2005-01-01

An Experimental Analysis of the Call Capacity of IEEE 802.11b Wireless Local Area Networks for VoIP Telephony

Brian Keegan
Technological University Dublin, brian.x.keegan@tudublin.ie

Funder: Enterprise Ireland

An Experimental Analysis of the Call Capacity of IEEE 802.11b Wireless Local Area Networks for VoIP Telephony

by

Brian Keegan
B.Eng.

A thesis submitted to the Dublin Institute of Technology
for the degree of

Master of Philosophy

Supervisor: Dr. Mark Davis

School of Electronic and Communications Engineering

August 2005

Abstract

The use of the Internet to make phone calls is growing in popularity as the Voice over Internet protocol (VoIP) allows users to make phone calls virtually free of charge. The increased uptake of broadband services by domestic users will further increase the use of VoIP telephony. Furthermore, the emergence of low cost wireless networks (namely IEEE 802.11a/b/g WLANs) is expected to bring wireless VoIP into the mainstream.

As the number of wireless hotspots increases more users will want to use VoIP calls wherever possible by connecting to open access points (AP). A major concern with VoIP is Quality of Service (QoS). In order for VoIP to be truly successful users must enjoy a similar perceived QoS as a call made over a traditional telephone network. There are many factors that influence QoS which include: throughput, packet delay, delay variation (or jitter), and packet loss.

This thesis is an experimental study of the call capacity of an IEEE 802.11b network when using VoIP telephony. Experiments included increasing the number of VoIP stations and also increasing the level of background traffic until network saturation occurs. Results show that the network is capable of supporting at least 16 VoIP stations. Due to the operation of the IEEE 802.11 medium access control (MAC) mechanism, the AP acts as a bottleneck for all traffic destined for wireless stations, in that significant delays can be incurred by VoIP packets which can lead to a poor perceived QoS by users. Consequently the performance of the AP downlink is the critical component in determining VoIP call capacity.

Declaration

I certify that this thesis which I now submit for examination for the award of ____________________, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for an award in any other Institute or University.

The work reported on in this thesis conforms to the principles and requirements of the Institute's guidelines for ethics in research.

The Institute has permission to keep, to lend or to copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature_________________________________Date_______________

Acknowledgements

I would first like to thank my supervisor, Dr. Mark Davis for his invaluable advice and guidance during the course of my research. He has been extremely supportive and motivating throughout out the entire project particularly during difficult periods.

I would like to acknowledge my gratitude to the late Professor John T. Lewis who as Principal Investigator at the Communications Networks Research Institute in the Dublin Institute of Technology provided me with the opportunity to carry out my research in a stimulating and refreshing work environment.

I would like to thank all my colleagues at CNRI, past and present for their stimulating and often humorous conversations which provided me with a pleasant working atmosphere. In particular I owe a debt of gratitude to David McGrath for giving up his valuable time and experience.

I would also like to express my thanks to Enterprise Ireland for providing financial assistance under the Informatics Initiative programme.

Finally, I would like to thank my friends and family who have all contributed to this project in some form. This project would not have been possible without their support and understanding.

List of Figures

List of Tables

Abbreviations and Acronyms.

ACK	Acknowledgment
AP	Access Point
ATM	Asynchronous Transfer Mode
BC	Backoff Counter
BPSK	Binary Phase Shift Keying
BSS	Basic Service Set
BW_{Access}	Access Bandwidth
BW_{Busy}	Busy Bandwidth
BW_{Free}	Free Bandwidth
BW_{Idle}	Idle Bandwidth
BW_{Load}	Load Bandwidth
CCK	Complementary Code Keying
CFP	Contention-Free Period
CP	Contention Period
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CTS	Clear To Send
CW	Contention Window
DA	Destination Address
DCF	Distributed Coordination Function
DIFS	DCF Inter Frame Space
DS	Distribution System
DSSS	Direct-Sequence Spread Spectrum

EB	Effective Bandwidth
ESS	Extended Service Set
ETSI	European Telecommunications Standards Institute
FHSS	Frequency Hopping Spread Spectrum
FIFO	First In First Out
GOB	Good Or Better
IBSS	Independent Basic Service Set
IEEE	Institute of Electrical and Electronic Engineers
IFS	Inter Frame Space
IP	Internet Protocol
ISM	Industrial, Scientific and Medical
ISO	International Standards Organization
ITU	International Telecommunications Union
LLC	Logical Link Control
MAC	Medium Access Control
MSDU	MAC Service Data Unit
NAV	Network Allocation Vector
NIC	Network Interface Card
OFDM	Orthogonal Frequency-Division Multiplexing
PAMS	Perceptual Analysis/Measurement System
PCF	Point Coordinator Function
PESQ	Perceptual Evaluation of Speech Quality
PHY	Physical Layer
PIFS	PCF Inter Frame Space
PLR	Packet Loss Rate/Ratio

POW	Poor Or Worse
PPS	Packet Per Second
PSTN	Public Switched Telephone Network
PSQM	Perceptual Speech Quality Measurement
QPSK	Quadrature Phase Shift Keying
QoS	Quality of Service
RF	Radio Frequency
RTP	Real Time Protocol
RTS	Request To Send
SA	Source Address
SIFS	Short Inter Frame Space
SIP	Session Initiation Protocol
STA	Station
TCP	Transmission Control Protocol
TSB	Telecommunications System Bulletin
UDP	User Datagram Protocol
VoIP	Voice over Internet Protocol
WEP	Wired Equivalent Privacy
WLAN	Wireless Local Area Network
Wi-Fi	Wireless Fidelity

1 Introduction

According to [1] global cellular subscriber growth rate will increase by 7 percent per annum over the next 5 years. As part of this growth a significant shift is predicted from existing wireless technologies to evolutionary and revolutionary technologies. This is due mainly to the success of mobile communications and will include a greater utilisation of Wireless Local Area Networks (WLANs) amongst others. WLANs have a number of advantages over conventional wired LANs with the most obvious advantage being mobility. Other advantages over LANs include ease of installation, flexibility, and lower costs. The IEEE 802.11b [2] is one of the most widely deployed wireless technologies available with IEEE 802.11a [3] and IEEE 802.11g [4] gaining ground. Bandwidth availability is limited due to physical layer restrictions. At present, line rates of 11Mbps for 802.11b and 54Mbps for 802.11a/g are available. However, overheads incurred due to the operation of the 802.11 medium access control (MAC) mechanism reduce the throughput.

1.1 Problem Statement

The 802.11 distributed co-ordination function (DCF) was designed to deliver best effort traffic for non-real time applications and is not well suited to delivering real-time applications such as VoIP. This is due to the fact that real-time applications require a specific amount of bandwidth in order to provide Quality of Service (QoS) guarantees. If a voice packet is dropped or delayed the end user will experience a reduction in perceived call quality. The operation of the DCF is such that all stations contending for access to the wireless medium do so with the same priority. This means that there is no differentiation between real-time and non-real time services and as such all traffic types will experience the same QoS. Furthermore, the access point (AP) undergoes the same contention process as all the stations connected to the medium. During multiple VoIP sessions the downlink from the AP will essentially contain the aggregate of all the station uplink loads. This is a well known problem which results in basic traffic asymmetry in the AP. In order for the AP to avoid delaying packets excessively or dropping packets it must gain sufficient access to the medium. This will effectively determine the upper limit on the network capacity and can be thought of as a bottleneck in the network. Likewise, as the number of VoIP users increases so too does the contention for transmission opportunities which results in larger delays being experienced by the VoIP packets. Another limiting factor on the network capacity will be due to protocol overhead. Every time a packet is transmitted, protocol headers are pre-pended which increase the wireless frame size. This may result in an overhead larger than the packet being transmitted resulting in an inefficient use of the medium. One solution would be to increase the size of the packet, however this would result in a greater loss of data if a frame is dropped or lost on the wireless medium.

These limiting factors have led to much research through both simulation and experiment into determining the capacity and QoS experienced by wireless networks employing VoIP [5]-[9]. Depending on the choice of audio codec and whether or not background traffic is present, results have shown that 802.11b networks can support approximately 5 VoIP users (using the G.711 codec) compared to 802.11a/g which can support approximately 40 VoIP users (and also using low bit rate codecs).

1.2 Objectives and Contributions

The majority of research publications in this area uses QoS performance metrics based upon end-to-end delay, throughput, and percentage packet loss during a VoIP call. The end-to-end delay is also known as the mouth-to-ear delay and is essentially the time taken from when the user on one end speaks to when the user on the other end hears them. There are a number of delays a packet may encounter, such as encoder delay, transport delay, and propagation delay. The emphasis in this thesis is to measure the packet transit time through the AP and in doing so to study the effects of the AP buffer dynamics. This is achieved by tracking a VoIP packet as it propagates through the AP. In order to do this it is necessary to first capture a packet on the uplink and then again on the downlink. By carefully examining the packet header information it is possible to identify individual packets as they transit the AP.

A tool known as a WLAN *Probe* is used to capture packets and determine their origin. Another feature of the WLAN Probe is to calculate the network utilisation on a per-station basis using a descriptive framework based upon the concept of MAC bandwidth components. The WLAN Probe calculates the amount of bandwidth that each contending station requires in order to gain access to the medium as well as to transmit its load. The remaining bandwidth available to each station is then deemed to be spare or “free” bandwidth which can be thought of as a reservoir of spare capacity that a station can draw upon should it require to do so. By using the WLAN Probe we aim to measure the MAC bandwidth components up to the point of network saturation.

To effectively study the dynamics of the AP buffer occupancy it is necessary to analyse the behaviour of two particular aspects of the queuing mechanism, namely the arrival rate and the service rate. For stability it is required that on average the service rate be greater than the arrival rate. In the short term, if the arrival rate exceeds the service rate the buffer should be capable of temporarily storing packets until they can be serviced. By introducing controlled amounts of background traffic it is possible to stress different aspects of the buffer operation at the AP. It is then possible to measure the transit delay through the AP under these different load conditions. Analysis of the AP delay is performed using a combination of probability density functions and cumulative density functions. Furthermore, by using a statistical method known as a *Watermark plot* (which is essentially the log of the inverse cumulative density function) it is possible to study the tails of the probability of delay distribution. The Watermark plot method allows us to determine if the AP suffers from long delays that may result in buffer overflows.

The results presented in Chapter 5 indicate that an 802.11b wireless LAN can support at least 16 VoIP callers when there is no background traffic present. As the number of stations accessing the medium was increased from 2 to 16 the contention for access to the medium also increased. As the number of stations contending increases so too does the network delay (due to the operation of the 802.11 MAC). This could be seen by an increased AP transit time. On average all stations displayed an increased probability in AP transit delay, however 99% of the delay times were found to be below 10 ms. Analysis shows that large peak loads in the aggregate are caused during periods of station double-talk (i.e. when there is a brief overlap in conversation between both talkers) resulting in an increase in the mean load.

Controlled background traffic enabled us to saturate the network using different packet sizes. By transmitting background traffic from a wireless source to a wired sink the contention for access to the medium was increased. Although VoIP stations would continue to transmit up to the point of network saturation it could be seen that the AP transit times had increased significantly before saturation had been achieved. When tests were conducted with background traffic transmitting from a wired source to a wireless sink the AP buffer contained both VoIP packets as well as traffic generator packets. Although a higher throughput could be achieved during these tests, results also showed that stations experienced longer delays and an increase in packet loss. The findings indicate that the operation of the AP limits the capacity of the network as well as adding to the network delay. In particular the performance of the AP downlink will ultimately dictate the capacity of the network.

1.3 Organisation

This thesis is organised as follows.

Chapter 2 describes the IEEE 802.11b in detail with particular emphasis on the MAC header information. This chapter also describes various technologies and protocols used during VoIP on a WLAN. The chapter finishes with a brief discussion on buffer dynamics and analysis.

Chapter 3 gives details of the characteristics of speech quality as well as quantification and measurement of speech quality. This chapter introduces the concept of MAC bandwidth components and describes each of them in detail. An example of a simple network capacity calculation is provided before giving an overview of relevant literature concerning the research.

Chapter 4 describes the various test bed scenarios before detailing the method of packet capture and processing involved before analysis.

Chapter 5 presents the results in 3 main sections. Each section is based on a different experimental set-up and follows a logical course of analysis. The first section is the results for increasing the number of VoIP stations whilst the second and third sections involve the use of background traffic.

Chapter 6 presents a summary of the main findings and conclusions from the work carried out. It also suggests areas of further research.

2 Technical Background

2.1 Wireless Local Area Network (WLAN)

Wireless networks come in many different varieties such as, Bluetooth, IEEE 802.15 and Cellular networks to name but a few. Quite simply a wireless network uses radio waves to connect devices together in order to communicate. The emphasis of this project will be on IEEE 802.11 Wireless LANs [10] in particular IEEE 802.11b [2]. The IEEE 802.11 Working Group (WG) is a member of the IEEE 802 LAN/MAN Standards Committee and as such shares much of the functionality with other IEEE 802 components [11]. The data link layer consists of two sub layers: the logical link control (LLC) and medium access control (MAC). Simple bridging from wireless to wired networks is achieved by employing the same 48-bit addressing as other 802 LANs, namely the 802.2 LLC. The difference between 802.11 and other 802 standards is at the MAC and Physical Layer. Figure 2.1 illustrates the relationship of 802.11 to the generally accepted TCP/IP reference model (as opposed to the OSI 7-layer model [12]). The original 802.11 standard published in June 1997 defined a 2.4 GHz system with a maximum data rate of 2 Mbps. Currently there are two basic categories of IEEE 802.11 WLAN standards [13]. First there are those that specify the fundamental protocols for the complete WLAN system. These are called 802.11a, 802.11b, and 802.11g. Second, there are extensions that address weaknesses or provide additional functionality to these standards. Amongst others, some examples include 802.11e which addresses QoS enhancements for 802.11 networks [14], 802.11f which addresses the issue of roaming between APs from different vendors [15], and 802.11i which addresses MAC enhancements for enhanced security [16]. The IEEE 802.11 standard originally specified two spread spectrum radio layers, frequency hopping (FHSS) and direct sequence (DSSS) as well as a lesser deployed infrared (IR) layer. However, the IEEE 802.11 working group later added a third radio technique called orthogonal frequency division multiplexing (OFDM). Table 2.1 compares the three basic 802.11 standards.

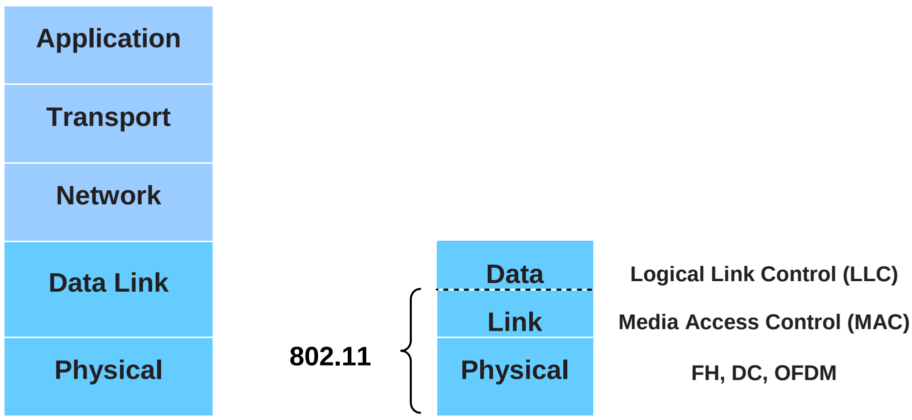


Figure 2.1: The communications protocol stack and 802.11.

Standard	Radio	Modulation	Max. Link Coverage	Max. Data Rate	Max. # Non-Overlapping Channels
802.11b	2.4 GHz	DSSS	100m	11 Mbps	3
802.11a	5 GHz	OFDM	50m	54 Mbps	12
802.11g	2.4 GHz	OFDM	100m	54 Mbps	3

Table 2.1: Comparison of 802.11 standards [17].

2.1.1 802.11 Architecture

The 802.11 standard specifies four major physical components. A Wireless Station (STA) which is typically a computing device equipped with a wireless network interface card (NIC). This would normally be a portable handheld computer but there is no reason why it should be limited to this. A Wireless

Medium which is used to transport data frames. How this is used depends on which radio frequency (RF) physical layer is used. An Access Point (AP) which performs wireless-to-wired bridging in order to connect to other networks. However, it may simply act as a relay in order to extend the range of a wireless network. A Distribution System (DS) which can be thought of as a system used to interconnect a set of basic service sets (BSSs) and integrated local area networks (LANs) to create an extended service set (ESS).

The BSS is the fundamental building block of an 802.11 LAN. This consists of two or more STAs equipped with wireless network interface cards in order to communicate with each other. When a BSS exists without an AP it is also referred to as an independent BSS (IBSS) or *ad-hoc* network. Another mode of operation is the infrastructure mode and is distinguished by the use of an access point. The AP acts as a relay in order to extend the range of the BSS and/or to connect to a LAN. These networks are referred to as infrastructure BSS. When two or more BSSs are linked together using a backbone network this is known as an extended service set (ESS). Figures 2.2 (a) and (b) illustrates the different network configurations.

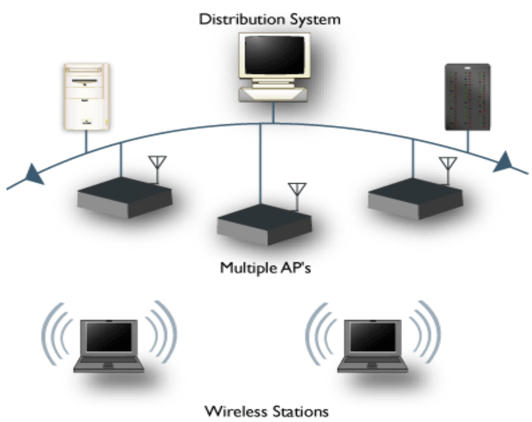


Figure 2.2(a): Extended Service Set.

Figure 2.2 (b): Basic Service Set (*ad-hoc*).

2.2 The Protocol Stack

Figure 2.1 shows the various layers of the communications protocol stack using the TCP/IP reference model. The network layer (layer 3) is known as Internet Protocol (IP). Two transport layer protocols (layer 4) are used, known as transmission control protocol (TCP) and user datagram protocol (UDP). The integration of these protocols on the communications protocol stack can be seen in Figure 2.3.

2.3 Packet Switched Networks & Multimedia Communications

Packet switched networks as opposed to circuit switched networks refer to a communications network in which data is broken down into smaller packets before being transmitted. Each data packet is pre-appended with a header that details its source and destination address. The main difference between datagram packet switched and circuit switched network is that once data is transmitted on a packet switched network it may take any number of available paths (which will impact on the transit delay and packet loss). In contrast a circuit switched network uses a dedicated fixed path from the transmitter to the receiver which is set up before communication takes place. For multimedia communications this means breaking up different media types such as text, images, speech, audio, and video into packets before transmission. Text and images are normally stored in digital format and would not be sensitive to timing. However, audio and video would normally be generated from an analogue source and would therefore require digitising. This can result in very large files being generated. As a consequence some form of compression algorithm is applied to these media types in order to reduce the volume of the file. Another problem arises when it is required to use audio and video for real-time applications. Real-time applications are very intolerant to delays incurred during transmission and as a result these effects must be minimised if the users are to experience an acceptable perceived quality of service.

For the effective transport of multimedia communications it is usual to implement a set of communications protocols. This ensures that information streams are interpreted in the same way from the transmitter to the receiver even if they are using different video conferencing software. Although there are alternative protocols (such as SIP [18]), during the course of this project we will be using the protocol defined by ITU-T Recommendation H.323 [19].

2.4 ITU-T Recommendation H.323

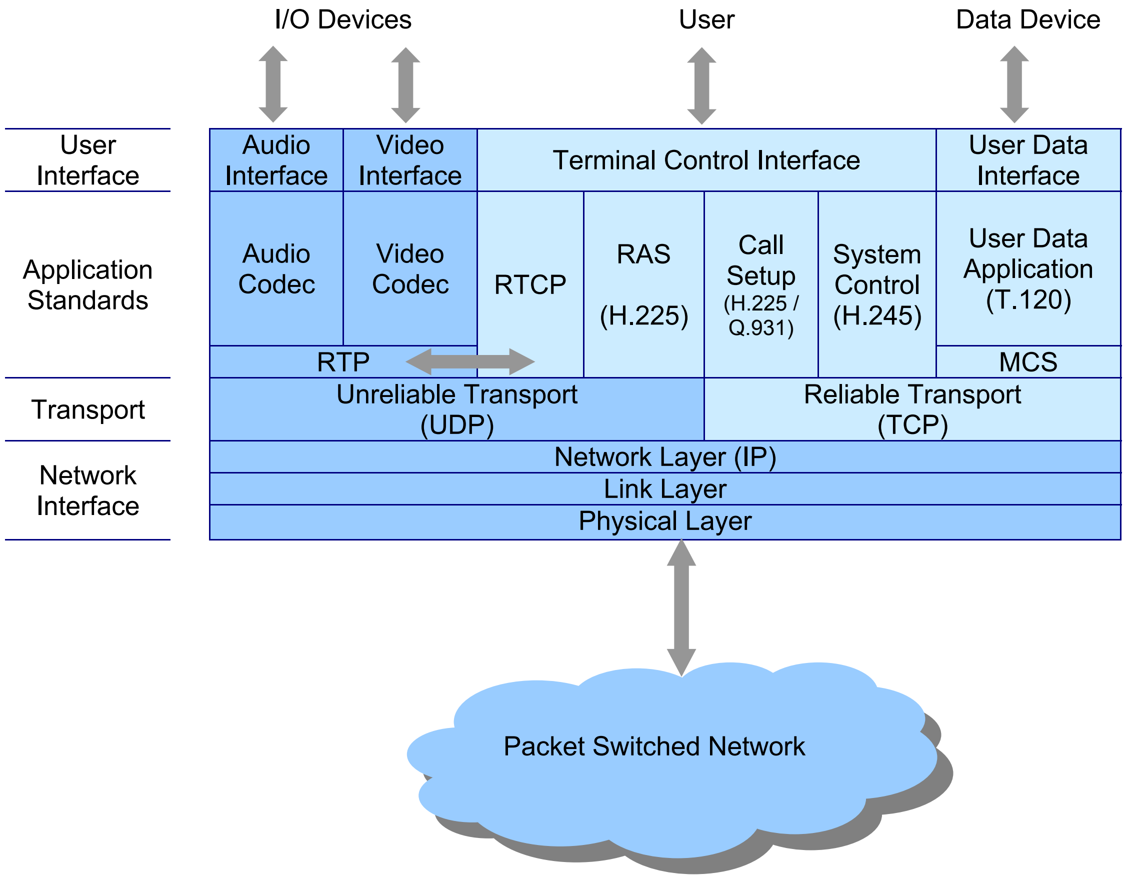


Figure 2.3: Structure of the H.323 recommendation in the protocol stack.

ITU-T Recommendation H.323 is a standard for packet-based multimedia communications. H.323 is intended for use with networks and communications systems that do not provide a guaranteed quality of service. Furthermore, the H.323 standard specifies the components, protocols, and procedures providing multimedia communication over packet-based networks that are used to build up such a communications system. It can be considered as an “umbrella standard” which aggregates standards for multimedia conferencing over packet-based networks.

Figure 2.3 shows the integration of the various H.323 application standards. The standard comprises components for the packetisation and synchronization of audio and video streams, a registration, admission and status (RAS) control procedure, multipoint conference control, as well as terminal control for interfacing with different types of circuit switched networks. The standard is independent of the underlying transport and network layers and can therefore be implemented with any type of network (i.e. LAN, WLAN, and Intranet). The figure shows the structure for the majority of LANs where the network layer protocol is the Internet Protocol (IP). The H.323 standard assumes that the transport layer provides both an unreliable service (provided in this case by the User Datagram Protocol, UDP) and a reliable service (in this case the Transmission Control Protocol, TCP).

2.4.1 H.323 Call Control and Signalling

A typical call is initiated as shown in Figure 2.4. STA 1 connects to STA 2 by sending a H.225 set-up message. This also includes the type of call to be established (e.g. audio only, audio/video) as well as caller and callee number and address. STA 2 alerts STA 1 of the connection establishment by sending a H.225 alerting message. STA 1 must receive this message before its set-up time expires. The call is established when STA 2 confirms by sending a H.225 connect message to STA 1. The H.245 control channel is now established between the two stations.

