a suite of simulations that are the symmetric equivalent as those used for the DL in Chapter 6.