STA 1 sends a H.245 TerminalCapabilitySet message to STA 1 to exchange its capabilities (media type, codec choices, etc). STA 2 then replies with a TerminalCapabilitySetAck to acknowledge its capabilities. A similar procedure takes place in the opposite direction. A media channel is then opened between STA 1 and STA 2. This is achieved by sending a H.245 OpenLogicalChannel message in order to set-up the voice channels over which the media will be streamed. This will include the transport address for the RTCP channel. STA 2 acknowledges the establishment of the channel by sending a H.245 OpenLogicalChannelAck message to STA 1. Included in the acknowledgement message is the port number to which STA 1 should send the RTP data as well as the port number for the RTCP data as previously specified by STA 1. Again a similar procedure takes place in the opposite direction so that a bi-directional media communication stream is established. At this point the stations exchange encapsulated RTP media streams. RTCP packets are sent during this exchange to monitor data transmission quality.

When a station hangs up, for instance STA 1, it sends a H.245 CloseLogicalChannel message to the connected station, STA 2. STA 2 replies with H.245 CloseLogicalChannelAck message. When this is received, STA 1 sends a H.245 EndSessionCommand. When STA 1 receives the same command from STA 2 both stations send a H.225 ReleaseComplete message which closes the channel and ends the call.

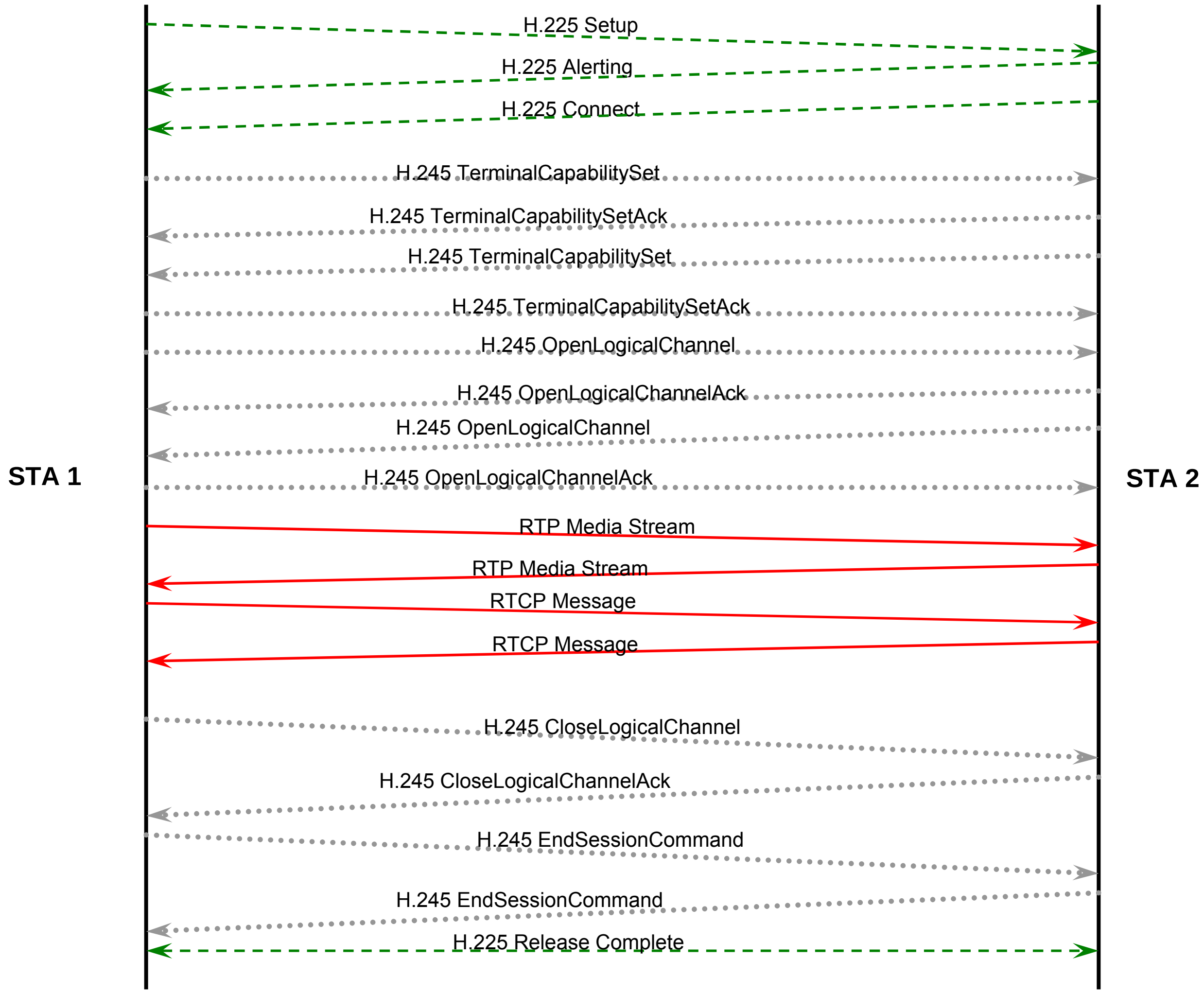


Figure 2.4: H.323 Call control and signaling.

2.4.2 Real-time Transport Protocol

Real-time transport protocol (RTP) provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video [20]. Services include payload type identification, sequence numbering, timestamping, and delivery monitoring. Applications typically run RTP on top of UDP to make use of its multiplexing and checksum services. RTP does not in itself guarantee real-time delivery of multimedia data (since this is dependent on network characteristics) nor does it assume the underlying network is reliable in delivering the packets in sequence; it does however provide the mechanism to manage the data as it arrives to best effect. RTP is augmented by a control protocol (RTCP) to monitor data delivery and network statistics. Together they resolve many of the problems a UDP network environment may experience, such as lost packets, jitter, and out of sequence packets. RTCP operates alongside RTP and shares information with it. However, each RTCP has a different port number associated with it so that it can operate independently of RTP. Figure 2.5 shows how the RTP header is incorporated into a Voice over IP (VoIP) packet. Along with the voice data is the UDP and IP header information.

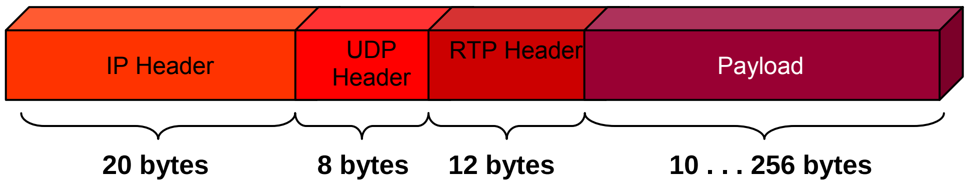


Figure 2.5: RTP Header used for VoIP packet.

The RTP packet structure can be seen in Figure 2.6. The first twelve octets are present in every RTP packet, while the list of CSRC identifiers is present only when inserted by a mixer. The fields have the following meaning;

- *Version (V)*: This is a 2-bit field that indicates the version of RTP being used.
- *Padding (P) & Extension (X)*: These are both 1-bit fields to allow extensions to the basic header to be defined and added in the future.

- *CSRC Count (CC)*: The CSRC count is a 4-bit field that contains the number of CSRC identifiers that follow the fixed header. This is used in a multicast session.
- *Marker (M):* This is a 1-bit field that allows significant events such as frame boundaries to be marked in the packet stream.
- *Payload Type*: This is a 4-bit field that indicates the type of audio or video codec being used. Moreover, since each packet contains this field it can be changed to allow for a change in the network quality of service.
- *Sequence Number*: This is a 16-bit field that increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence by buffering a number of packets before playout of the data they contain starts. Later we will show how this number can be used to track a packet through the network.
- *Time Stamp*: This 32-bit field indicates the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations. This information is sent via the RTCP to the sending RTP which may modify the resolution of the compression algorithm being used depending on the network performance.
- *Synchronization Source (SSRC) Identifier*: This is a 32-bit field that identifies the source device that has produced the packet contents. The data generated from a video conferencing call may come from different

sources (i.e. audio and video) and so the identifiers should be random so that no two sources should have the same identifier.

- *Contributing Source (CSRC) Identifier*: This is a 32-bit field that is used during multicast sessions and is typically the IP address of the source.

<table>
<tr><td>1 2</td><td>3</td><td>4</td><td>5 6 7 8</td><td>9</td><td>10 16</td><td>17 32</td></tr>
<tr><td>V</td><td>P</td><td>X</td><td>CC</td><td>M</td><td>Payload Type</td><td>Sequence Number</td></tr>
<tr><td colspan="7">Time-Stamp</td></tr>
<tr><td colspan="7">Synchronization Source (SSRC) Identifier</td></tr>
<tr><td colspan="7">Contributing Source (CSRC) Identifiers</td></tr>
</table>

Figure 2.6: Real-time transport protocol packet format.

2.4.3 Audio and Video Codecs

Codecs are used to convert an analogue voice or video signal to a digitally encoded version. Codecs vary widely in the quality, the bandwidth required for their operation, and their computational requirements. They work by using an encoder that takes an incoming analogue stream, compressing the signal by deleting extraneous or duplicate information based on a certain compression algorithm, digitising the stream and then transmitting it. At the receiving end, a decoder decompresses the signal before playing it back. A “waveform codec” preserves the incoming signal’s waveform characteristics and operates by sampling the signal. Examples of audio waveform codecs are the A-Law, μ-Law (collectively known as G.711 [21]), G.721 [22], and G.726 [23]. Low bit rate codecs (such as G.723.1 [24], G.729A [25], and G.729 [26]) work differently to these by using a model of human speech to encode and compress speech signals. These are often referred to as “vocoders” and instead of transmitting the signal, they simply send the speech model parameters hence reducing the transmitted bit rate. ITU-T Recommendation G.113 [27] provides a comprehensive list of some of the most popular waveform and low bit rate codecs. The list also includes figures for Equipment Impairment Factors, Ie (i.e. impairments introduced by speech compression and packet loss). This list can be seen in Table 2.2 where it can be shown that G.711 has the lowest Ie value, for this and other similar reasons, G.711 is often used as a reference [28].

The main difference between all codecs is how much or how fast they can compress the stream. Tight compression results in a smaller amount of bandwidth being used for transmission or storage, but may require more processing power than is available in a given amount of time. This results in the inability to compress the signal in real-time. Faster compressors may take shortcuts to get to the end result, leaving a signal that does not retain all the original quality or isn't compressed as tightly. The trade off is to find the balance between small files and real-time compression that gives the best quality.

Codec Type	Reference	Operating Rate kbit/s	Ie Value
Waveform Codecs			
PCM	G.711	64	0
ADPCM	G.726, G.727	40	2
	G.721, G.726, G.727	32	7
	G.726, G.727	24	25
	G.726, G.727	16	50
Speech Compression Codecs			
LD-CELP	G.728	16	7
		12.8	20
CS-ACELP	G.729	8	10
	G.729-A + VAD	8	11
VSELP	IS-54	8	20
ACELP	IS-641	7.4	10
QCELP	IS-96a	8	21
RCELP	IS-127	8	6
VSELP	Japanese PDC	6.7	24
RPE-LTP	GSM 06.10, Full-rate	13	20
VSELP	GSM 06.20, Half-rate	5.6	23
ACELP	GSM 06.60, EFR	12.2	5
ACELP	G.723.1	5.3	19
MP-MLQ	G.723.1	6.3	15

Table 2.2: Comparison of codecs [27].

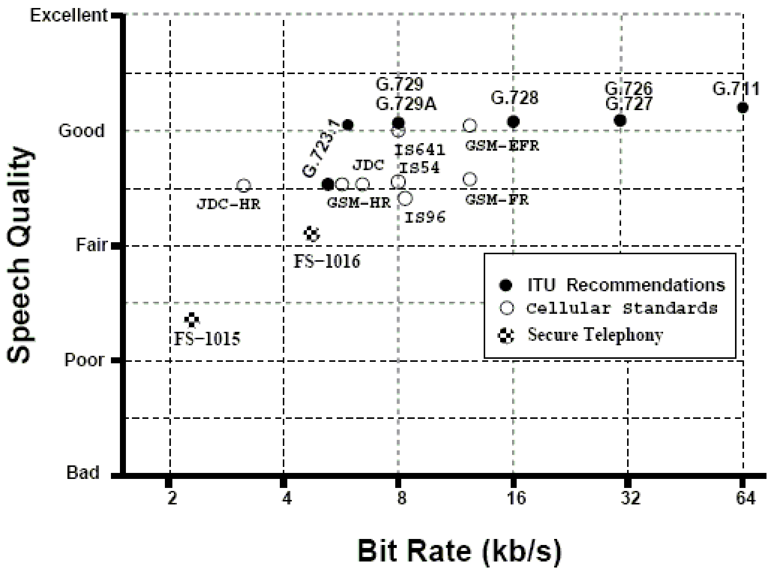


Figure 2.7: Comparison of speech codecs for multimedia communications [29].

2.5 Voice over IP (VoIP)

Traditionally telephone calls are carried through the Public Switched Telephone Network (PSTN) which provides high-quality voice transmission between two or more parties. On the other hand, data such as email, web browsing etc. are carried over packet-based data networks like Internet Protocol (IP) or Asynchronous Transfer Mode (ATM). In recent years, there has been an increased move towards using data networks to carry both the telephone calls and the data. The convergence of voice and data networks is very appealing to many service providers, mainly for two reasons; lower costs and increased functionality. VoIP systems encode, compress, and transmit analogue voice signals as a stream of packets over a digital data network. VoIP technology ensures proper reconstruction of voice signals, compensating for echoes due to the end-to-end delay, jitter, dropped packets, and signalling required for making telephone calls.

The IP network used to support IP telephony can be a standard LAN, a network of leased facilities or indeed the Internet. VoIP calls can be made or received using standard analogue, digital and IP phones. VoIP gateways serve as a bridge between the PSTN and the IP network. A call can be placed over the local PSTN network to the nearest gateway server which moves it onto the Internet for transport to a gateway at the receiving end. With the use of VoIP gateways, computer-to-telephone calls, telephone-to-computer calls and telephone-to-telephone calls can be made with ease.

To ensure interoperability between different VoIP manufacturers, VoIP equipment must follow agreed procedures for setting up and controlling the telephone calls. ITU-T H.323 is one such family of standards that define various options for voice (and video) compression and call control for VoIP.

2.6 Quality of Service

A primary concern of IP telephony has been ensuring Quality of Service (QoS), i.e. making sure that a VoIP call has a similar perceived QoS by the user as a call made over a traditional telephone network. Quality of Service when referred to a communications channel or the rating of telephone communications quality is a performance assessment in which listeners judge transmissions by qualifiers, such as excellent, good, fair, poor, or unsatisfactory. Examples of qualifiers and scales are dealt with in more detail in Chapter 3. QoS is often used as a general umbrella term that incorporates bandwidth, delay, delay variation (jitter) and information loss to describe a network's ability to customise the treatment of specific classes of data. For example, traffic engineering techniques can be used to prioritise video transmissions over Web-browsing traffic. Advanced networks can offer greater control over how data traffic is classified and offer greater flexibility as to how the treatment of that traffic is differentiated from other traffic. The key parameters impacting a user are outlined below;

- *Throughput*: This is a measure of the amount of data that will be sent across a network per unit time. This is usually measured in bits per second (bps) or packets per second (pps). Limitations on throughput are linked to the available bandwidth.
- *Delay*: Delay manifests itself in a number of ways, including the time taken to establish a particular service from the initial user request and

the time to receive specific information once the service is established. This may include (amongst others) coder delay, packetisation delay and queuing delay.

- *Delay Variation*: Also known as "*Jitter*", delay variation becomes an important performance parameter in packet switched networks as the arrival times become inherently variable as packets may take different paths as they propagate through a network. These effects may be counteracted by buffering packets on reception, with the added expense of an increased end-to-end delay.
- *Information Loss*: This is normally the ratio of information received to the amount of information transmitted. This is usually defined as the Packet Loss Ratio (PLR). However, it can also include the effects of any degradation introduced by media coding.

The International Telecommunication Union (ITU) gives guidelines on performance targets for audio and video applications [30]. Table 2.3 below outlines some of the main performance metrics of concern.

Application	**Degree of symmetry**	**Typical data rates**	**Key performance parameters and target values**		
			One-way delay	**Delay variation**	**Information Loss (Note 2)**
Conversational voice	Two-way	4 – 64 kbit/s	<150 ms preferred (Note 1) <400 ms limit (Note 1)	< 1ms	< 3% Packet loss ratio (PLR)
Videophone	Two-way	16 - 384 kbit/s	< 150 ms preferred (Note 4) <400 ms limit		< 1% PLR
NOTE 1 – Assumes adequate echo control NOTE 2 – Exact values depend on specific codec, but assumes use of packet loss concealment algorithm to minimize effect of packet loss NOTE 4 – These values are to be considered as long-term target values which may not be met by current technology.					

Table 2.3: Performance targets for audio and video applications [30].

2.7 IEEE 802.11 in Detail

In Section 2.1 a brief outline was given of WLAN components and architecture. This section will address in more detail some of the underlying components. The most important aspect from the point of view of this project is the function and structure of the MAC layer and also how the medium access and contention mechanism operate.

2.7.1 IEEE 802.11 and 802.11b Physical Layer (PHY)

The 802.11 LAN system uses spread-spectrum technology, a wideband radio frequency technique. This technology is the foundation for wireless communications in the Industrial, Scientific & Medical (ISM) bands for data use. Traditional radio communications focus on occupying as narrow a band as possible with as much signal as possible. Spread spectrum works by using mathematical functions to diffuse the signal power over a large range of frequencies. When the receiver performs the inverse operation, the smeared out signal is reconstituted as a narrow band signal which makes the data much less susceptible to electrical noise than conventional radio modulation techniques. Spread-spectrum is designed to trade off bandwidth efficiency for immunity to interference, integrity, and security. Spread spectrum modulators use one of two methods to spread the signal over a wider area: frequency hopping spread spectrum (FHSS), or direct sequence spread spectrum (DSSS).

FHSS works very much as the name implies. It takes the data signal and modulates it with a carrier signal that hops from frequency to frequency as a function of time over a wide band of frequencies. On the other hand direct sequence combines a data signal at a sending STA with a higher data rate bit sequence, thus spreading the signal in the whole frequency band.

In this project we consider WLANs based on DSSS technology as given by the IEEE 802.11b standard [2]. The IEEE 802.11 WLAN based on DSSS was initially aimed for the 2.4 GHz band designated for ISM applications as provided by the regulatory bodies worldwide. The DSSS system provides a WLAN with 1 Mbps, 2 Mbps, 5.5 Mbps and 11 Mbps data payload communication capability. This is accomplished by chipping the baseband signal at 11 MHz with an 11-chip pseudo random noise (PRN), code (Barker sequence). The DSSS system uses baseband modulations of differential binary phase shift keying (DBPSK) and differential quadrature phase shift keying (DQPSK) to provide the 1 and 2 Mbps data rates, respectively. Complementary code keying (CCK) is used to provide the 5.5 Mbps and 11 Mbps rates.

2.7.2 IEEE 802.11 MAC Layer

IEEE 802.11 adopts many of the features of Ethernet MAC by adapting the carrier sensing multiple access (CSMA) scheme. However, a collision detection (CSMA/CD) scheme would be impractical on a wireless network, so instead a collision avoidance (CSMA/CA) scheme is implemented. To ensure reliable data delivery and to prevent collisions, the IEEE 802.11 includes a frame exchange protocol. When a station transmits a data frame it must receive an acknowledgement (ACK) from the destination within a certain period of time. If the ACK is not successfully received, either because the original data was damaged or the ACK was damaged then the station re-transmits the data frame. The exchange is treated as atomic. As such 802.11

allows stations to lock out contention during atomic operations, so that other stations attempting to use the transmission medium do not interrupt atomic sequences. A four-frame exchange may also be used to enhance reliability and prevent collisions. In this scheme a source first transmits a Request to Send (RTS) frame which silences any stations that hear it. These stations then refrain from transmitting. The destination station responds with a Clear to Send (CTS) message frame. Like the RTS frame, the CTS also alerts other stations in its vicinity that an exchange is under way so that they may cease transmitting in order to avoid collisions. Once this exchange has completed, frames must be positively acknowledged as before.

Access to the wireless medium can be controlled by one of two coordination functions: a distributed coordination function (DCF) which distributes the decision to transmit over all stations, and a point coordination function (PCF) which involves a centralised decision maker to regulate transmission. The DCF makes use of the basic CSMA algorithm by sensing the medium to see if it is idle. To avoid collisions a random back-off counter is used after each frame is transmitted. The station counter that reaches zero first may then transmit. PCF is built on top of DCF and allows priority to be gained over other stations by initiating a contention free period (CFP) where polling is used to coordinate access to the medium.

2.7.3 Distributed Co-Ordination Function (DCF)

To ensure fair operation, the DCF includes a set of standardised delays that amounts to a priority scheme. These are known as interframe spaces (IFS) and are used in order to coordinate access to the transmission medium. Carrier sensing is used to determine if the medium is available. IEEE 802.11

provides two different types of carrier sensing: Physical carrier sensing, performed at the physical layer; and virtual carrier sensing, provided by the network allocation vector (NAV) at the MAC layer. The NAV makes use of 802.11 duration fields in order to reserve the medium for a fixed period of time. The NAV is a timer that indicates the amount of time the medium will be reserved. Stations set the NAV to the time for which they expect to use the medium. Other stations count down from the NAV to zero. When the NAV is non-zero, the virtual carrier sensing mechanism indicates that the medium is busy. When the NAV reaches zero, the virtual carrier sensing mechanism indicates the medium is idle.

Virtual Carrier Sensing and the Network Allocation Vector

Stations receiving a valid frame update their NAV with the information received in the Duration/ID field. This occurs only when the new NAV value is greater than the current NAV value and only when the frame is not addressed to the receiving station. Operation of the NAV can be seen in Figure 2.8.

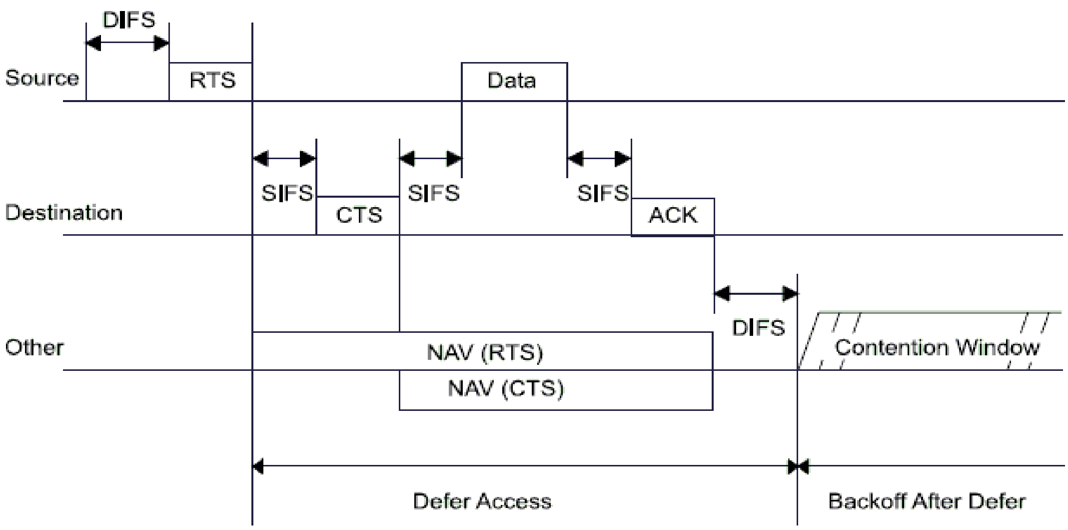


Figure 2.8: RTS/CTS/Data/ACK and NAV setting [10].

To ensure that the atomic frame exchange (outlined in Section 2.7.2) is not interrupted, the source STA sets the NAV in its RTS to block access to the medium while the RTS is being transmitted. Access to the medium is deferred until the NAV reaches zero by all stations that hear the RTS. As all stations may not hear the source STA and hence the RTS, the receiving STA replies with a CTS that includes a shorter NAV. Other stations are prevented from accessing the medium until the transmission is complete. Once the sequence is complete any station may transmit after they have waited for a period equal to a distributed interframe space (DIFS).

Interframe Spacing and Medium Access

As well as avoiding collisions, controlling transmission deferrals can provide different priority levels for different types of traffic. This is achieved by using different interframe spaces. The reasoning behind this is that when the medium is idle, high priority traffic can access the medium before low priority traffic gets a chance to by using shorter interframe spaces.

Figure 2.9 illustrates the basic access method for a contending station. There are 3 types of IFS specified. The short interframe space (SIFS) is used for high priority such as ACK and RTS/CTS control frames. Frames transmitted after SIFS have elapsed have priority over all other frames. PCF interframe spaces (PIFS), used by the PCF to provide contention free operation. DCF interframe spaces (DIFS) is the minimum idle time a contending station has to wait in order to gain access to the medium.

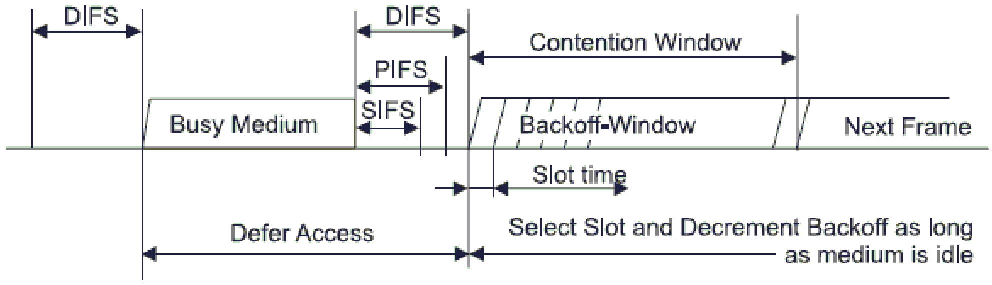


Figure 2.9: Basic access method [10].

If the medium has been idle for longer than DIFS, transmission can begin immediately. If the medium is busy, the station defers transmission and continues to monitor the medium until the current transmission is over. The station waits for the medium to be idle for the DIFS and prepares for the exponential back-off procedure.

After frame transmission has completed and the DIFS has elapsed, stations may attempt to transmit. A period called the contention window; CW (or back-off window) follows the DIFS. The CW is divided into slots, with the slot length being medium dependent (higher speed physical layers use shorter slot times). The CW size is initially set to *number of slots* $= (2^5\text{-}1)=31$. A station picks a random slot before attempting to access the medium; all slots are equally likely selections. When several stations are attempting to transmit, the station that picks the first slot (i.e. the station with the lowest random number) wins.

If a transmission fails, a retry counter is set and the back-off time is selected from a larger range. A binary exponential back-off mechanism operates such that CW sizes are always 1 less than an integer power of 2 (e.g. 31, 63, 127 etc). Each time the retry counter increases; the CW increases to the next greatest power of 2. When the maximum size has been reached it will remain at this value until it can be reset. The contention window is reset to 31 when the frames are successfully transmitted or when the associated maximum retry counter has reached its limit and frames have been dropped.

2.7.4 Point Co-Ordination Function (PCF)

As an optional access method, the 802.11 standard defines the PCF which enables the transmission of time-sensitive information. With PCF, a point coordinator within the access point controls which stations can transmit during any given period of time. The point coordinator makes use of PIFS when issuing polls. Because PIFS is smaller than DIFS, the point coordinator can seize the medium and lock out all asynchronous traffic while it issues polls and receives responses. Within a time period called the contention free period (CFP), the point coordinator will step through all stations operating in PCF mode and poll them one at a time. For example, the point coordinator may first poll STA 1, and during a specific period of time STA 1 can transmit data frames (and no other station can send anything). The point coordinator will then poll the next station and continue down the polling list in a round robin-fashion, while allowing each station configured for polling to have a chance to send data. Thus, PCF is a contention-free protocol and enables stations to transmit data frames synchronously, with regular time delays between data frame transmissions.

If this function were to be implemented on its own, PCF would lock out all asynchronous traffic by repeatedly issuing polls. Timing mechanisms within 802.11 make use of an interval known as a superframe to ensure that stations on the WLAN alternate between the use of DCF and PCF. As a result, the WLAN can support both asynchronous and synchronous information flows. For a period of time, stations will contend for access by using DCF. For the following time period, the stations will wait for a poll from the point coordinator before sending data frames. An example of the PCF using a superframe can be seen in Figure 2.10

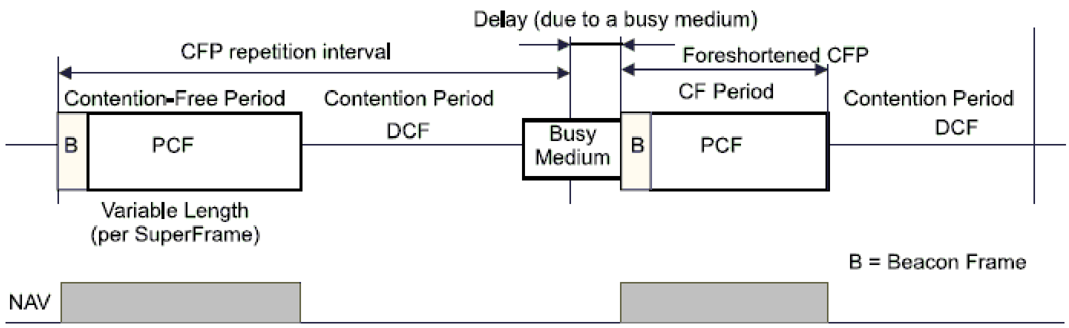


Figure 2.10: PCF Superframe Construction [10].

The optional PCF has been largely ignored (and has been avoided) by most of the major equipment manufacturers. This is partially due to the additional production costs involved. As such it is almost impossible to obtain hardware from vendors incorporating it. However, extensive research and analysis through simulation has shown the PCF to be largely ineffective [31], [32].

2.7.5 MAC Frame Format

The 802.11 MAC frame format is shown in Figure 2.11 below. This is a generic MAC frame format as not all frames use all the address fields. The values assigned to the address fields may change depending on the type of MAC frame being transmitted. Frames are transmitted from left to right with the most significant bits appearing last.

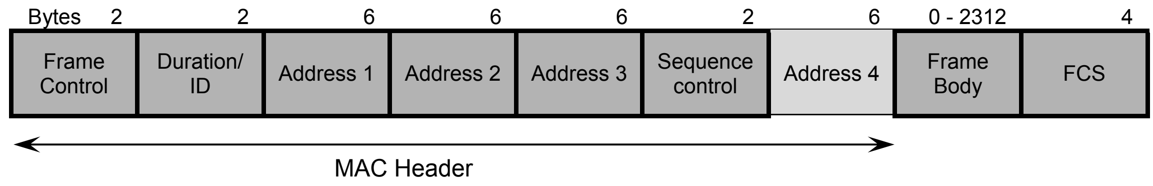


Figure 2.11: 802.11 MAC frame format.

Frame Control Field.

Each MAC frame begins with a 2-byte frame control field. A description of the frame control fields is outlined below and its format is illustrated in Figure 2.12

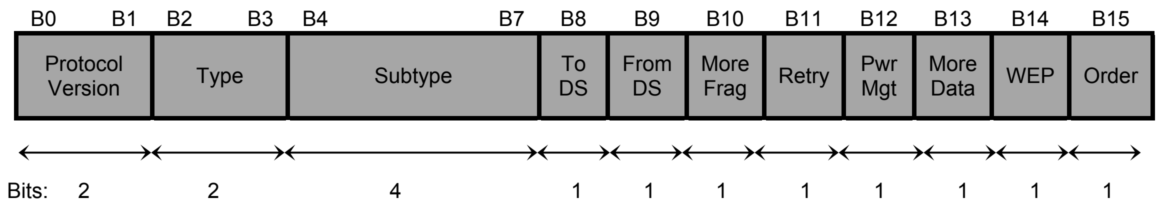


Figure 2.12: Frame Control Field.

- *Protocol Version*: The first two bits indicate the 802.11 version contained in the rest of the frame.
- *Type*: The next two bits indicate whether the type of frame in use is control, management or data.
- *Subtype*: This is a four-bit field which when used in conjunction with the Type field, further identifies the function of the field.
- *To DS*: This bit is set to 1 when the frame is addressed to the AP for forwarding it to the Distribution System (including the case where the destination station is in the same BSS and the AP is to relay the frame). The bit is set to 0 in all other frames.

- *From DS*: This bit is set to 1 when the frame is coming from the Distribution System.
- *More Fragment*: This bit is set to 1 when there are more fragments belonging to the same frame following this current fragment.
- *Retry*: This is set to1 if the frame is a retransmission of a previous frame.
- *Power Management*: This is set to 1 if the transmitting station is in sleep mode.
- *More Data*: Indicates that the station has additional data to send. Each block of data may be sent as one frame or a group of fragments in multiple frames.
- *WEP*: This bit indicates that the frame body is encrypted according to the optional wired equivalent privacy (WEP) algorithm.
- *Order*: This bit is set to 1 if the data frame is sent using the Strictly Ordered service which tells the receiving station that frames must be processed in order.

Duration/ID Field

This is a 2-byte field that follows the frame control field, its function depends on how bits 14 and 15 are set. In power-save poll messages this is the station ID. In all other frames this is the duration value in microseconds used to set the NAV.

Address Fields

IEEE 802.11 uses the same 48-bit convention for addressing as other IEEE 802 networks. If the first bit is set to 0 then this represents a unicast address.

If the first bit is set to 1 then this will represent a multicast address. If all bits are 1’s, then the frame is a broadcast and is delivered to all stations connected to the wireless medium. The address fields are numbered, as their meaning may change depending on the frame type (i.e. depending on how the To DS and From DS fields are set). Address types include source address (SA), destination address (DA), transmitting station, and receiving station. Generally Addresses 1 – 3 are set, address 4 is only set when wireless bridging is in use.

Sequence Control Field

The Sequence Control Field is used to represent the order of different fragments belonging to the same frame and to recognise packet duplications. It consists of two sub-fields, Fragment Number and Sequence Number which define the frame and the number of the fragment in the frame.

Frame Body

The frame body contains a MAC service data unit (MSDU) or a fragment of an MSDU. This is also known as the Data field and its purpose is to move higher level payloads from station to station.

Frame Check Sequence (FCS)

The 802.11 frame ends with a 32-bit FCS field which contains a 32-bit cyclic redundancy check (CRC). The FCS allows stations to check the integrity of received frames respectively.

2.8 Statistical Procedures

Network engineers often face the problem of resource allocation. For network operators this also involves dealing with the uncertainty inherent in delivering services across packet switched networks. Ideally, network engineers and service providers require their networks to provide controlled levels of loss, delay, and jitter. They may choose to push the performance limits of their existing assets at the cost of any certainty about the quality they may achieve or they may aim to provide a predictable performance by over provisioning the network according to peak requirements of all the applicants that are using it. Figure 2.13 shows the components of network operation that impact IP-Certainty, namely network resources (such as bandwidth), traffic, and quality, and are all intrinsically linked. Changes in one affect the relationship of the other two; for example, the bandwidth required to meet a delay target depends not only on the load on the network, but also on whether it carries VoIP or video or data traffic [33].

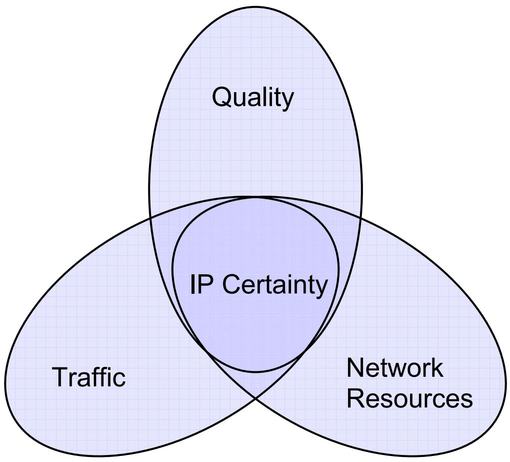


Figure 2.13: Relationship between network components and IP-Certainty [33].

It is possible to model the effects of traffic usage under different network conditions. The effects of data as it propagates through network nodes can then be analysed. This enables the engineer to design and dimension buffers and to estimate the required Effective Bandwidth (EB). Effective bandwidth is a concept for characterising network traffic streams in order to dimension buffers in routers and switches [34], [35], [36]. Another consideration that must be taken into account is the nature of the data transmitted across the network. This is of particular importance in real-time applications, as the characteristics of the traffic stream may fluctuate considerably. For instance, the peak rate of a video stream may be up to10 times larger than its mean rate. To prevent losses, the network should allocate to each stream a bandwidth equal to the streams peak rate. This type of allocation is overly conservative and it may be possible to allocate a bandwidth much closer to the mean rate rather than the peak rate. There are two fundamentally different techniques a network may employ in other to reduce the bandwidth allocation. One way is to multiplex various different sources. The other method is to use buffering.

2.8.1 Statistical Multiplexing

Statistical multiplexing is a method widely used in ATM networks. However, general observations show that these methods may be employed when analysing the downlink of an AP which will essentially be the aggregate of the individual uplink streams. The analysis of the multiplexing method determines how many sources must be multiplexed and the transmission rate per source required, so that the probability that the total rate of the sources exceeds the transmitter rate is smaller than some specified value, say 10^{-9} [37].

For example consider N sources. For $n = 1, . . . , N$, let Y_n be the rate at some time t of source number $n.$ We assume that the sources are stationary, independent and identically distributed. That is, the rates $\{Y_1, . . . , Y_n\}$ are independent random variables that have a common distribution that does not depend on t. We want to find the rate c such that

$$P\{Y_1 + \ldots + Y_N > cN\} \leq 10^{-9} \quad (2.1)$$

In other words if a node transmits the superposition of the N sources with that rate c, then it drops at most a fraction 10^{-9} of the bits [37].

The multiplexing method makes use of the fact that separate sources fluctuate independently from each other. The effect of this might be such that when the rate of a source is larger than the average, the rate of another source may be less than the average. Consequently, the aggregate sum of many sources may be closer to its average value. This manifests itself as a reduced variance in the rate of the aggregate load.

2.9 Buffers and Delays

A buffer holds data for the purpose of caching before it can be forwarded on to its next destination. As data packets arrive at a buffer they are queued up. [38] Briefly outlines some basic servicing mechanisms for a queue that has a limited buffer space. In the case of a first in, first out (FIFO) buffer the first value placed in a buffer queue is subsequently the first value read out or serviced (Figure 2.14). For this type of buffer there are two main effects concerning the buffer size. If the buffer is relatively small the buffer will quickly fill up and the probability of packet loss will increase. On the other hand if the buffer size is relatively large the probability of packet loss will decrease however the buffer transit time will increase as the buffer fills up. In order to study the buffer behaviour we must take into consideration two separate parameters, namely,

- Arrival rate X
- Service rate Y

Consider the case of a single server queue with stationary arrivals (X) and service rate Y. For a stable system we require that

$$\boldsymbol{E}\,[Y] > \boldsymbol{E}\,[X] \tag{2.2}$$

In other words, if the average rate of arrival is much greater than the service rate the buffer will soon fill up and overflow, preventing any guarantee of quality of service. By analysing the tails of the distribution of the probability of delay times, it is possible to determine the characteristics of rare events such as the probability of buffer overflow. In the case of traffic engineering this may give an indication of the probability of long delays. The theory of large deviations [39], [40] tells us that if the tails of the queue length distribution are asymptotically log-linear, then

$$\lim_{q \to \infty} \frac{1}{q} \log P(Q > q) = -\delta \tag{2.3}$$

Where $P(Q > q)$ is the frequency with which the queue length exceeds some level q. If we plot the $\log P(Q > q)$ against q then the asymptotic slope can be shown to be $-\delta$. This type of graph is sometimes known as a *Watermark plot* and with sufficient data the plot will typically have a linear region with an asymptotic slope close to $-\delta$ before running out of data at levels which are rarely exceeded.

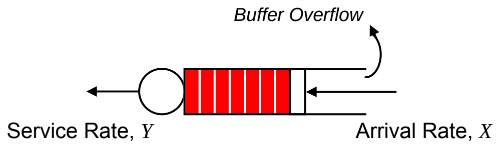


Figure 2.14: FIFO buffer queue.

If we therefore assume a simplified buffer model with a constant service rate we can show that,

$$\mathrm{P}(overflow) \approx e^{-\delta} \tag{2.3}$$

For a large slope the tails of the distribution will decay rapidly. This will indicate a small probability of finding large values of delay and hence a small probability of buffer overflow. For a small slope the tails of the distribution will decay slowly. This result will indicate that there is a significant probability of finding large values of delay and hence a significant probability of buffer overflow. This method can be used in order to analyse the AP transit times for VoIP packets. Examples of a PDF with heavy and light tails and the corresponding Watermark plot are shown in Figures 2.15 and 2.16.

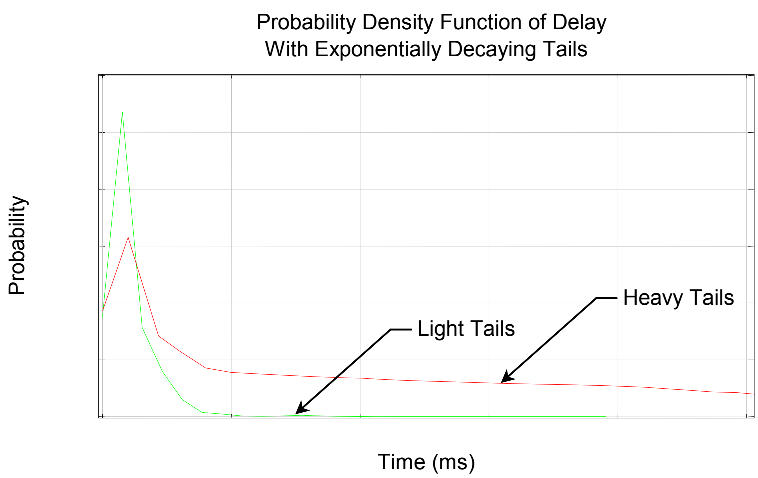


Figure 2.15: Example of PDF with heavy and light Tails.

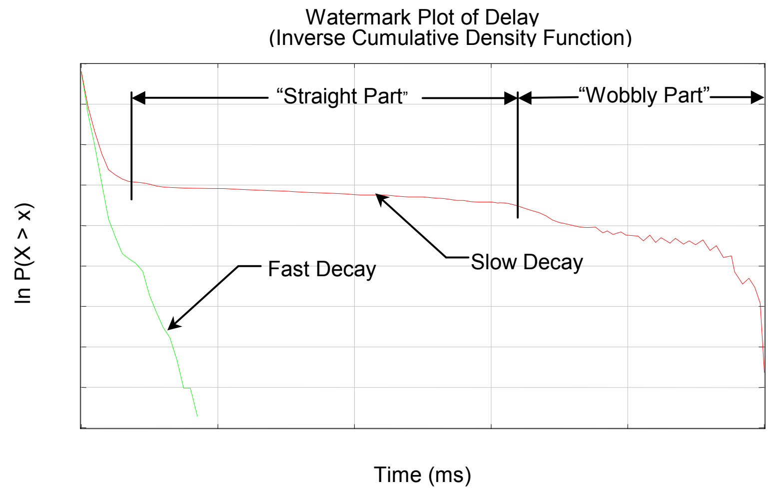


Figure 2.16: Watermark plot for distributions shown in Figure 2.15.

2.10 Network Analysers

Network analysers are useful tools when planning network deployment. Many WLAN network analysers are readily available and for some commercial suppliers it was a simple extension to their catalogue of wired network products. However, 802.11 is not Ethernet and has a number of additional protocol features, each of which can cause its own set of problems. Network analysers allow network administrators to view low level details and to quickly focus on problem areas for troubleshooting.

Wireless network analysers are no different in this respect. However, because of the added security risk from rogue stations and vulnerability of Wired Equivalent Privacy (WEP), wireless network analysers are possibly crucial for any wireless infrastructure.

Most commercially available (and some freely available) wireless network analysers are essentially protocol analysers. They all work in a similar way with the main difference being on which layers are being analysed. Some can decode higher-level protocols, including the entire TCP/IP stack. Others concentrate decoding lower layers such as detecting interference sources by focusing on the physical layer. Many centre on security issues and the detection of rogue stations. Others include planning and mapping tools and sometimes allow interfacing with GPS devices in order to pinpoint hotspots. Most analysers provide some sort of data output that can be saved to a file for post analysis processing whilst others can generate statistical reports and graphs. During the course of the experimental set-up the wireless network analysers *Airopeek NX* [41] and *Ethereal* [42] were used for general planning and verification of network operation.

3 QoS Analysis

3.1 Speech Quality

3.1.1 A General Definition of Speech Quality

According to ETSI Guide on Speech processing, Transmission and Quality aspects; Mouth-to-ear speech quality (also "end-to-end speech quality") is defined as the degree of speech quality that a listener perceives at his terminal with a talker at the far end [43]. The guide goes on to mention that this definition is not absolute and further clarifies its definition by raising the following points;

- An absolute physical definition of "speech quality" does not exist; the only "baseline" we have is the subjective perception of human listeners.
- Speech quality ultimately is a psycho-acoustic phenomenon involving a complex interaction of many parameters within the process of human perception, although many of the individual parameters can be measured purely electrically.
- Mouth-to-ear in this context implies that there is a transmission of the speech signal by some kind of network; it is to be defined what that network consists of.
- In today’s liberalized environment a network provider can no longer prescribe the terminal equipment being used by his customers; his reach and therefore his responsibility is limited to his network and ends at the outlet on the customer's premises.

- Speech quality is but one component of the overall quality perceived by a telecommunications user.

So in brief, perception of human speech quality will ultimately depend on the person listening. The quality of the transmitted speech signal will also depend greatly on the network which it uses. Furthermore, although perceived quality is dependent on the listener many of its parameters can be measured and analysed objectively.

3.1.2 Human Perception Characteristics of Speech Quality

Although it may not be possible to predict the perception of human speech quality it is however possible to measure physical factors and aspects of a speech signal transmitted through a network. Therefore, when we talk about speech quality we will be referring to a statistical mean. The many physical characteristics can impact differently on how the individual perceives speech and can be determined broadly by the following parameters; Intelligibility, Naturalness, and Loudness. Impacts of these can be seen in [43].

As mentioned previously, differences in perception exist between individuals, such as hearing sensitivity and how aware the individual is (auditive cognition). As well as this, an individual will not always perceive speech in the same way, but instead will depend on the individuals' mood and interest. Another consideration which must be taken into account is language. There is a direct correlation between frequency range and spectra of consonants.

3.1.3 Speech Samples

Although there is no commonly agreed set of speech samples for either objective or subjective experiments, there are a number of characteristics and parameters which the samples should include. The International Telecommunication Union (ITU) provides a set of clean speech samples with both male and female speakers included. The signals are sampled at 16 kHz with a 16-bit resolution using PCM. It is also possible to use artificial speech for objective measurements, however it is still necessary to use real speech as a reference.

The characteristics of speech should meet the following criteria.

Physical Characteristics	
Frequency range:	300 Hz to 3400 Hz (narrowband) 100 Hz to 7000 Hz (wideband)
Duration:	5 s to 10 s (not including header and trailer sequences)
Density:	70% speech, 30% pauses.

Phonetic Characteristics	
Language(s)	English, German, French, Swedish, Italian
Distribution of phonemes:	The distribution within the sample should represent the standard distribution of phonemes in the chosen language
Female / male speakers:	50% / 50%

Recording Characteristics	
Sample rate:	≥ 8 kHz (narrowband systems, at network insertion point); ≥ 16 kHz (wideband systems, at network insertion point)
Digital resolution:	> 12 + 1 bit (original stored sample)
Recording resolution:	> 95 dB (i.e. 16 bit)
DC offset:	0

Table 3.1: Guidelines for speech samples [43].

3.1.4 Artificial Speech

Artificial speech is a method of generating a signal that closely resembles the characteristics of human speech. This signal can then be used for repeating objective performance measurements. As conversational speech is not one continuous stream, the artificial speech should reflect this characteristic. ITU-

T Recommendation P.59 [44] outlines the characteristics of conversational speech and defines a method for generating artificial speech based on an artificial voice described in ITU-T Recommendation P.50 [45]. The recommendation states that the signal should reflect parameters of human speech such as the length of talk-spurt, pause, double talk and mutual silence. The recommendation also states that artificial conversation should be at least 10 minutes in duration. The following figures show the transition state diagrams for the model, p_1, p_2 and p_3 denote transition probabilities of 40%, 50% and 50% respectively.

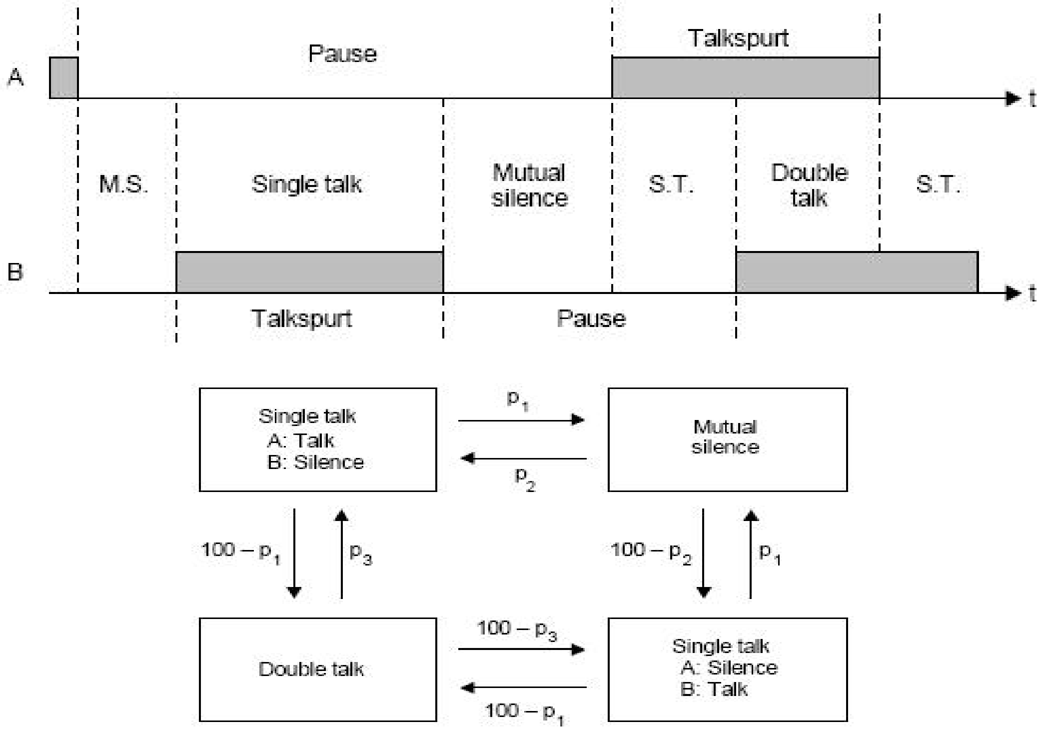


Figure 3.1: State transition model for conversational speech [44].

3.2 Quantification & Measurement of Speech Quality

As there is no commonly agreed objective definition of "speech quality", assessment of this attribute is necessarily subjective. As outlined above this would have to be a statistical mean of any data assessed. In order to quantify speech quality in an objective and repeatable manner, individual aspects need to be eliminated from the assessment. The various methods for achieving this can be broadly grouped into two main categories, namely *Subjective Measurement Methods* and *Objective Measurement Methods*.

Subjective measurement methods involve carefully designed experiments where a statistically valid number of individuals listen to the speech samples. Each individual rates the speech sample according to a scale. An average of the individual scores are taken to give a Mean Opinion Score (MOS). The MOS results in a more accurate representation than the individuals' perception. In contrast, objective comparison measurement methods are more recent. This method involves computing a quality value and estimating the MOS. This method is much preferred as it can be automated and it takes less effort in setting up, compared to subjective methods.

3.2.1 Subjective Measurements

As mentioned previously, subjective measurements can be extremely time consuming to conduct and take a great deal of effort to set up. However, it is necessary to carry out these tests in order to have a reference for objective measurements. ITU-T Recommendation P.800 [46] contains guidelines on different methods for subjective measurements and how to conduct these tests. The results of subjective listening-opinion tests are influenced by a wide variety of conditions. Some of the factors to be controlled are [46]:

- *Speech material:* Perception depends on the gender of talkers, their pronunciation, the language, length and content of samples, the recording room and equipment characteristics.
- *Experiment set-up:* Results can depend on nationality and gender of listeners, recent previous experience with listening tests, instruction of listeners about the experiment, duration of test sessions, and order of presentation of speech samples.
- *Listening conditions:* Loudness of presented speech samples and choice of equipment (headphones/telephone handsets) can influence the rating.

ITU-T Recommendation P.800 [46] gives guidelines for four main methods of subjective testing; Conversation-Opinion Tests, Listening-Opinion Tests, Interview and Survey Tests and SIBYL (based largely on Interview & Survey method)

- Conversational Opinion Tests: This involves two individuals engaged in conversation across an audio connection. Typical parameters that can be assessed using this method are: overall quality, speech quality, and difficulty in talking or listening. This type of test is important as it provides the closest simulation of real telephone interactions between two subscribers.
- Listening Opinion Tests: These types of tests involve individuals listening to a set of speech samples and rating them based on a predefined marking scheme. ITU-T Recommendation P.800

recommends a method for listening-only tests known as Absolute Category Rating (ACR).

- Interview and Survey Tests: If the rather large amount of effort needed is available and the importance of the study warrants it, transmission quality can be determined by "service observations". Recommended ways of performing these, including the questions to be asked when interviewing customers, are given in ITU-T Recommendation P.82 [47]. To maintain a high degree of precision a total of at least 100 interviews per condition is required.
- SIBYL: This allows a small proportion of a user’s ordinary calls to be modified and for the quality to be distorted according to a test program. If a particular call has been so treated, the volunteer is asked to vote by dialing one of a set of digits to indicate his opinion. This allows for a certain amount of automation.

3.2.2 Objective Measurements

In order to consistently measure speech transmission on a regular basis it is necessary to use objective measurement methods [48]. Objective determination of speech quality is based on two distinct methods: Signal-based methods and Parameter-based methods.

Signal based method: This is an automated method that involves distorting the original signal in a known way prior to transmission. The received signal is compared to the original undistorted signal. A rating is then derived from an algorithm in order to predict a perceived MOS. ITU-T Recommendation P.862 [49] recommends a method known as Perceptual Evaluation of Speech

Quality (PESQ). Other signal based methods include, Perceptual Analysis/Measurement System (PAMS) and Perceptual Speech Quality Measurement (PSQM) amongst others. PESQ is the most popular choice as its performance shows a high correlation with subjective scores on a large number of databases covering a large number of conditions [43].

Parameter-based methods: These types of methods are represented mainly by the E-model described by ITU-T Recommendation G.107 [50]. The E-Model is a transmission-planning tool that can estimate the end-to-end voice quality, taking the IP Telephony parameters and impairments into account. The model can be used for estimating the user satisfaction of a narrowband, handset conversation, as perceived by the listener. It is not intended for predicting absolute user satisfaction. Instead, the intention is to model the performance of an unknown connection relative to a connection with known performance. The E-Model has proven to be a versatile tool that has adapted well to the impairments of IP telephony [51].

The input to the E-model consists of parameters which are available at the time of planning. This makes the E-model very flexible as planning can be carried out either before or after installation of a network. These parameters include factors such as noise, delays, and echoes. These factors are subject to internationally accepted standards and recommendations, or are known from prior measurements. Additionally, the E-model weights the influence of low-rate codecs, multiplexers etc. on communication quality. The E-model is based on the assumption that transmission impairments can be transformed into "psychological factors", and that these factors are additive on a "psychological scale". In other words, the subjective perception of speech quality is supposed to be equal to the sum of transmission impairments. The E-model first computes a "base value" for quality which is determined from network noise. Each further impairment is expressed as an impairment value which is subsequently subtracted from the base value. This results in a predicted speech quality for a specific network. Finally, the resulting value for the speech quality can be used to estimate what fraction of the user population would rate the quality as "good or better" (GOB) and "poor or worse" (POW) [43].

The output of the E-Model is a scalar called the “Rating Factor”, the “R-value”, or simply R. The scale is typically from 0 to 100 [51]. The equation for the transmission-rating factor R is:

$$R = Ro - Is - Id - Ie + A \tag{3.1}$$

Where,

Ro – effects of noise and loudness ratio

Is – simultaneous speech transmission impairments

Id – delayed impairments relative to speech signal

Ie – equipment impairment factor

A – advantage factor, compensation for some additional convenience.

A comparison of the E-model output and MOSs is shown in Figure 3.2 below.

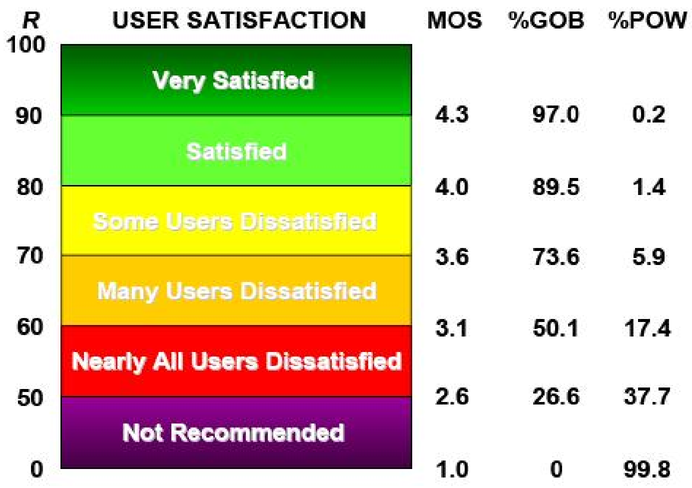


Figure 3.2: Comparison of E-model output scales [51].

A typical curve displaying R *vs delay* is shown in Figure 3.3. The plot also includes the E-model scale and indicates how delay is in an important factor in IP telephony. The ITU-T Recommendation G.113 [27] supplies a list of Equipment Impairment Factors (Ie) for various codecs. As can be seen in Figure 3.3, codecs with speech compression (G.729A and G.723.1) have larger Ie values and can therefore tolerate less one-way delay for a given voice quality level. The Telecommunications System Bulletin TSB-116 [51] makes the following voice quality recommendations for IP telephony,

Delay Rec. #1: *Use G.711 end-to-end because it has the lowest Ie-value and therefore it allows more delay for a given voice quality level.*

Speech Compression Rec.#1: *Use G.711 unless the link speed demands compression.*

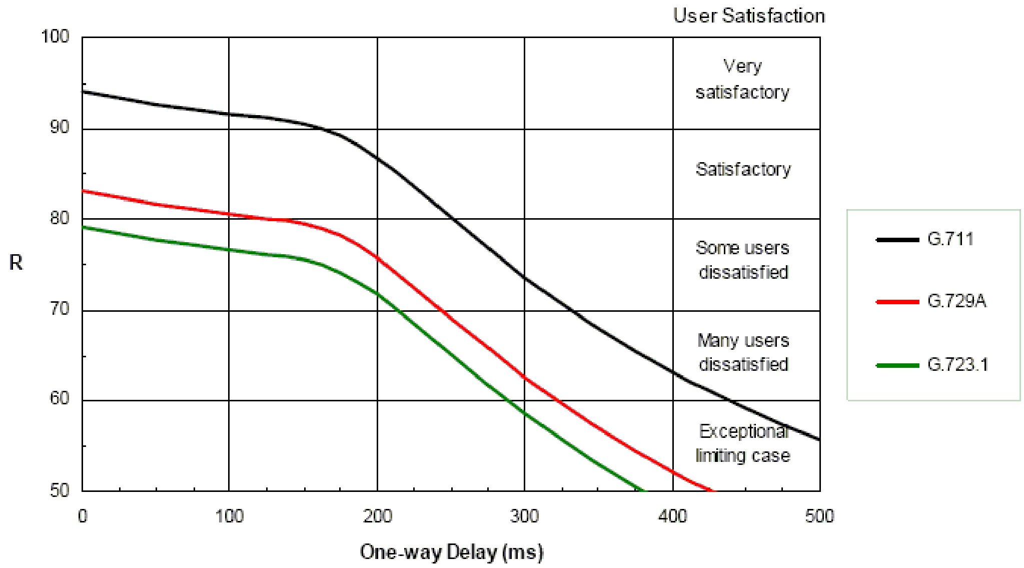


Figure 3.3: Speech Compression Impairment [51].

3.3 Understanding The MAC Bandwidth Components

The basic operation of the MAC in the IEEE 802.11 Standard [10] is described in Section 2.7. We now describe a framework for performance characterization and resource utilization that is based upon the concept of *MAC Bandwidth Components* as presented in [52] and also in [53]. There is a direct relationship between these MAC bandwidth components and the line rate (e.g. 11Mbps for 802.11b). The MAC bandwidth components are indicative of how the bandwidth of the wireless medium is utilized by all contending stations including APs.

The 802.11 MAC operation describes various timing mechanisms called Inter Frame Spaces (IFS). The basic idea of the MAC bandwidth components is to transform the relationship between IFS and MAC service data units (MSDUs) into a fraction of the line rate. By doing this, the bandwidth utilization can be identified in a more intuitive fashion. In particular [52] and [53] put emphasis on three MAC bandwidth components; BW_{Load}

which is related to throughput, BW_{Access} which is related to the 802.11 contention mechanism and BW_{Free} which is related to the availability of spare capacity and is an indication of likely QoS. Two other components are also addressed, namely BW_{Busy} which is related to overall network usage and BW_{Idle} which is related to the time the medium is not being utilized.

As can be seen in Figure 2.9 (page 28) there are times when the network is “busy”. This corresponds to periods when data frames and their associated acknowledgements are being transmitted or simply the network load. These periods are complemented by periods when the network is idle, i.e. there is no traffic on the medium.

A station utilizes these idle intervals in two ways. If a *STA* wishes to transmit, it must first wait DIFS and then an additional random amount of time based on the 802.11 back-off algorithm. This period is known as the access time when a *STA* is actually contending for access to the medium. If a *STA* does not try to gain access for all of the available idle time, then the remainder is seen as spare capacity known as “free” time. This spare capacity, as seen by a *STA*, is closely related to QoS, in that the more available, then better the QoS should be (e.g. if a *STA* is contending less, then it stands to reason that it will have lower delays). The timing diagram in Figure 3.4 illustrates the interaction between the various components when two *STAs* are contending for access to the medium.

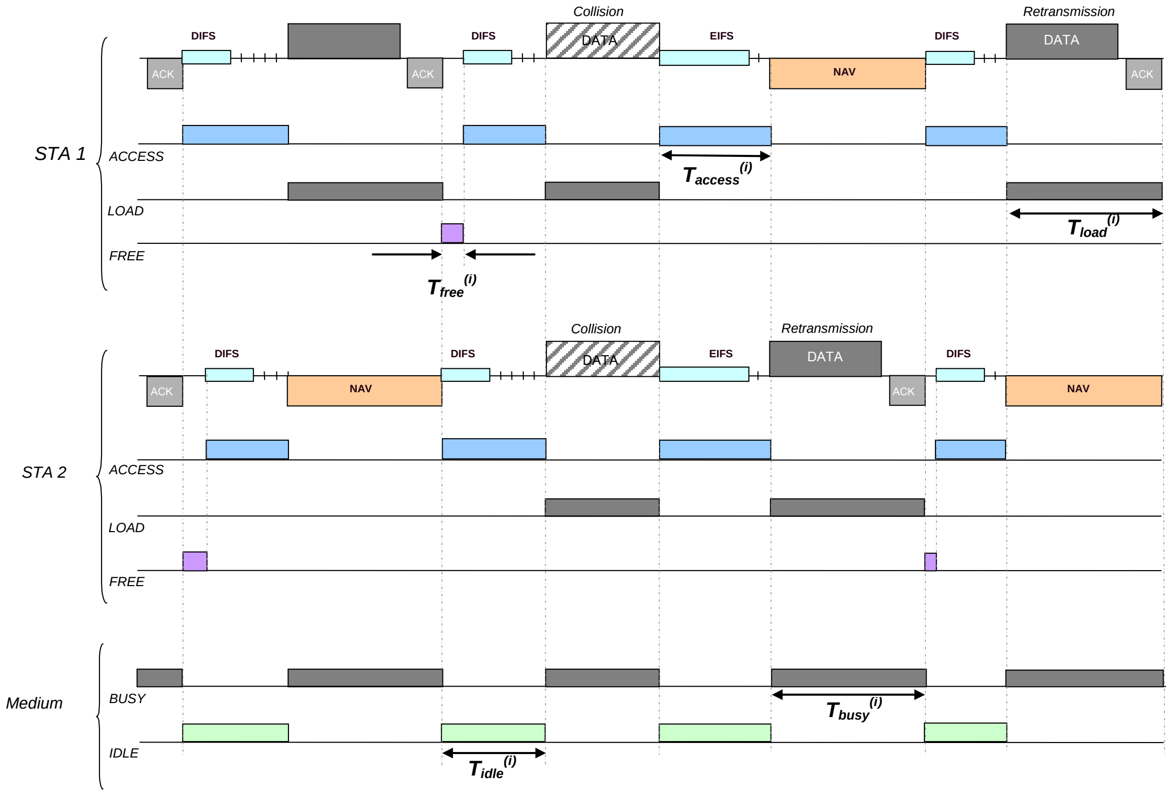


Figure 3.4: Interaction of the MAC timing components.

Busy and idle times are summed (over some specified measurement time interval) as follows;

$$T_{Busy} = \sum_i T_{Busy}{}^{(i)} \tag{3.2}$$

and

$$T_{Idle} = \sum_i T_{idle}{}^{(i)} \tag{3.3}$$

Where $T_{Busy}{}^{(i)}$ and $T_{Idle}{}^{(i)}$ are durations of the i^{th} busy and idle intervals respectively, within the specified time period. These periods are then converted into more intuitive bandwidth values by normalizing them and relating them to the line rate.

$$BW_{Busy} = \frac{T_{Busy}}{T_{Busy} + T_{Idle}} \times Line\,Rate \tag{3.4}$$

and

$$BW_{Idle} = \frac{T_{Idle}}{T_{Busy} + T_{Idle}} \times Line\,Rate \tag{3.5}$$

so that,

$$BW_{Busy} + BW_{Idle} = Line\,Rate \tag{3.6}$$

By analysing the MAC header information it is possible to identify the period a particular station is transmitting its load. This can then be translated into a MAC bandwidth component that is directly related to a *STAs* throughput. So for a particular *STA (k)*, the BW_{Load} (including its associated collisions) may be calculated as follows,

$$BW_{Load}(k) = \frac{T_{Load}(k)}{T_{Busy} + T_{Idle}} \times Line\,Rate \tag{3.7}$$

For the single station case BW_{Busy} and BW_{Load} will be identical, but for multiple stations there will be an inevitable waste of bandwidth due to collisions ($BW_{Collisions}$) when two or more stations are present;

$$BW_{Busy} = \sum_k BW_{Load}(k) - BW_{Collisions} \tag{3.8}$$

As stations are independent, once an individual station has gained access to the medium, then that station alone has exclusive use of the medium and therefore load bandwidths are also independent of each other. However, in contrast to this all stations will perceive the same idle bandwidth. So for any *STA (k)* the following applies,

$$BW_{Access}(k) + BW_{Free}(k) = BW_{Idle} \tag{3.9}$$

or,

$$BW_{Busy} + BW_{Access}(k) + BW_{Free}(k) = Line\,Rate \tag{3.10}$$

All that remains now is to determine the BW_{Access} or the BW_{Free}. As the BW_{Free} is essentially spare capacity it makes sense to first define the BW_{Access}. As outlined previously the access time can be considered in two parts. Access to the medium involves waiting a certain amount of time (DIFS) deferring for access which may or may not involve more than one deferral and also a random amount of time based on the 802.11 back-off counter. Because these two periods are essentially random, a mean value is obtained over a specified interval, such that;

$$\overline{T}_{Access} = \overline{T}_{Defer} + \overline{T}_{Back-off} \tag{3.11}$$

This mean value allows BW_{Access} (and consequently BW_{Free}) for a particular $STA\ (k)$ to be obtained;

$$\overline{BW}_{Access}(k) = \frac{\overline{T}_{Access}(k)}{T_{Busy} + T_{Idle}} \times Line\,Rate \tag{3.12}$$

In [52] and [53] it is suggested that these parameters can be used to detect the on-set of saturation. This may be classed when either $BW_{Free} = 0$ (i.e. no more spare capacity available) or when $BW_{Access}(k) = BW_{Idle}$ (i.e. $STAs$ are using all available bandwidth to win access opportunities for their loads). The WLAN *Probe* as presented by [52] is a C/C++ application that operates on a Linux platform. The WLAN Probe captures frames and processes the MAC bandwidth components after which it creates a log file of the processed data. Further details of this operation are given in Section 4.1.3.

3.4 Network Capacity

One may consider that for an 802.11b network an 11Mbps line rate would be sufficient to accommodate most users in a lightly used network. Indeed many manufacturers will say that such a network will support approximately 50 – 100 users. However, the situation is different when considering real-time traffic such as VoIP which is highly sensitive to delays and packet loss. To calculate the capacity of a network it is necessary to consider not only the data load but also the associated protocol overhead to transmit this data. Let us consider the G.711 audio codec sampled at 64 kbps which supports three payload types as shown in the table below.

Codec: G.711	
Payload (bytes)	**Frame size (ms)**
80	10
160	20
240	30

Table 3.2: G.711 fixed payload sizes.

We will also assume a fixed protocol overhead of 40 bytes (i.e. IP 20 bytes, UDP 12 bytes and RTP 8 bytes). It is then possible to calculate the bandwidth requirement for a single VoIP call and consequently the maximum number of simultaneous pairs as follows;

Assumptions:

- *G.711 Codec (64 kbps/ 30 ms): Payload = 240 bytes every 30 ms*
- *Line Rate = 11 Mbps*
- *Percentage of Line rate available due to protocol overhead = 70%*
  (Max % of Line rate available)
- *Voice activity factor = 0.43*

$$packets\ per\ second = \frac{1}{(frame\ size)} \tag{3.13}$$

$$= \frac{1}{30 \times 10^{-3}} = 33.33\ pps$$

$$\begin{array}{l} Throughput\ requirement \\ at\ network\ layer \end{array} = (packet\ sze + Overhead) \times pps \tag{3.14}$$

$$= [(240 \times 8) + (40 \times 8)] \times 33.33$$

$$= 74.67 kbps$$

$$\begin{array}{l} No. of\ simultaneous \\ pairs\ of\ calls \end{array} = \frac{Line\ rate \times 0.70}{2 \times Throughput\ requirement \times voice\ activity\ factor} \tag{3.15}$$

$$= \frac{11{,}000 \times 0.70}{2 \times 74.67 \times 0.43} = 119$$

Similar calculations can be performed for payloads of 160 bytes and 80 bytes with the number of calls being 111 and 92 respectively. Indeed, similar calculations could be performed for any voice codec. There are two main points of interest here. One is that the choice of codec will determine the maximum number of VoIP calls permissible on the medium. The other point is that the calculated number of calls would appear far higher than anything reported in the literature. This second point then poses the following questions:

- What is the maximum number of VoIP calls?
- Why is it not as high as these calculations would show?

3.4.1 VoIP over WLAN Literature Overview

In [54] *Anjum et al* investigate the capacity of IEEE 802.11b networks using VoIP in the presence of data traffic. They also investigate methods for improving performance by implementing back-off control and priority queuing (BC-PQ) in the access point. This enhanced mode access point uses a combination of marking VoIP packets as high priority and allowing them to "jump" the queue as well as setting the back-off counter for VoIP packets to zero (enabling them to be sent as soon as they reach the top of the queue).

In this paper, the maximum number of VoIP calls is quantified by only considering conversations with less than 2% packet loss for both uplink and downlink traffic. It should also be noted that this experiment uses simulated speech in accordance with [44], emulating the G.711 codec with 10ms packetisation intervals with a payload of 80 bytes. The experimental set-up allowed for the number of VoIP streams to be increased from 1 to 10 and background traffic to be increased from 1Mbps to 4Mbps, as well as employing a BC-PQ scheme.

The experiment indicated that in the absence of background traffic 802.11b could support at least 10 VoIP calls. When "*light background traffic*" was introduced at 1Mbps this figure was reduced to 7 VoIP calls. They then considered the effects of introducing "*heavy background traffic*" at 2 Mbps and 4Mbps. Under these conditions (and without using the BC-PQ) they show that the maximum number of VoIP calls capable of being supported is 5 with a background load of 2Mbps. However, with a load of 4Mbps their experiment shows that the medium is not capable of supporting any VoIP users.

The paper concludes by stating that background data traffic contributes significantly to the downlink traffic. Voice traffic will suffer a high percent packet loss as large data packets contend with voice data packets at the access point. As such, the downlink traffic usually dominates. They show that performance can be greatly increased by implementing Back-off Control and Priority Queuing (BC-PQ) at the access point.

In [55] *Casetti* and *Chiasserini* consider the throughput of real-time traffic in WLANs using the IEEE 802.11e Enhanced Distributed Channel Access (EDCA). Although our experimental set-up does not use 802.11e the findings are still relevant as the test scenario described here uses a line rate of 11Mbps. The testing scenario involves the use of simulated speech. The model used mimics two-speaker interaction determined by a Pareto-distributed random variable. The codec used is the G.729A with a 20 ms frame duration. This particular codec has an encoded bit rate of 8 Kbps and an average bit rate of 2.8 kbps (resulting from the use of Voice Activity Detection, VAD). The experiment stipulates QoS requirements comprising delays less than 150ms with a packet loss rate less than 3% for good quality and delays between 150-400 with a maximum packet loss rate of 7% for medium quality. The number of simulated wireless stations is increased from 4 to 40 and also includes light TCP background traffic (simulating web browsing). Furthermore, the tests first allow traffic to be differentiated by type only and then by traffic type and direction.

These tests show that when priority is given to real-time traffic over TCP background traffic, the maximum number of simultaneous VoIP users capable of being supported on an 11Mbps WLAN is between 28 and 30. When priority is given to traffic type and direction (i.e. downlink traffic) this figure is increased slightly to 32 VoIP users.

The paper concludes that the reason for the increase in VoIP users is that traffic congestion occurs at the downlink. By giving priority to real-time traffic on the downlink the number of VoIP users can be increased while still operating within QoS requirements.

In [56] *Elaoud* and *Agrawal* investigate the ability of wireless applications to deliver high quality voice services. The paper is an experimental evaluation of VoIP capacity in IEEE 802.11b in an attempt to identify the contributing factors that limit the capacity (in their view up to 10 users per access point. See also [57] and [58]). This paper extends the work of [54] by incorporating both packet loss and delays. They suggest performance metrics of up to 200ms and a packet loss rate of up to 2% [59] [60]. Their experiments determine the capacity as well as the effects of packet header overhead. The authors also introduce a performance metric called the Packet Success Ratio (PSR) that measures the percentage of voice packets successfully delivered within a specified time (i.e. a deadline). Packets that do not meet this deadline are not considered.

The experiment's test bed involves 15 laptops connected to an AP and four desktops connected to the AP via a router. Again, simulated speech is used in accordance with [44], emulating the G.711 codec with 10ms packetisation intervals with a payload of 80 bytes. The experiments show that the maximum number of VoIP users that the network can support is 10. This figure does not include performance based on delay but only the effects of packet loss. The deadline for maximum permissible delay (as used by the PSR) is varied to identify the effects of "LAN deadline" and "WAN deadline". This results in the number of VoIP users being reduced to approximately 6 and 4 for LAN and WAN respectively.

The authors then consider the effects of packet overhead and suggest that the average back-off mechanism delay [10] is more than 20 times larger than the payload transmission time. As a consequence the back-off delay constitutes the largest overhead and represents a congestion point that limits the capacity of the network. The paper concludes by suggesting that the access point requires a greater share of access to the medium over contending stations in order to increase the number of VoIP users.

In [61], *Medepalli et al* investigate analytically and through simulation, the capacity for voice data in IEEE 802.11a/b/g networks operating in infrastructure mode. They set out to determine the capacity and the limiting factors for such networks. They present findings which indicate that the capacity of VoIP is a strong function of the channel bandwidth, voice codec packetisation interval, and data traffic in the system.

The experimental set-up involves full duplex VoIP from wireless to wired stations via an access point and router. Experiments use simulated speech which is modelled on a four state Markov chain between two users as recommended by [44]. The model emulates the G.711 codec with a sample rate of 64 kbps with the packetisation intervals used between 10 ms and 50 ms in steps of 10 ms in different test scenarios. Testing scenarios consider different 802.11 physical layers (i.e. the different line rates for 802.11a, 802.11b & 802.11g). Voice-only traffic is initially considered before introducing constant bit rate (CBR) video traffic and analysing its effects on the capacity. The authors refer to maximum capacity when QoS requirements of maximum round-trip delay of 200ms and maximum packet loss rate of 2% are met.

The results compare analytical values of capacity to simulated experimental results. Their findings show a close link between calculated values of maximum capacity and simulation. In their findings they show that the maximum number of VoIP users on an 802.11b infrastructure to be 11, 21, 30, 38 and 44 with packetisation intervals of 10, 20, 30, 40 and 50 respectively (closely matched by analytical results allowing for approximately one additional user). Their observations were that in infrastructure mode, the downlink (from the access point to a wireless station) is the bottleneck and routinely dictates the capacity as also indicated in [54]. They also observed that larger packetisation intervals achieve higher capacity. When video traffic was introduced, results showed that voice capacity is lowest when the video packet size is small and increases as the video packet size is increased.

The authors conclude that by increasing the audio codec packetisation interval, the channel capacity efficiency is increased by decreasing the number of packets generated per second; thereby reducing the number of times the larger protocol overhead is incurred.

In [62] *Garg* and *Kapes* study the inherent limitations of the 802.11a/b distributed co-ordinated function (DCF) in supporting VoIP calls. In particular, the authors look at the upper bound on the number of simultaneous VoIP calls that can be placed on a single 802.11a/b access point. The upper bound is determined by increasing the number of new callers until the quality of the network falls below a certain level. They determine that the upper bound is a function of the choice of VoIP codec and the length of the audio payload.

The experimental set-up involved a number of wireless enabled PCs connected to an 802.11b access point which in turn was connected to a 100 Mbps LAN. Full duplex calls were made between a wireless PC and a wired PC using IP phones using ITU G.711 A-Law codec with 10ms frame size. As well as this the paper investigates the maximum number of VoIP calls through calculation for G.711, G.729 and G.723 audio codecs with different frame durations. This analysis included the overhead incurred from protocol overhead.

The paper states that packet loss ratio, round trip delay, and jitter were used as performance metrics, however they fail to mention what value of these performance metrics quantify an acceptable or un-acceptable level of quality. Their results found that the network could maintain 6 simultaneous calls under these circumstances. However, they found that when a seventh call was placed all calls on the downlink (i.e. wired to wireless) suffered a packet loss rate of approximately 16% and therefore stated a maximum of 6 calls could be maintained. These figures were closely matched by calculation. Calculated values showed that the network could support either 6, 12 or 17 calls when using the G.711 codec with frame sizes of 10ms, 20ms and 30ms respectively. Similar calculations were carried out for low bit rate codecs (G.729 and G.723) which yielded even larger figures. The study also included the effects of spatial distribution of 802.11b clients.

The paper concludes that if a client is not receiving full signal strength (either due to interference or distance from the access point) they will not enjoy the same quality as other users. They suggest that the physical location of an access point is crucial as far as supporting VoIP connections is concerned. They also state that given the choice of payload size in IP phones, 802.11b access points are not capable of supporting a large number of VoIP calls. Furthermore, they also state that 802.11b channel inefficiency at smaller frame sizes limits the number of VoIP calls (The authors base the channel efficiency on the ratio of time taken to transmit the payload to the time taken to transmit the payload plus the sum of any additional overheads). The authors then state that a larger payload per frame will increase the number of VoIP capable of being sustained. In contrast to other reports, they also state that codec selection does not contribute much to 802.11 networks (in comparison to Ethernet). This would only be apparent when comparing the codecs G.729 and G.723.

In [63] product specifications and practical implications as well as security issues are outlined for a standard VoIP industry deployment. The Netlink system is deployed over an IEEE 802.11b infrastructure using H.323 for call set up. The system utilizes a proprietary protocol known as Spectralink Voice Priority (SVP). The SVP Server is a dedicated network appliance that works with the wireless LAN access points to guarantee QoS by utilizing the protocol. It is intended that the network deployment will be solely used for VoIP and as such the SVP protocol ensures that clients accessing the medium for voice calls will be given priority. SVP can then be replaced with IEEE 802.11e when the standard has been ratified.

This system is said to be capable of supporting up to 16 users and up to 8 simultaneous calls. Nevertheless, they point out that the number of simultaneous calls supported by a single access point depends on the data rate, audio codec, protocol overhead, and performance of a specific manufacturer's AP. They further state that in general, approximately twelve simultaneous calls are possible if all twelve are operating at 11 Mbps. The authors further recognize that, depending on the location of the client to the access point, their IP phones may be operating at a lower data rate.

The document also addresses bandwidth utilization for VoIP calls and states that when a call is in progress the bandwidth is a function of the transmission data rate and the type of voice encoding used. The document gives an example of bandwidth utilization for a call using a G.711 codec at a sample rate of 64 kbps. According to their calculation such a call will use about 4.5% of the AP bandwidth at 11 Mbps and about 12% at 2 Mbps. They then stress that only 70% of the access point bandwidth is available for clients, as 30% of the bandwidth is needed for retransmissions and channel allocation.

3.4.2 Summary

The majority of research publications in VoIP over WLAN has involved using simulated speech. In [54] the number of VoIP users supported drops from 10 to 5 when background traffic is used. These figures are based on a G.711 codec with a payload of 80 bytes. In [61] a simulation of an 802.11b network showed that 30 calls could be supported when using a G.711 codec with a payload of 240 bytes. From [61] and [62] it can be shown that increasing payload greatly increases channel capacity efficiency. Findings from [56] and [63] point out that protocol overhead and back-off delay limits the capacity of the network, stating that the back-off time is 20 times larger than payload transmission time. The main findings from [54], [55], [56], and [61] state that the AP downlink will ultimately dictate the capacity of the network. The research shows that the AP acts as a bottleneck and would require greater access to the medium in order to increase the number of VoIP calls capable of being supported. Finally, [62] states the importance of the physical location of the AP. According to their work the proximity of a wireless station to the AP will influence the QoS.

3.5 Types of Delay

End-to-end delay comprises several contributing factors, the sum of which constitute the overall delay endured. A definitive guide is provided in [64] on the sources of delay as well as its effects of end-to-end one-way delay. According to [51] one-way delay has three components:

- Encoding/decoding/packetisation & jitter buffer delay (delay variation),
- Transport delay,
- Propagation delay.

3.5.1 Packetisation Delay

Packetisation delay in a codec is comprised of several components. On the encoder side there is a delay for the time taken to sample speech and then build a frame. If the particular codec uses compression it will require some time to carry this out before it inserts the speech frame into a packet. Then there is the time taken to transfer the packet to transport mechanism including any hardware delays. As well as this some voice codecs (vocoders) use a look-ahead function on order to gain information about how to compress the signal. This requirement adds additional waiting time to the compression process. The opposite process occurs at the other end on the decoder and as such delays are incurred at each stage of decompression and reconstruction. Furthermore, some codecs have an add-on packet loss concealment algorithm that creates additional delay.

In [64] guidelines are provided on calculating the maximum and minimum codec related delays. There is always a trade-off between header-to-payload efficiency and packetisation delay. However, it is clear that as more speech frames and larger speech frames are inserted into each packet, the packetisation delay increases [51].

3.5.2 Jitter Buffer Delay

If a jitter buffer is used on the decoder side this will have an associated delay. The function of the jitter buffer is to compensate for the variation in arrival time of sequential packets from the transport facility. Although this may lessen the effects of delay variation and help improve QoS, it adds to the overall end-to-end delay. Frame based jitter buffers can increase delay dramatically if the frame size is large [51].

3.5.3 Transport Delay

This comprises a fixed and variable delay. There is a fixed delay related to the time required to send a packet onto the network interface. The time required to achieve this is a function of the packet size and link throughput. Other delays encountered here are due to the medium and the equipment used on it. There will be a fixed delay for propagation of packets as they traverse the medium. This type of delay is related to the distance a packet has to travel. Variable delays are also incurred due to buffering and queuing delays in switches. Such delays can increase jitter across a network and can be difficult to predict [51].

3.5.4 Example of a Delay Budget

Figure 3.5 illustrates the various components and associated delays to establish a VoIP call.

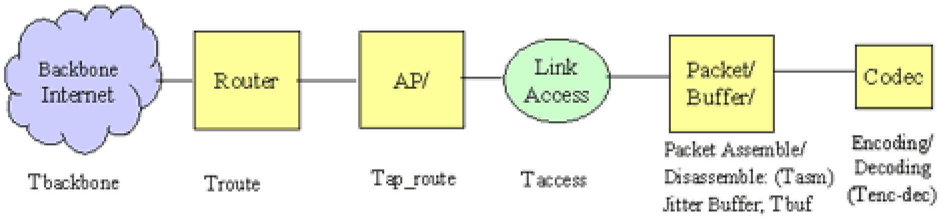


Figure 3.5: Example of VoIP connection [65].

This is a typical scenario which may have a delay budget as shown in Table 3.3 below when using the G.711 or G.726 codec.

Round Trip Delay Source	Delay Symbol	G.711 (5msec)	G.726 (10msec)
Encode / Decode	Tenc_dec	5.5	5.5
Assemble / Disassemble	Tasm	10	20
Jitter Buffer (1)	Tbuf	10	20
WLAN Access Delay	Taccess	10	20
Access Point Routing	Tap_route	10	10
Enterprise Routing	Trouting	5	5
Backbone Delay (residual)	Tbackbone	**99.5**	**69.5**
	Total	**150**	**150**

Table 3.3: Example of delay budget when using G.711 and G.726 [65].

In [66] a similar delay budget calculation is carried out for VoIP in a LAN for a single hop, however they allow 18ms for encoding/decoding. This figure also includes look-ahead time. Taking this into account the figures approximately match up as far as the AP routing is concerned, with a difference of about 3ms. These values for delay budget can be used as a reference when measuring the AP transit time. By comparing our measured values to tabulated examples of delay times we can indicate an upper limit on the delay before perceived QoS is reduced.

4 Framework for Analysis

The basic aim of the analysis was to incrementally increase the number of VoIP calls on an IEEE 802.11b network, operating in infrastructure mode, until the network had reached saturation point (i.e. no further calls could be supported). During these tests all MAC bandwidth data was recorded as well as the AP transit delay.

4.1 Test Bed Scenarios

The experimental test bed comprised 16 wireless enabled desktop PCs, all capable of operating under *Windows 2000* or *Linux Fedora Core 3*. VoIP calls were established under the ITU-T H.323 protocol using the G.711 codec at 64 kbps with a payload of 240 bytes. Simultaneous VoIP calls were established via a Cisco AP1200 Access Point. Data was captured using a laptop running a modified version of the WLAN Probe [52] over a 10 minute period with a bandwidth sampling interval that varied between 1 and 10 seconds depending on the test (generally speaking the bulk of the tests used a 10 second sampling period).

The first set of tests involved recording data for N VoIP STAs over a t-seconds test duration. N was increased from $N = 2$ in steps of 2 up until $N = 16$, (i.e. the pairs of VoIP STAs switched on). The next set of tests set $N = 14$ and recorded over $t = 600$ seconds. These tests also included background data in the form of the UDP traffic generator (TG) MGEN [67]. These tests can be further divided into two groups.

- Traffic generator operating from WLAN to LAN
- Traffic generator operating from LAN to WLAN

This involved changing how the traffic source and traffic sink connected to the medium (i.e. using the wired or wireless network interface). In doing this it was possible to distinguish between the effects of traffic on the downlink (from the AP to a wireless STA) and the uplink (from the wireless STA to the AP). This enabled us to study the properties of the MAC contention mechanism in the AP and to control the arrivals to the buffer in the AP. Figure 4.1 illustrates the general experimental set up. All test equipment was situated in a room measuring L 5.85m x W 4.37m x H 2.68m. Eight wireless STAs were located under a central table (L 2.44m x 1.22m x H 0.92m) with the remainder located around a perimeter workbench (H 0.9m).

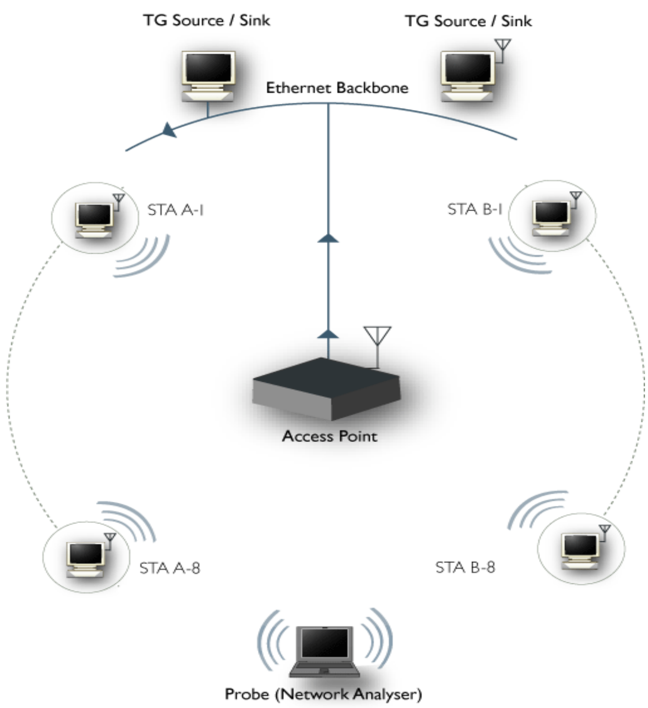


Figure 4.1: Experimental Set-up.

The traffic generator enabled control over the type of traffic transmitted across the network. The toolset generates real-time traffic patterns so that the network can be loaded and stressed in a variety of ways. In this case UDP traffic was generated with a payload of 128 bytes, 256 bytes and 512 bytes (to simulate light, medium, and heavy background traffic). The traffic generator initially produced packets at an average rate of 100 pps which was increased for each subsequent test by 100 pps to a maximum of 1000 pps. Packets were generated according to a Poisson distribution, i.e. with exponentially distributed inter-packet times.

In Figures 4.2 (a) and (b) we illustrate the significance in the direction of UDP traffic through the AP. Notice the difference in traffic flow for the TG packets in each diagram. By changing the direction of the TG flow it is possible to alter the nature of the AP buffer occupancy. As packets enter the AP buffer it fills up. When traffic from the TG is flowing from wired-to-wireless, additional data packets will fill up the buffer along with the voice packets. Depending on the rate and TG payload the buffer may fill up very quickly. The effects of this are twofold; received packets may suffer large delays, the buffer may overflow and packets will be lost.

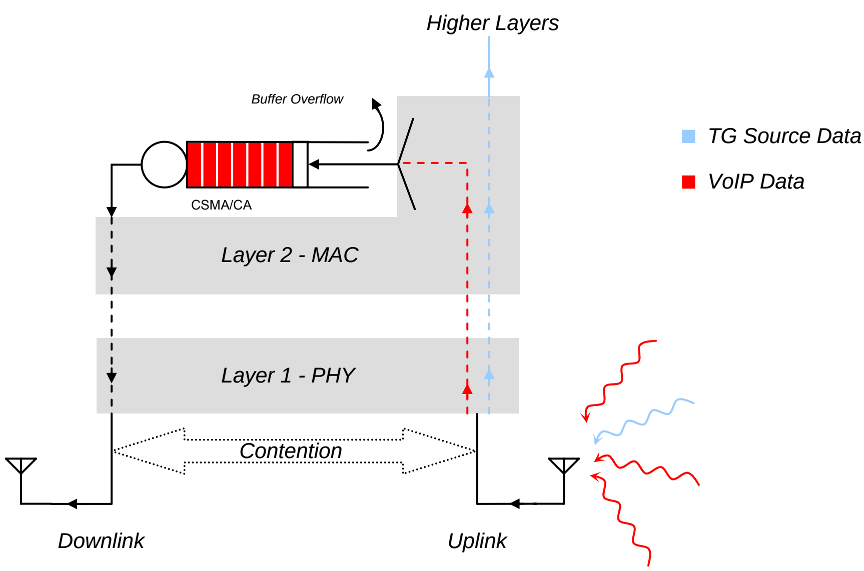


Figure 4.2(a): AP with TG packets. Buffer contains voice only data packets.

In Figure 4.2(a) the AP buffer should contain VoIP packets only. However, the service rate for the AP buffer will be reduced owing to the increased uplink TG traffic. Contention from the uplink traffic will reduce transmission opportunities for the downlink of the AP and hence reduce the buffer service rate.

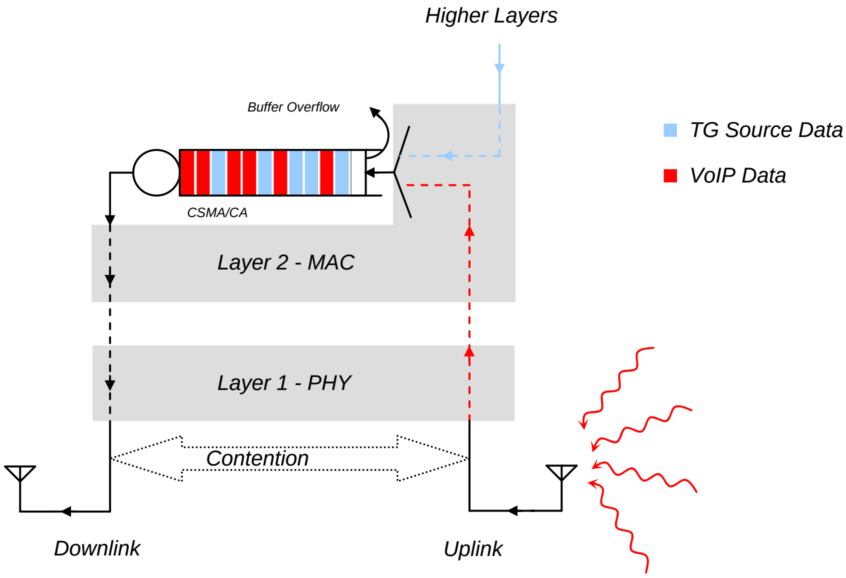


Figure 4.2(b): AP with TG packets. Buffer contains voice & TG packets.

In Figure 4.2(b) The AP buffer will contain both VoIP as well as TG packets. This results in an increased AP buffer occupancy and hence the possibility of packets being lost. However, this time the TG does not compete against uplink traffic for access to the medium.

In order to control the AP buffer dynamics we must take into account two separate criteria, namely,

- Arrival rate X
- Service rate Y

In section 2.9 we discussed how for stability we require that,

$$\mathbf{E}[Y] \geq \mathbf{E}[X] \quad [4.1]$$

Through our experimentation we attempt to manage the arrival rate X and service rate Y via the downlink and uplink traffic respectively.

4.1.1 Hardware

A total of 17 desktop PCs were used for experiments, one of which was a dedicated PC for capturing data from the wireless network (i.e. hosting the WLAN Probe). Table 4.1 outlines the equipment used. All PCs had a dual boot operating system. PCs used as VoIP STAs operated under *Windows 2000 SP4*. When the TG was in use, STAs 15 and 16 operated under *Linux Fedora Core 3* and utilized the traffic generator MGEN [67]. As discussed above either STA 15 or 16 operated as a source or a sink for the TG. All desktops were fitted with Buffalo WLI-PCI-OP WLAN PCI adapters in order to interface between the PCMCIA NIC and the desktop. The majority of PCs use Netgear dual band wireless cards utilizing the Atheros chip set. In addition to this, a pair of STAs (9 and 10) communicated via Belkin 802.11b wireless cards and one pair of STAs (11 and 12) used one of each card. The reason for this was to observe any differences in wireless cards. In addition to the equipment listed below initial set-up made use of *Wild Packets Airopeek NX* [41] and *Ethereal* [42] network analysers in order to verify capture results.

Name	Model	Specification	Operating System	Radio	Reference
STA 1	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker A1
STA 2	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B1
STA 3	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker A2
STA 4	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B2
STA 5	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker A3
STA 6	Dell Optiplex GX260	2.4 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B3
STA 7	Dell Optiplex GX150	1.0 GHz P3	W2K & Linux FC 3	Netgear WAG511	Talker A4
STA 8	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B4
STA 9	Dell Optiplex GX150	1.7 GHz P3	W2K & Linux FC 3	Belkin F5D6020	Talker A5
STA 10	Dell Optiplex GX150	1.7 GHz P4	W2K & Linux FC 3	Belkin F5D6020	Talker B5
STA 11	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Belkin F5D6020	Talker A6
STA 12	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B6
STA 13	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker A7
STA 14	Dell Optiplex GX240	1.7 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B7
STA 15 /Source/Sink	Dell Optiplex GX260	2.4 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker A8 /Source/Sink
STA 16 /Sink/Source	Dell Optiplex GX260	2.4 GHz P4	W2K & Linux FC 3	Netgear WAG511	Talker B8 /Sink/Source
AP	Cisco AP 1200		Firmware V 1.22	2.4GHz 802.11b	AP
Probe	Dell Latitude D610	2.0GHz P M	Linux FC 3	Allied Telesyn AT-WR2411	Net. Analyser

Table 4.1: Test-bed equipment.

4.1.2 Speech Sample

Section 3.1.3 discussed the lack of a commonly agreed-upon set of speech samples, however guidelines are provided for desirable characteristics of the sample [43]. In this instance, the original source sample consists of two male talkers participating in a mutual conversation recorded off a radio show. The conversation was captured and recorded using *Creative Labs Wave Studio* (WS) with characteristics as shown in Table 4.2 below.

Recording Characteristics	
Format:	PCM Wave
Sample Rate:	44.1 kHz
Sample Size (Resolution):	16 bit
Audio Channels:	2 (Stereo)
Duration:	≈ 90 seconds

Table 4.2: Sampled conversation recoding characteristics.

After the conversation was captured it was necessary to separate the individual talkers. By using WS it is possible to view the complete capture (over time or samples) for a selected sample. The software allows us to zoom in to a higher resolution and view the sample in fine detail. In order to extract the signal it is necessary to listen to the playback to gauge were each talker starts and finishes their dialogue. At this point we can increase the resolution in order to view the signal on a smaller scale and identify exactly when the signal (i.e. the speech) begins or ends. A rough edit allows us to cut out the sections of each talker into individual components, leaving silent sections in each talker's speech sample. The next step was to identify instances of double-talk (i.e. instances when both parties talked at the same time). This is a more complex and timely process and involves using filtering tools in WS to identify one talker over another. The end result proved to be satisfactory, with the characteristics of both samples closely matching each other (as can be seen from the duration of each talker's conversation as well as the individual file size) as well as displaying the required characteristics as outlined in [43] and recommendations in [44]. Figure 4.3 displays a section of each speech sample over time. As can be clearly seen the two samples are interleaved with each other as well as containing a density of approximately 70% speech with 30% pauses. The figure also displays periods of mutual silence and double-talk.

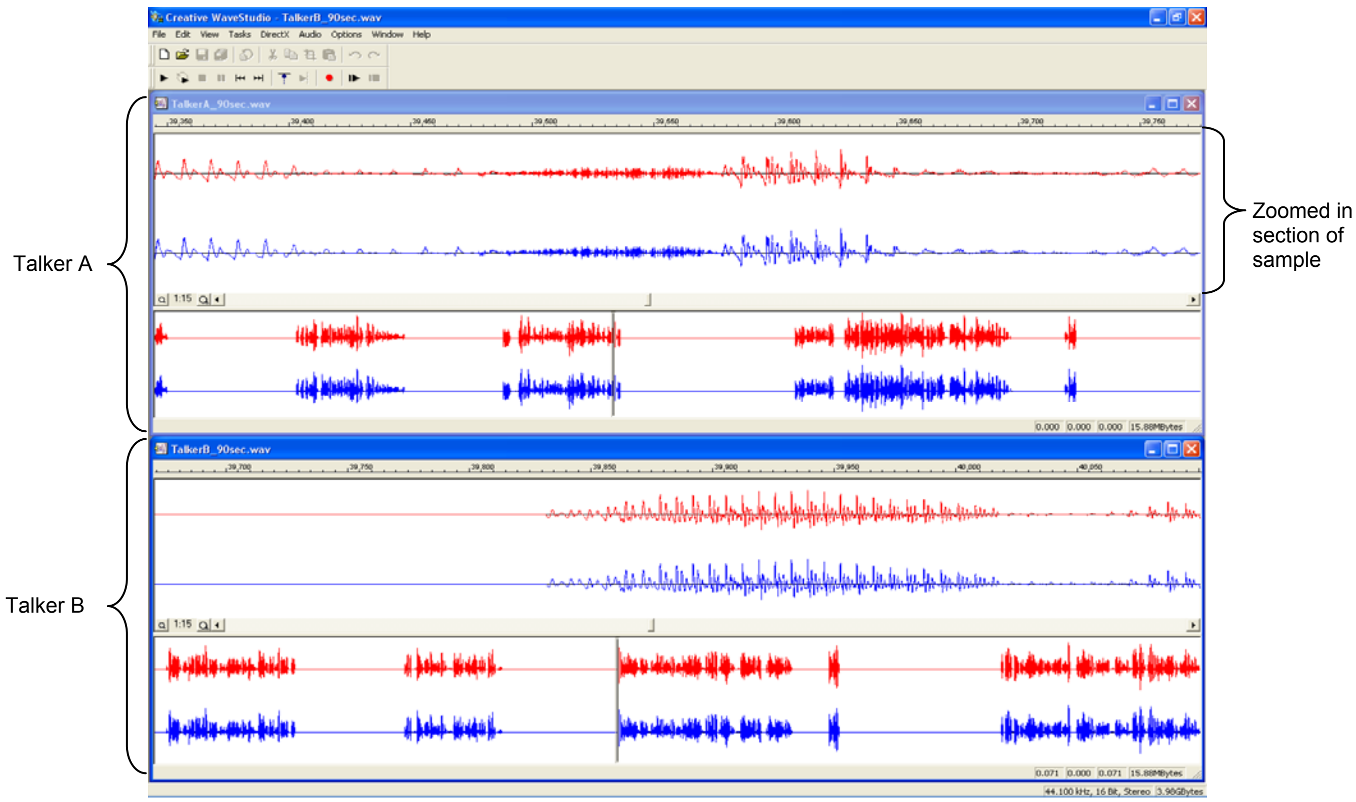


Figure 4.3: Audio spectrum of speech sample.

4.1.3 Packet Capture

Packet capture and analysis occurs over a series of stages, from the initial capture stage to processing data, formatting data, and then generating plots. Packet capture is made possible by the use of a modified WLAN Probe. Once the packets have been captured they are passed over to the WLAN Probe for processing in order to determine the MAC bandwidth components. This generates a graphical display of the network activity in real-time. The WLAN Probe also generates output files so that post capture analysis can be made possible. These files need to be formatted in such a way so as to allow them to be displayed as meaningful results. The files go through a series of processes in order to be displayed and read in by a graphing application.

4.2 Wireless Probe and Packet Capture

The network analyser passively "sniffs" the wireless medium to obtain and capture wireless packets. In order to do this a PC is set up with an 802.11b network interface card capable of operating in promiscuous mode. This basically means that the card can receive data and will not interfere with traffic on the medium. At the heart of the Probe is the open source packet capture library *libpcap*. *Libpcap* operates under *Linux* (although a *Windows* version is available "*winpcap*") and essentially acts as a hardware interface between the wireless LAN card and the operating system.

Once a data packet has been captured, it is then processed by the Probe along with the captured packet's associated protocol headers as well as the capture time. (Note: this is not the same as the RTP time stamp). The Probe calculates the time intervals (as outlined in Section 3.3) based on the MAC header information as well as the transmission rate. The network busy interval is calculated based on the frame type and size as well as the transmission rate. By analysing the MAC header it is possible to determine the sender of each frame. From this it is possible to distinguish between different stations and then determine the intervals for T_{Load}, T_{Access} and T_{Free} and consequently the associated MAC bandwidth components (namely BW_{Load},

BW_{Access} and BW_{Free} respectively). The frame count is also used in order to calculate the percentage packet loss over a given period. This data is also saved to a file for each capture time to allow for post capture analysis.

In order to calculate the delay through the access point it is necessary to make some modifications to the Probe. These modifications were necessary in order to allow us to distinguish between packets going to the access point (Uplink) and those coming from the access point (Downlink). This is achieved by checking the To DS and From DS flag (see Section 2.7.5) in the MAC header. It was decided that the modified Probe would capture all data in raw format as real-time processing could prove to be extremely processor intensive, given the possible number of stations transmitting and total number of packets involved. If a frame was of type "Data" it was output to one of two files depending on how its To DS / From DS flag was set. These output files contained all data packets as they were captured. The file output was structured with the following information;

To DS, From DS, Source Port, Dest. Port, Address 1, Address 2, Address 3, MAC Seq. No., RTP Seq. No., Capture Time

As mentioned previously in Section 2.7.5, the To DS and From DS flags indicate the direction of the traffic and hence which file the capture packet is to be saved to. They also serve another purpose. The condition of these flags will indicate how the MAC address fields are set (i.e. source address (SA), destination address (DA) and basic service set identification (BSSID)). Assuming the infrastructure mode architecture, the address fields are set according to the following table.

Function	ToDS	FromDS	Address 1 (receiver)	Address 2 (transmitter)	Address 3	Address 4
To AP	1	0	BSSID	SA	DA	Not used
From AP	0	1	DA	BSSID	SA	

Table 4.3: MAC address field settings.

4.3 Post Capture Analysis

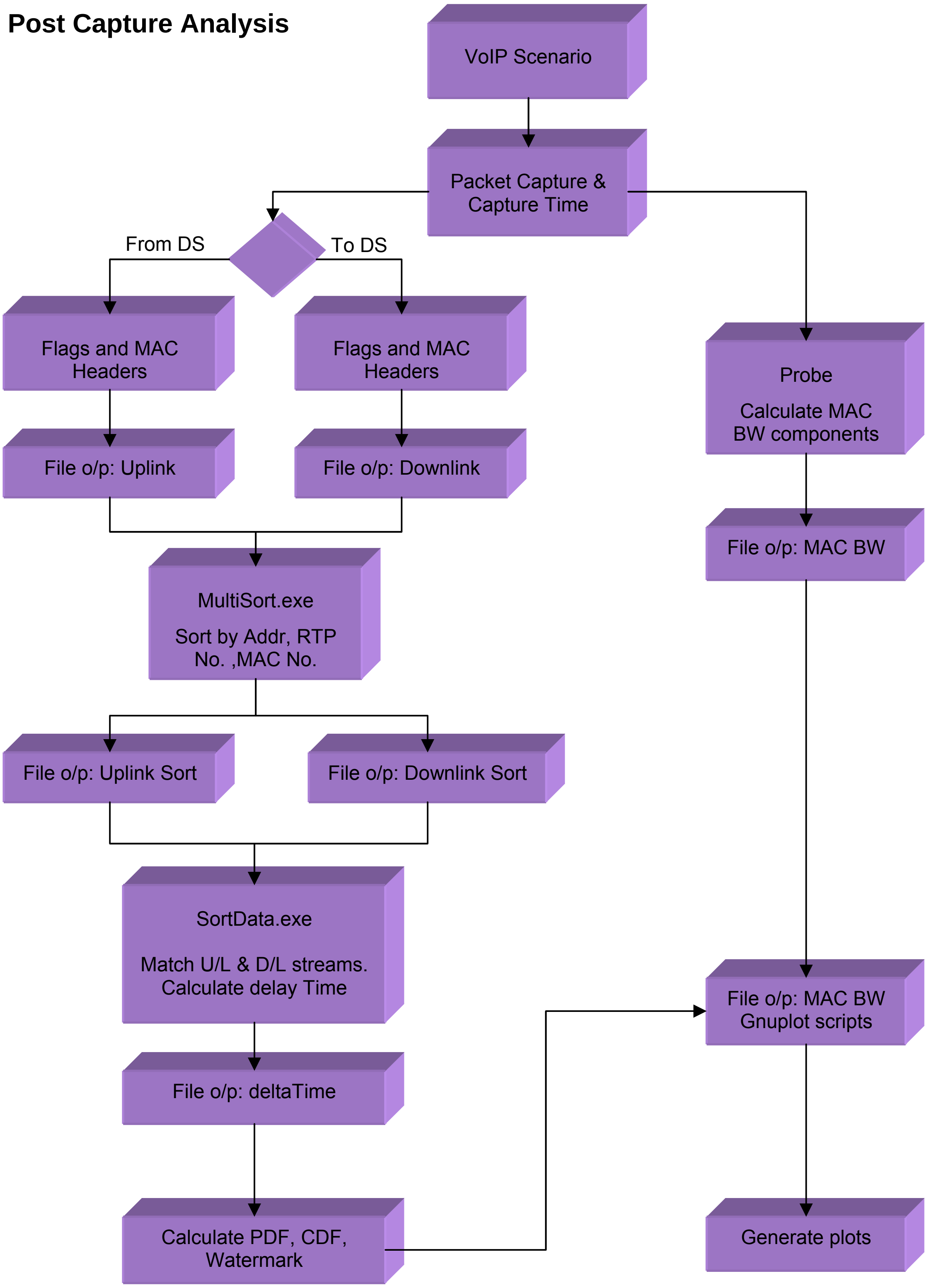


Figure 4.4: Capture and analysis flow chart.

4.3.1 Sorting Data

After the modified WLAN Probe had recorded the transmission data and created all the output files it was necessary to sort the data before processing in order to improve the efficiency of the next stage. A short program enables the uplink and downlink files to be read in and sorted according to the following criteria; sort files according to *source address* first, then by *RTP sequence number* and finally by *MAC sequence number*. The output of this program consisted of two separate files (an uplink and downlink file) containing all the data of the input files, the only difference being that the data had been re-organised. The input files were in comma separated value (csv) format. The approximate file size for each of the uplink and downlink files for a ten-minute capture was between 10 Mb and 100 Mb (This figure greatly depended on whether the TG was sending packets on the wireless uplink or downlink). Processing and re-ordering these files took approximately ten minutes on a 2.7 GHz Pentium machine. At this point the data was ready for processing by the next stage in order to match packets.

4.3.2 Identifying Packets

The data at this point was parsed into two separate (and somewhat large) text files. It was now necessary to match the packets in each file in order to calculate the delay time through the access point. When a packet leaves a wireless station its RTP header contains a sequence number. Also its MAC header will contain a sequence number which is incremented for every transmitted packet. The RTP sequence number is only incremented when a packet has been successfully transmitted and acknowledged. Therefore the

RTP sequence number will remain the same on the uplink and downlink stream for the initial transmission of a packet as well as any subsequent retransmissions of that particular packet (if of course the retransmission limit is set greater than one).

If one were to compare uplink and downlink packets based on RTP sequence number alone you may experience multiple packets with the same sequence number on both the uplink and downlink. (Remember, the modified Probe captures all packets when outputting to the “To DS” and “From DS” files, regardless of whether they are successful or not). To counter this we make use of the MAC sequence number. On the uplink for example, by checking the lowest value of MAC sequence number for packets with the same RTP sequence number we can be sure to have the first instance of a transmitted packet. Likewise on the downlink, by checking the highest value of MAC sequence number for packets with the same RTP sequence number we can be sure to have the last instance of a transmitted packet.

We can then simply discard any duplicate packets with the same RTP sequence number in a given stream. A short program reads in the files from the previous stage and stores them in vectors. These are similar to link lists only it is possible to traverse across various lists laterally. Each vector is ordered according to source address. Every time a source address is read in its data is stored in the associated vector. The vectors are created for new address fields. The previous stage greatly improved the efficiency of this process.

In order to match data streams, a particular packet found in the uplink file was searched for in the downlink file. First the vector with the matching source address was located then a vertical search was performed for the matching sequence number. If a matching packet was located the access point delay time was calculated as follows,

$$\textit{delta time} = \textit{downlink capture time} - \textit{uplink capture time} \tag{4.2}$$

Upon successful matching, the capture details were output to a new file (again in comma separated value file format) containing the source address, RTP sequence number, downlink capture time, uplink capture time and delta time. At this stage delay times had been calculated, however it was still necessary to perform a statistical analysis of the data on a per station basis. A copy of all source code used is included on an accompanying CD ROM.

4.3.3 Displaying Data

In order to transform the data into a meaningful format it was first necessary to separate data for each MAC address. The output of the previous stage was read into a short program which temporarily stored the data in a two-dimensional array, the rows of which were based on each MAC address. The column index served as a reference to the associated delta time while the contents of each row corresponded to the frequency of each delta time. From this information it was possible to calculate a Probability Density Function (PDF), a Cumulative Density Function (CDF) and an Inverse Cumulative Density Function (i.e. Watermark Plot). This data was then output to a new file. From here script files were written in order to read the data in a graphing application called Gnuplot in order to display the data. Gnuplot was also used in order to read in script files written for the data generated by the WLAN Probe containing the MAC bandwidth component information. A copy of all plot files is included on an accompanying CD ROM.

5 Results and Analysis

In Section 4 we discussed the experimental set-up and the process of gathering data. The experiments conducted can be divided into two main groups. Those without background traffic present and those with background traffic present. For tests without background traffic the number of pairs of VoIP stations was increased from 1 to 8. Tests involving background traffic can be further divided depending on how the traffic generator (TG) was configured. This involved connecting the TG up initially with a wireless source and a wired sink and then reversing the set-up with a wired source and a wireless sink. Approximately 500 experiments were conducted over a 4 month period which amounted to almost 6 GB of raw captured data (in comma separated value format). All plot files are included on an accompanying CD ROM.

5.1 Increasing The VoIP Stations

The following tests involved using pairs of VoIP stations engaged in a two-way conversation as outlined in Section 4.1. The number of STA pairs was increased to the maximum available of 8 (i.e. $N = 16$ stations in total). Each conversation pair utilized the same conversation (i.e. odd number STAs transmitting as "Talker A" and even number STAs transmitting as Talker B). Although the possibility of two separate parties (or more) engaged in the same conversation is highly improbable, their speech patterns would be similar over time. To help counteract this improbable event, the interval between starting times of each conversation was staggered according to the following,

$$conversation\,interval \approx \frac{conversation\,length \times 2}{N} \qquad (5.1)$$

Where N is the number of stations used during a particular test.

MAC bandwidth components were recorded and plotted against time. Each plot will normally include the access point (AP) components as well as the individual STAs. Due to the relatively large number of plots per graph it made sense to display the station BW_{Load}, BW_{Access} and BW_{Free} separately against the AP and network components. We will then isolate a section of the plot to analyse some of the results finer detail.

5.2 Results for 2 Stations Talking

5.2.1 MAC Bandwidth Components

Figure 5.1 is an example of a typical plot of the MAC bandwidth components over time. This plot illustrates the STA BW_{Load} behaviour for a single two-way VoIP conversation utilising a G.711 codec with a frame duration of 30 ms.

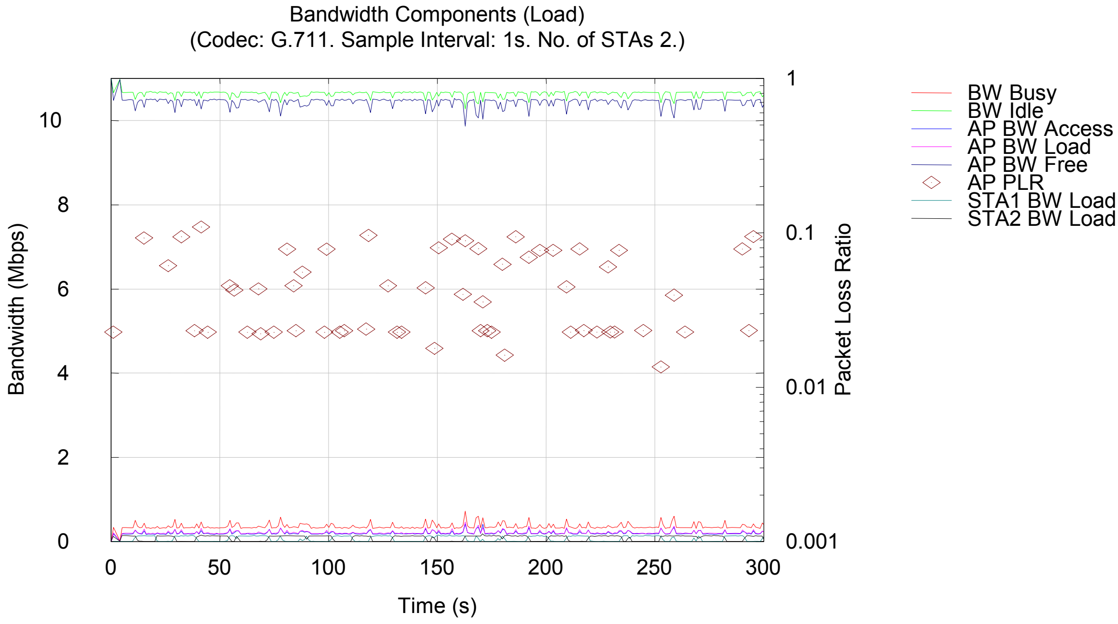


Figure 5.1: MAC BW components. BW_{Load} for STA 1 and STA 2.

The plot in Figure 5.1 can be used as a reference for how the medium will operate under light VoIP traffic. We can observe from the graph that the BW_{Busy} is quite low as expected due to only two STAs transmitting. As a result the BW_{Idle} is quite high in comparison, almost at 11 Mbps and is closely matched by the BW_{Free} as seen by the AP. If we take a closer look at the lower end of the bandwidth scale we can observe the STA BW_{Load} components in more detail. Figure 5.2 shows a 50 second section of the capture. Note how both STA 1 and STA 2 exhibit an on-off speech pattern as well as some cross over for double-talk. The average throughput for a STA when transmitting is approximately 120 kbps. We can also see that the AP BW_{Load} is approximately the sum of the two transmitting STAs at around 210 kbps. Furthermore, the BW_{Busy} has a mean value of around 360 kbps. This figure should be the sum of all BW_{Load} values not including collisions. Packet loss due to collisions will be low due to only two STAs contending. Note the packet loss ratio in the AP is sporadic and infrequent over the duration of the capture. Therefore we can say,

$$BW_{Busy} \approx BW_{Load}(AP) + (BW_{Load}(1) + BW_{Load}(2)) \tag{5.2}$$

$$\approx 210\ kbps + (240\ kbps) = 450\ kbps$$

The reason these figures do not correspond to the observed BW_{Busy} is that normally only one station is transmitting during a conversation. If we allow for only one station or use the mean value during on-off periods for all STAs we get,

$$BW_{Busy} \approx 210\ kbps + (140\ kbps) = 350\ kbps$$

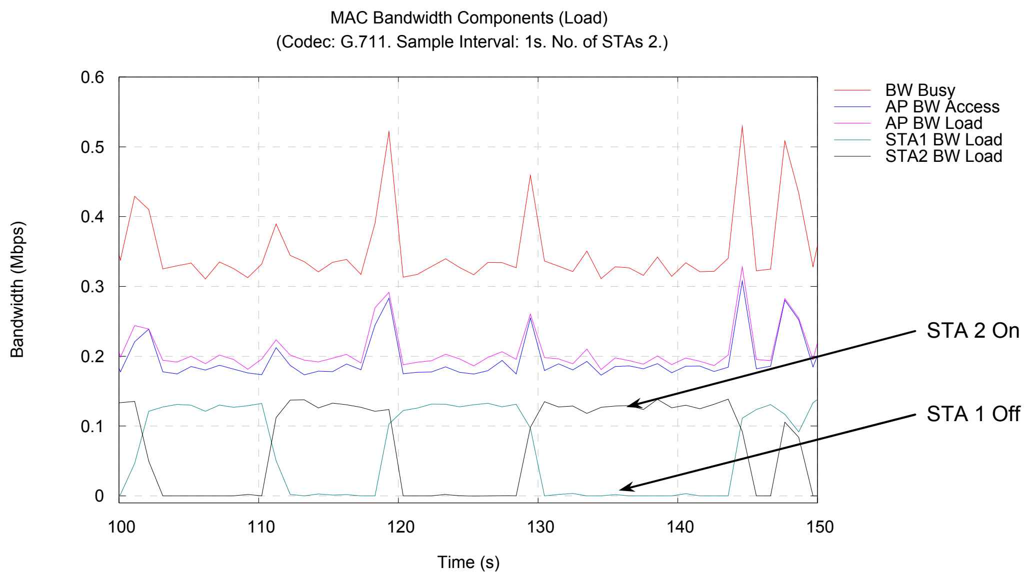


Figure 5.2: Zoomed in section of Figure 5.1.

We can also see from this graph that the AP BW_{Load} is greater than the AP BW_{Access}. We can express the access efficiency as the ratio of the BW_{Load} to the $BW_{Access.}$ Therefore by maximising the BW_{Load} and minimising the BW_{Access} we could increase the access efficiency. The graph would suggest an efficient use of the medium (however only slightly as the ratio is close to 1). Another point of interest here is the similar characteristics of the MAC bandwidth components. That is, the occurrences of high and low peaks take place simultaneously. We can see from the graph these peaks occur during cross over periods (double-talk) when one talker is finishing and the other is just beginning. During these periods both STAs are transmitting and as a result the overall throughput is increased. This also has the added effect of decreasing the BW_{Free}. This is best observed by looking at the AP BW_{Free} during the same period as before. Figure 5.3 illustrates this by observing the upper end of the bandwidth scale for the same 50-second period.

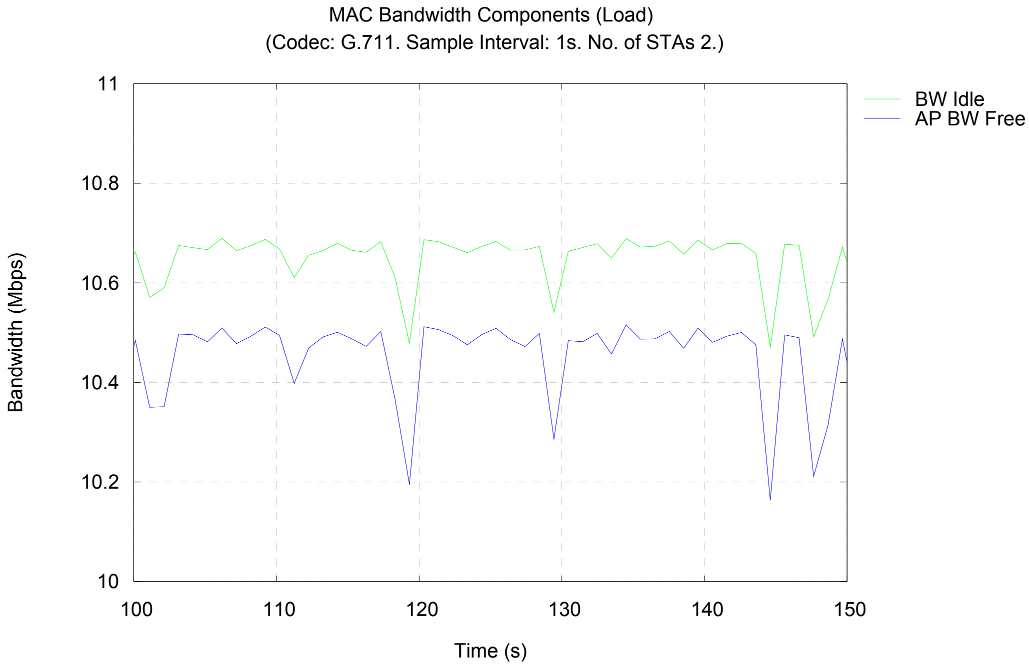


Figure 5.3: AP BW_{Free}. Dips occurring during double-talk.

Similar graphs are produced for the station access and free bandwidth components.

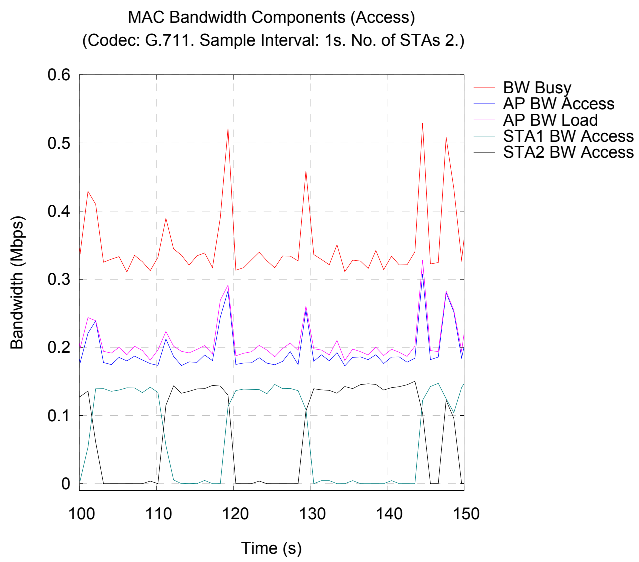


Figure 5.4(a): STA BW_{Access}.

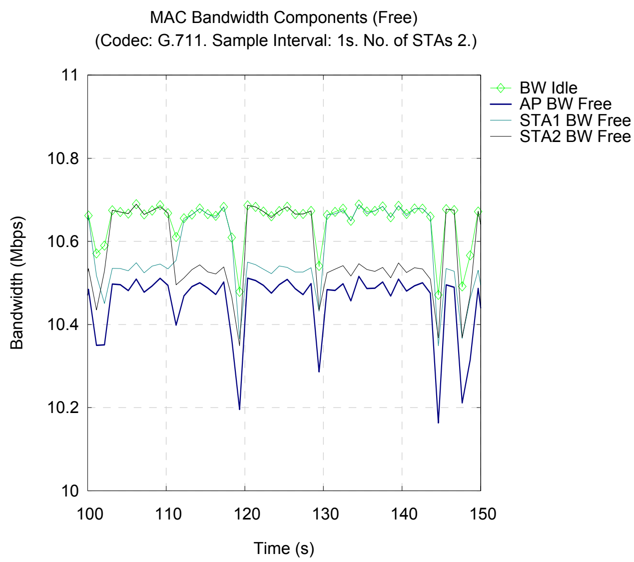


Figure 5.4(b) STA BW_{Free}

The two graphs in Figures 5.4(a) and (b) are sections for the same period as shown in Figure 5.2 and Figure 5.3. The difference being, that instead of the

STA BW_{Load} components they show STA BW_{Access} and BW_{Free} respectively. The STA BW_{Access} has a similar characteristic to the STA BW_{Load}. The STA BW_{Free} has a similar but inverse effect, in that for pronounced increases in STA BW_{Load} the STA BW_{Free} exhibits a similar pronounced decrease.

Table 5.1 outlines the main findings for the MAC bandwidth components when a single two-way VoIP conversation (using the G.711 codec with a 30 ms frame) is analysed.

	BW_{Busy} (Mbps)	BW_{Idle} (Mbps)	BW_{Access} (Mbps)			BW_{Load} (Mbps)			BW_{Free} (Mbps)		
			AP	STA1	STA2	AP	STA1	STA2	AP	STA1	STA2
Max	0.72	10.98	0.41	0.16	0.26	0.46	0.15	0.18	10.97	10.98	10.97
Min	0.02	10.28	0.01	0.00	0.00	0.01	0.00	0.00	9.87	10.14	10.14
Mean	0.36	10.64	0.20	0.07	0.09	0.21	0.06	0.08	10.44	10.57	10.56
S.D.	0.06	0.06	0.04	0.07	0.07	0.04	0.06	0.06	0.10	0.10	0.10

Table 5.1: MAC Bandwidth components for two-way VoIP on 802.11b.

It is worth noting at this stage the standard deviation (SD) and variance of the BW_{Busy} and AP BW_{Load}. The BW_{Busy} is given in Section 3.3 as,

$$BW_{Busy} = \sum_k BW_{Load}(k) - BW_{Collisions} \tag{5.3}$$

We can see from this equation that a plot of BW_{Busy} will be the aggregate of all the other load bandwidths including the AP. Similarly the AP BW_{Load} will be the sum of all the STA BW_{Load}. A comparison of these values as the number of STAs increases will be indicative of whether the network enjoys statistical multiplexing, i.e. we should observe a decrease in the standard deviation and variance as the number of STAs increase.

	Peak-to-Mean Ratio	Standard Deviation	Variance
BW_{Busy}	2.02	0.06	0.004
AP BW_{Load}	2.17	0.04	0.002

Table 5.2: SD and variance with 2 VoIP STAs in use.

5.2.2 AP Delay Characteristics

The delay characteristics through the access point may be first considered by examining the probability density function (PDF) for the delay times. This will give us an indication of the spread of delays (i.e. how they vary) and will identify the most probable delays expected. We will then look at the cumulative density function (CDF) for the delay times which will allow us to determine with some degree of certainty, what the probability of a delay occurring below a certain value will be. The PDF and CDF of STA delay times is illustrated in Figures 5.5 (a) and (b). We can see from the figures that most delays occur on the lower end of the time scale (i.e. small delays ≤ 5 ms are typically encountered).

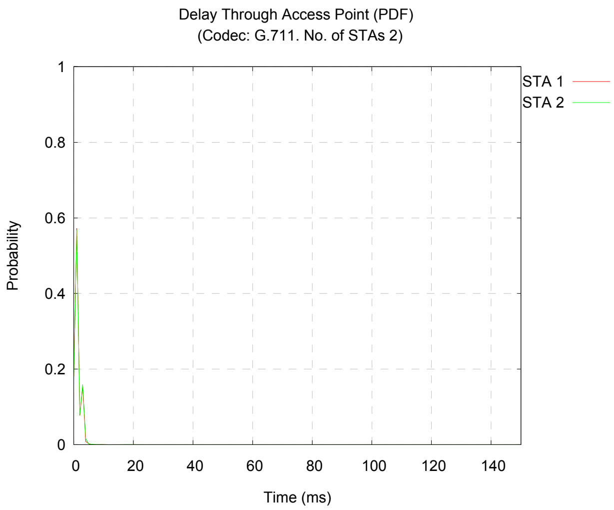


Figure 5.5(a): PDF of STA delays through AP.

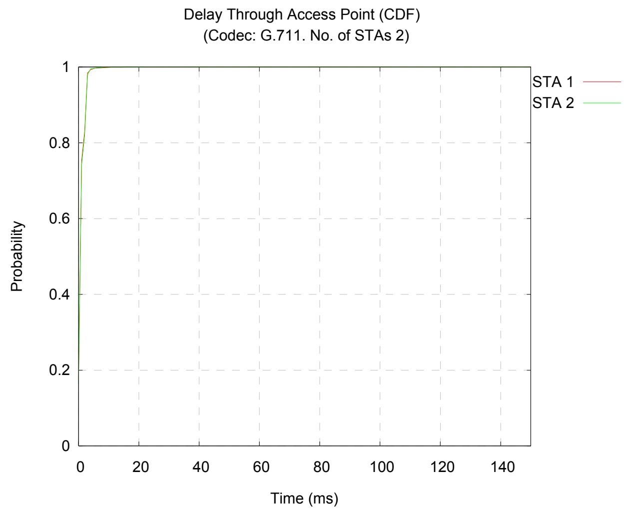


Figure 5.5(b): CDF of STA delays through AP.

We can notice that a large peak in PDF occurs below 10 ms and also from the CDF, it would appear that there are almost no delays after 20 ms. A closer look at these regions may help to clarify. Figure 5.6 shows a close up section of the PDF in Figure 5.5(a) for the region 0 – 25 ms. Likewise, Figure 5.7 illustrates the CDF for the same region.

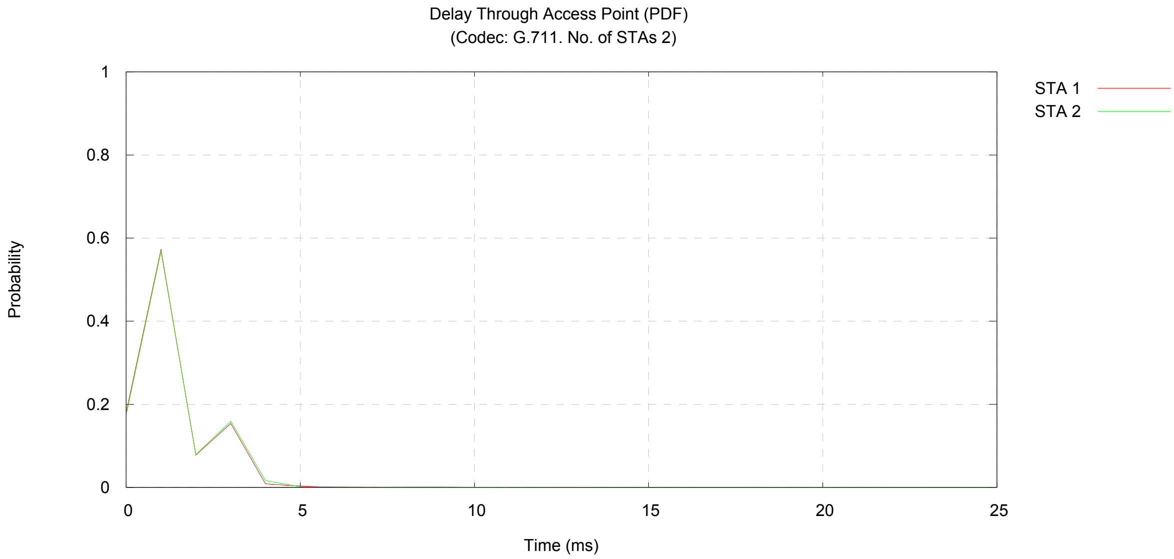


Figure 5.6. PDF showing maximum delay around 1 ms.

We can see in Figure 5.6 that under these conditions the delay for a VoIP packet through the AP is of the order of 1ms ($P_{AP\ delay}(1ms) \approx 0.57$). Furthermore, the CDF in Figure 5.7 would indicate that practically all VoIP are passed through the AP within 5 ms ($P(D \le 5ms) \approx 0.997$).

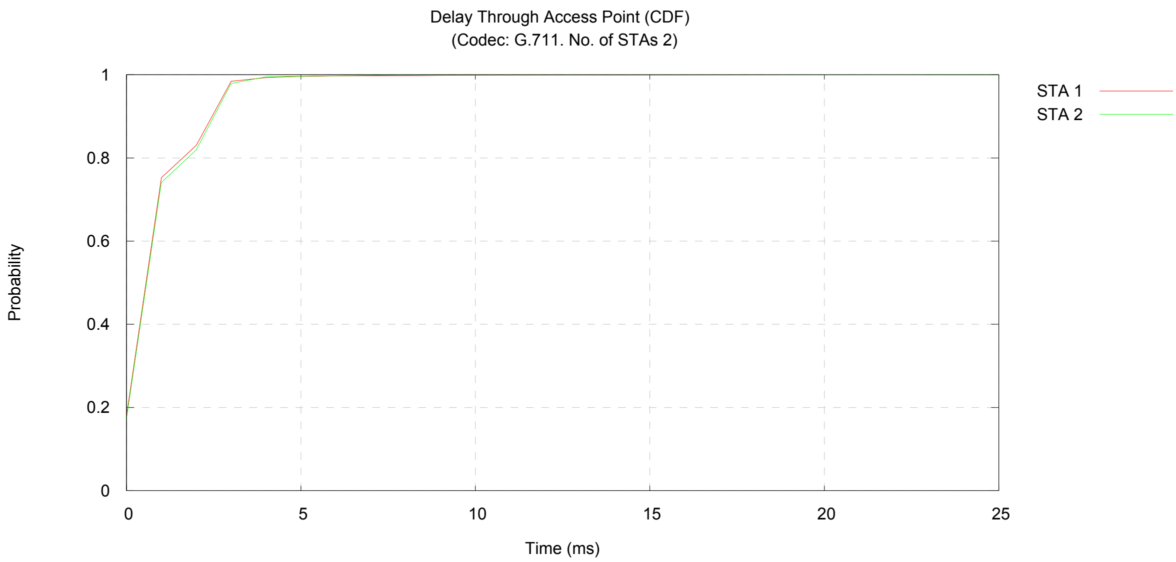


Figure 5.7: CDF of AP delay. Majority of delays are less than 5 ms.

We can see from both of these graphs that the delay characteristics exhibited by both STAs is almost identical. A summary of delay values for a single VoIP conversation is given in Table 5.3.

CDF			PDF		
Delay, d (ms)	STA 1 P(D < d)	STA 2 P(D < d)		STA 1 Peak	STA 2 Peak
1	0.75243	0.74072	Delay:	1ms	1ms
2	0.83031	0.81946	Probability:	0.57	0.57
3	0.98417	0.97878			
4	0.99304	0.99519			
5	0.99599	0.99685	Standard Deviation:	1.24	1.10
10	0.99894	0.99917	Variance:	1.53	1.22
20	0.99958	0.99983			
30	0.99979	1	Mean:	1.28	1.31
40	1	1			

Table 5.3: Comparison of values for two STA test.

Finally, if we view the delay times over the duration of the capture we may be able to observe if there is any correlation between changes in MAC bandwidth component values over time.

Figure 5.8(a): AP delays with MAC BW values.

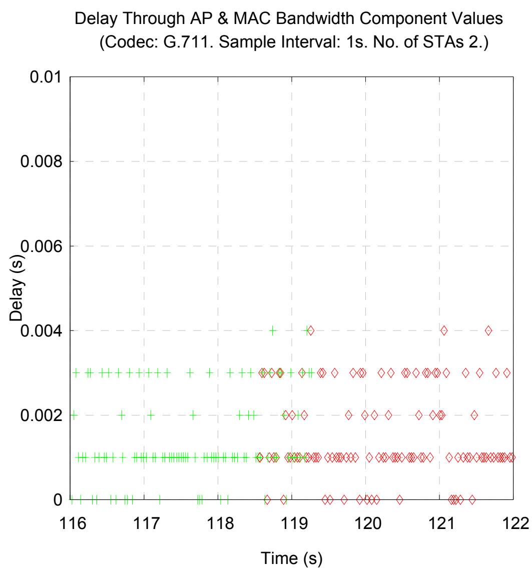


Figure 5.8(b): AP delay over time for 2 STAs.

By looking at the delay over time alongside the busy, load, and access MAC bandwidth components we can again see that the marked increase in BW_{Load} and BW_{Access} occurs during intervals of double-talk. This is indicated on the delay plot, as the on-off characteristic is clearly visible (i.e. periods of delays for one STA followed by periods of delays for another). When we take a closer look at the delays we can see that the peaks in the MAC bandwidth components occur during the transition from one STA to another.

5.2.3 Delay Distribution Tails

The plot of the probability density function gives us an insight into what happens around the mean value. In other words, it allows us to quickly see what the most likely delay will be and the extent of the spread of delays. However, we are also interested in rare events, such as the instances of very large delays that may result in buffer overflow and subsequent packet loss. These can be seen as values that occur in the tails of a distribution. Section 2.9 described a method, known as a *Watermark Plot*, for identifying the characteristics in the tails of a distribution as given in [40]. This method is basically an inverse cumulative density function (ICDF) which allows us to look at the asymptotic slope of the linear part of this plot and hence determine how the network responds to large delays. The asymptotic slope gives a measure of the rate function and the Watermark plot can be graphed by plotting;

$$\ln P(D > d)\; vs\; d \tag{5.4}$$

Or alternatively,

$$\ln(1 - CDF)\; vs\; d \tag{5.5}$$

In Figure 5.9 a typical plot of this sort can be seen. The distinct “straight part” which represents the asymptotic slope and the “wobbly” part where there is insufficient data for reliable statistics are present [40].

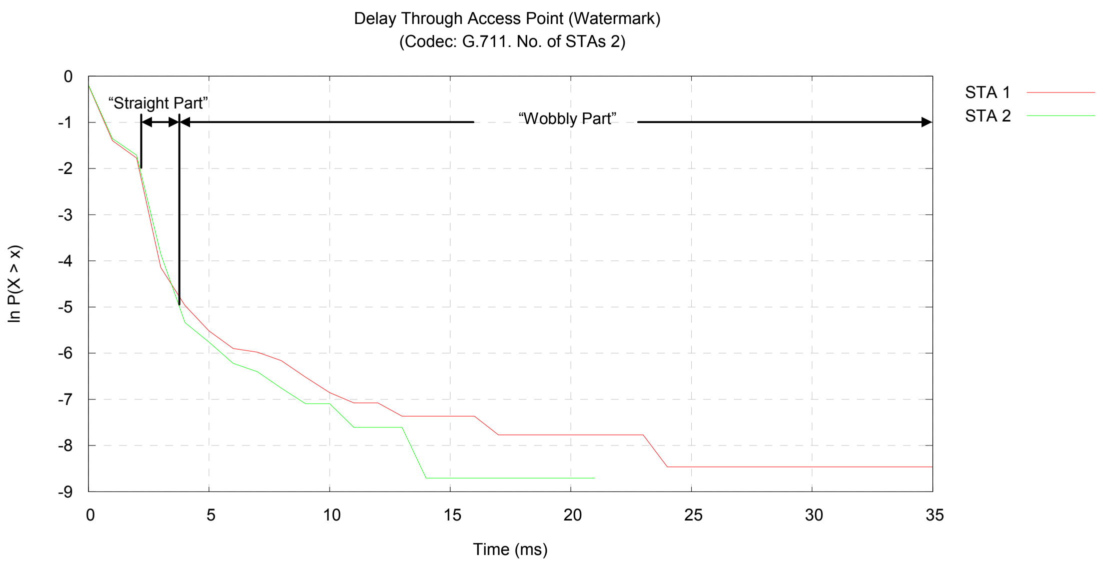


Figure 5.9: The inverse CDF for each STA.

We will later observe the difference in asymptotic slope as the number of stations increases and also as background traffic is introduced.

5.3 Results for 16 Stations Talking

5.3.1 MAC Bandwidth Components

In Figure 5.10 the BW_{Load} components for a 16 STA experiment are displayed. When we compare this plot to Figure 5.1 our attention is drawn to the increased packet loss ratio as well a significant change in the BW_{Busy}, BW_{Idle} (mean values $\approx$ 2.31 Mbps and 8.69 Mbps respectively) and AP MAC bandwidth components.

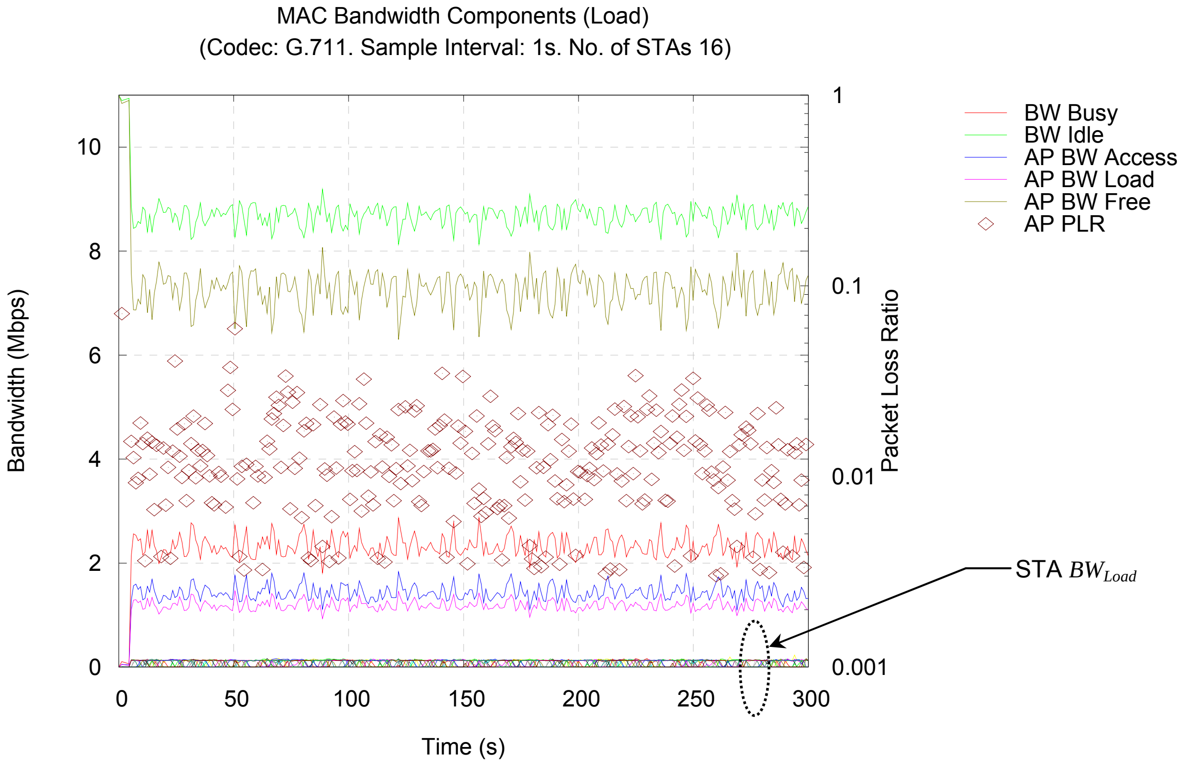


Figure 5.10: BW components for 8 conversation pairs.

The individual STA BW_{Load} characteristics are comparable to those in the 2-station test in Figure 5.2. Again when we take a closer look (Figure 5.11) at the BW_{Busy}, BW_{Access} and BW_{Load} we observe the similar large peaks. However, they now occur more frequently, on account of there being more STAs on the network and hence an increase in the amount of double-talk. Another point to note is the ratio of AP BW_{Access} to AP BW_{Load} as this will give us an indication of how efficiently the medium is being used. In Figure 5.2 we could see that the AP BW_{Load} is greater than BW_{Access}. However, in Figure 5.11 we can now observe that the opposite is true. Some of the main findings are summarised in Table 5.4.

	BW_{Busy}	BW_{Idle}	BW_{Access}	BW_{Load}	BW_{Free}
Mean	2.31	8.69	1.42	1.18	7.26
Standard Deviation	0.26	0.26	0.18	0.13	0.44

Table 5.4: Summary of WLAN and AP BW components in Mbps.

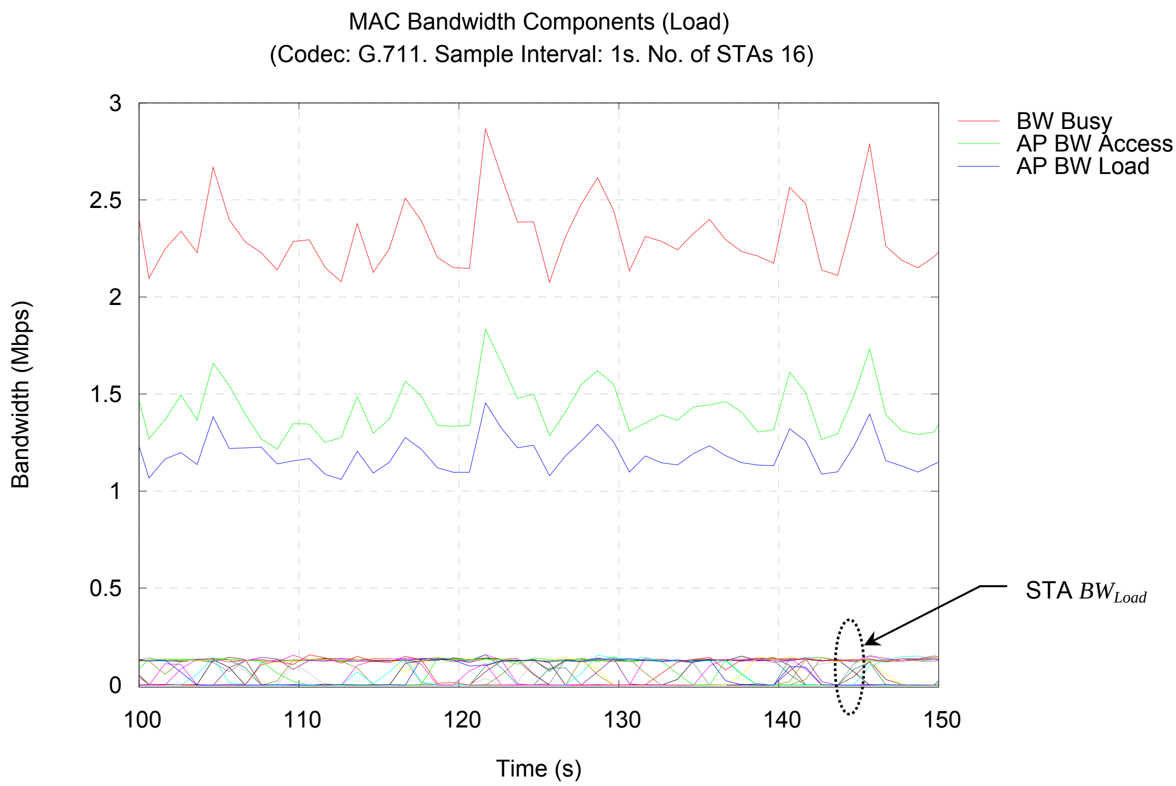


Figure 5.11: STA BW_{Load} for 16 STA test.

Figures 5.12 (a) and (b) take a close look at the access and BW_{Free} components of each STA. The BW_{Access} for each STA indicates a slight increase over the two-station test while the BW_{Free} components show a marked decrease closely matching the BW_{Idle} at approximately 8.60 Mbps. All STAs appear to see the same available BW_{Free} as expected. It is also interesting to note here that in Figure 5.12(b) we do not see the on-off pattern as displayed for the two-station test in Figure 5.4(b).

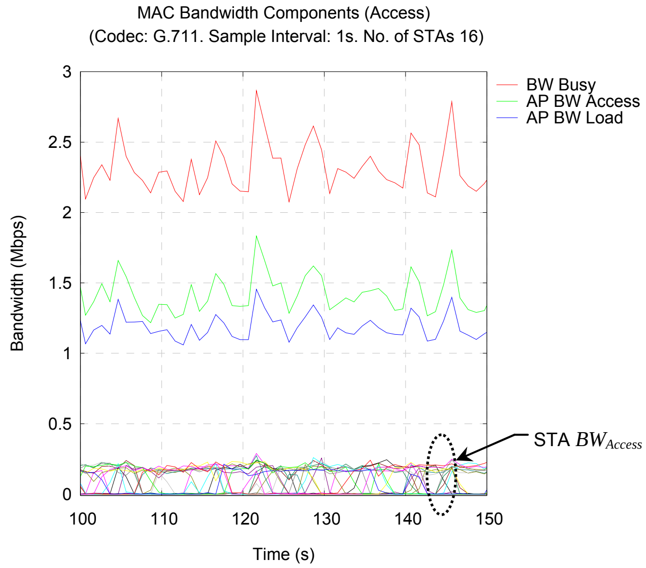


Figure 5.12 (a): BW_{Access} for 16 STA test.

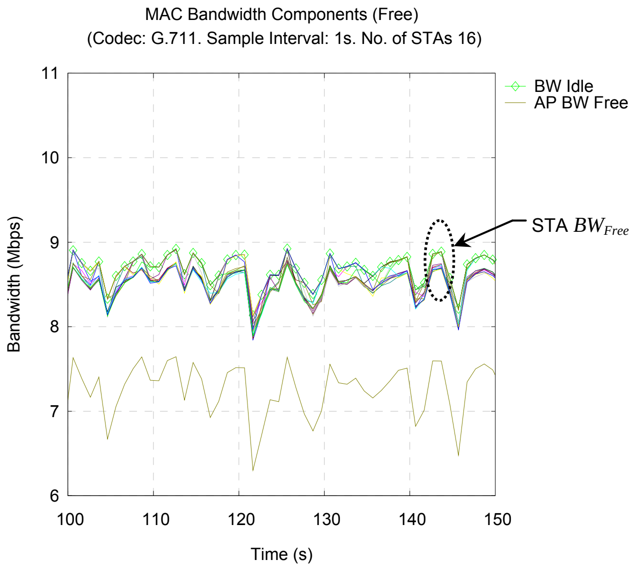


Figure 5.12 (b): BW_{Free} for 16 STA test.

As STA 1 and STA 2 were the only stations involved in all tests, we will look at the mean MAC bandwidth components of this pair over each successive test. In Figure 5.13 we can observe that the BW_{Free} available for both STAs is in the region of 10.50 Mbps, when only two STAs are transmitting. As the number of pairs is increased, the STA BW_{Free} decreases by about 0.25 Mbps for each successive pair added, until all 8 pairs are transmitting on the medium. This results in both STAs (and indeed all the other stations) indicating an available BW_{Free} of approximately 8.60 Mbps. Table 5.5 summarises the main findings for the MAC bandwidth components for STA 1 and STA 2.

	BW_{Busy} (Mbps)	BW_{Idle} (Mbps)	BW_{Access} (Mbps)			BW_{Load} (Mbps)			BW_{Free} (Mbps)		
			AP	STA1	STA2	AP	STA1	STA2	AP	STA1	STA2
Max	2.87	10.94	1.83	0.29	0.30	1.46	0.16	0.16	10.90	10.93	10.93
Min	0.06	8.13	0.04	0.00	0.00	0.03	0.00	0.00	6.30	7.84	8.00
Mean	2.31	8.69	1.43	0.09	0.12	1.18	0.06	0.08	7.26	8.60	8.57
S.D.	0.26	0.26	0.18	0.09	0.09	0.13	0.06	0.06	0.44	0.29	0.28

Table 5.5: Summary of STA 1 & STA 2 MAC BW values for 16-station test.

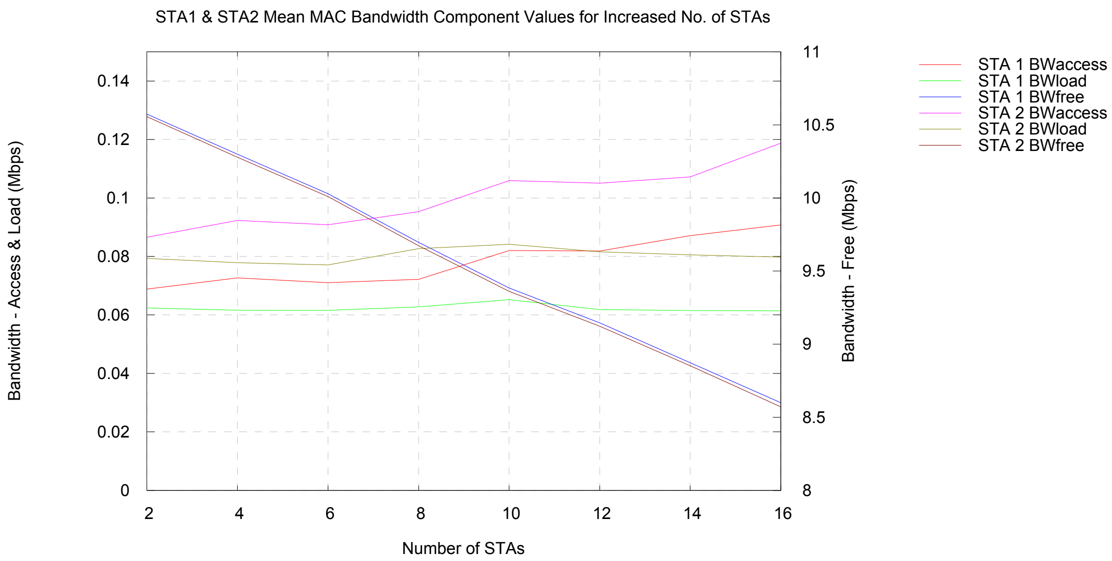


Figure 5.13: MAC BW values for STA 1 and 2 as number of STAs increase.

From Figure 5.13 it would appear that the load remains relatively constant for each STA, while the BW_{Access} shows a slight increase as a result of the increased contention, resulting in an increased number of deferrals and hence an increased bandwidth requirement. These findings are consistent with the results for the AP MAC bandwidth components as shown in Figures 5.2 and 5.11. This is further supported by Figure 5.14 which displays the standard deviation of the MAC bandwidth components as the number of station pairs is increased. We can see that the BW_{Load} remains relatively constant while the BW_{Access} and BW_{Free} show an increase and decrease respectively. We mentioned previously that the STA BW_{Free} does not exhibit the on-off characteristic in the 16-station test (Figure 5.12(b)) as seen during the 2-station test (Figure 5.4(b)). Figure 5.14 further supports this by indicating the increase in standard deviation. This essentially means that as the number of stations on the medium increases, the BW_{Free} will exhibit a greater fluctuation about the mean. So not only does the mean BW_{Free} decrease but it also has an increased range of values that experience frequent changes.

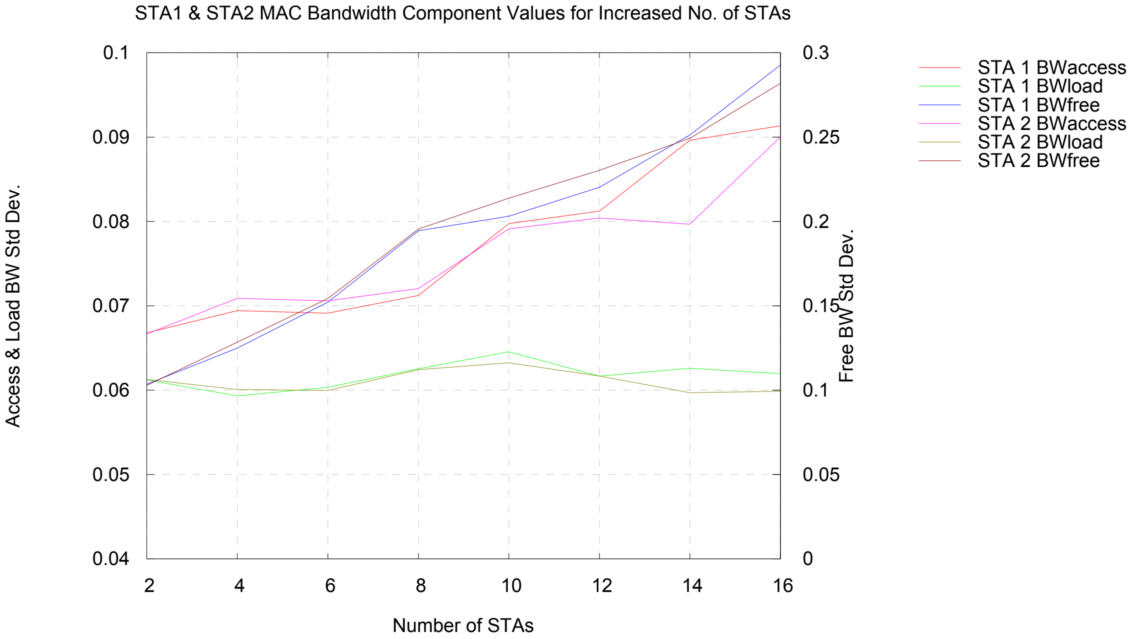


Figure 5.14: Standard deviation of MAC BW values for STA 1 and STA 2.

To round up the investigation of the MAC bandwidth components we will take look at the effects on the AP as the number of station pairs on the medium is increased (Figure 5.15). As the AP is essentially a relay for all the contending STAs, its downlink MAC bandwidth components will be the aggregate of all the station uplink components. We can clearly see that as the number of stations on the medium increases, the BW_{Idle} and BW_{Free} components decrease while the BW_{Busy}, BW_{Access} and BW_{Load} components all increase. Figure 5.15 also shows the ratio of packet losses due to collisions. We can see from the graph that as the number of stations on the medium increases there will be a greater chance of stations colliding. Collisions are comparatively low for 2-8 stations and then reach as high as 1% for 16 stations.

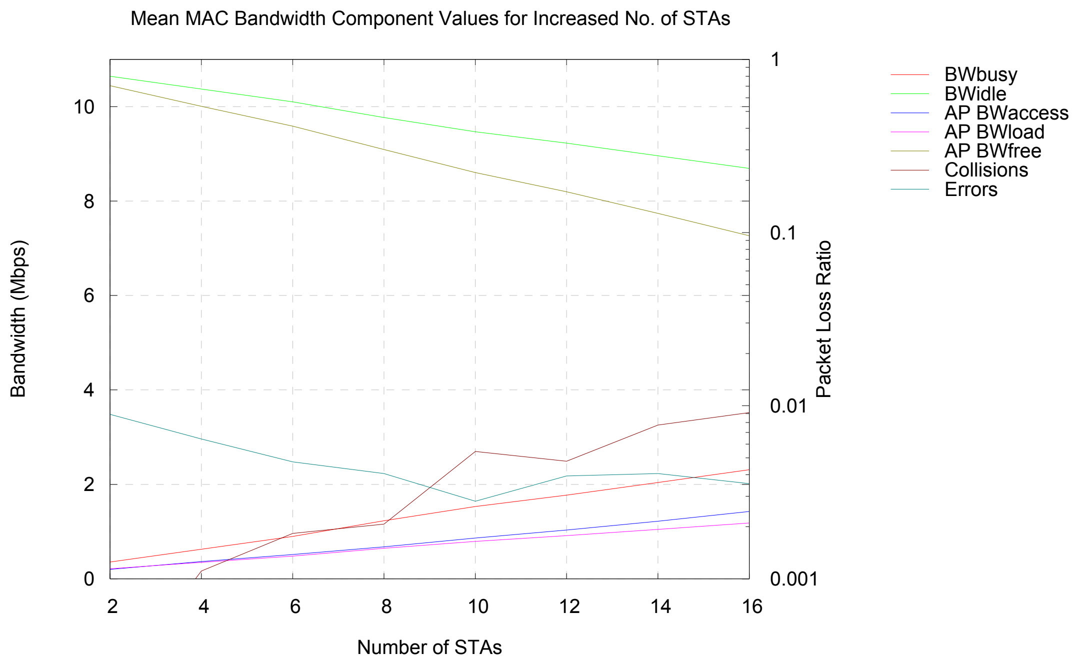


Figure 5.15: MAC Bandwidth values for increased number of STAs.

5.3.2 AP Delay Characteristics

We first take a brief look at the PDF and CDF for all 16 stations before analysing them in more detail. Figure 5.16(a) shows us that the majority of delays through the AP are concentrated below approximately 10 ms. When we observe the CDF for the same experiment, we can see that the probability of a station delay through the AP being under 10 ms is quite large (i.e. well over 90% of delays are below 10 ms). However, it is also clear that STA 13 and STA 14 endure longer delays compared to the other stations. A closer look at the PDF and CDF plots is then warranted.

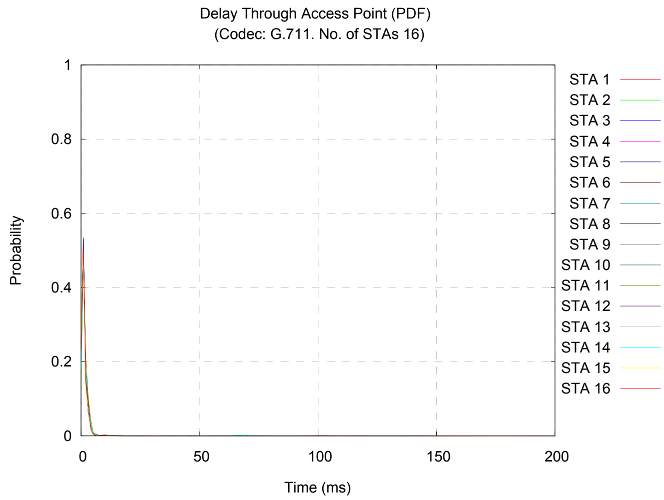


Figure 5.16(a): PDF for 16-station test.

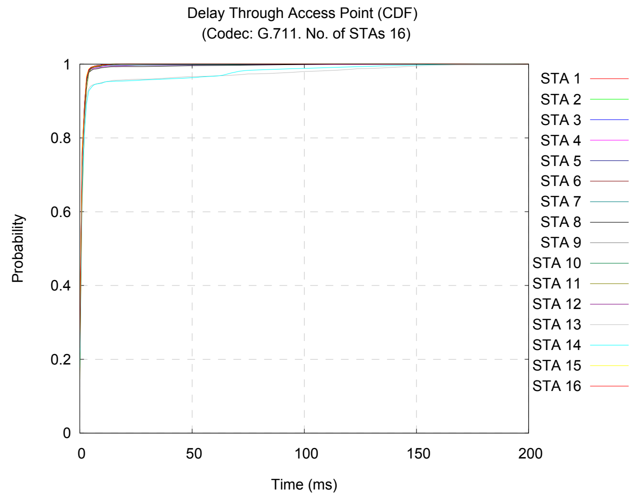


Figure 5.16(b): CDF for 16-station test.

A closer inspection of the PDF as shown in Figure 5.17 indicates that all stations bear similar delay characteristics (i.e. approximately 50% of delays centred around 1 ms). Compared to the two-station test this is a decrease in probability of about 0.07. Again when we look at the CDF we can clearly see however, that STA 13 and STA 14 suffer longer delays compared to all other stations. From Figure 5.18 we observe that approximately 99% of delays (with the exception of STA 13 and 14) are below 10 ms. In the case of STA 13 and 14 approximately 95% of their delays are below 10 ms.

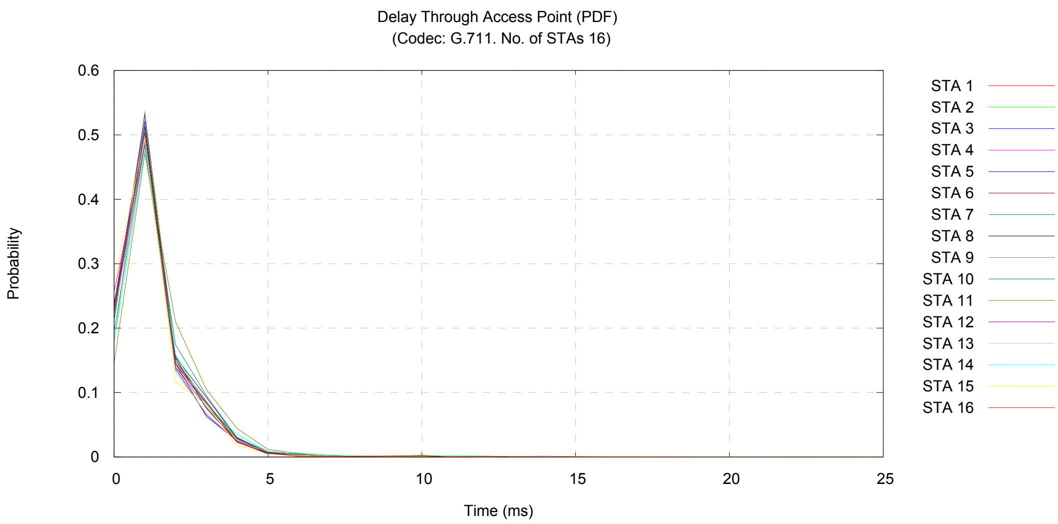


Figure 5.17: Zoomed in section of Figure 5.16(a)

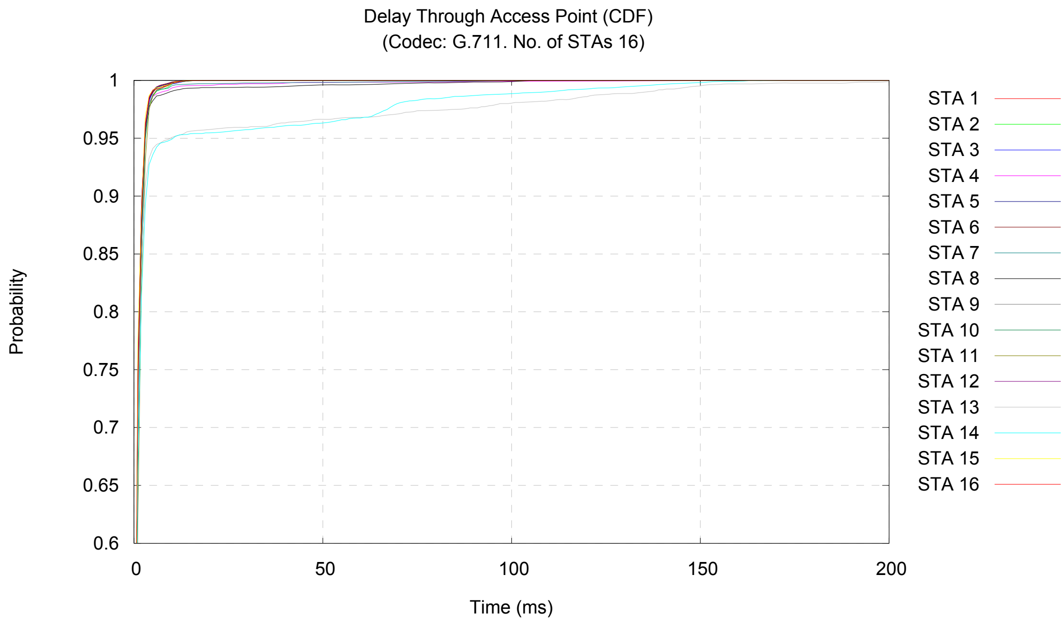


Figure 5.18: Zoomed in section of Figure 5.16(b).

Table 5.6 compares the delay values for STAs 1, 2, 13 and 14. All other STAs can be shown to have similar delay characteristics as STA 1 and STA 2 (depending on which voice sample was used, i.e. “Talker A” or “Talker B”). By comparing either the standard deviation or variance of the PDF for each STA, it becomes apparent that there is a significant difference in AP delay for the pair STA 13 and STA 14 in contrast to the pair STA 1 and STA 2. Clearly STA 13 and STA 14 are exhibiting longer delays than the other STAs. If we compare the CDF probability values we can see that STA 13 and STA 14 continue to experience large delays in excess of 20 ms, compared to STA 1 STA 2 which experience no AP delays after this point. Comparing the mean delay for each of the pairs also further supports this. We can see that the mean delay for either STA 13 or STA 14 is almost 5 times greater than STA 1 and STA 2 respectively.

CDF					PDF				
Delay, d (ms)	STA 1 P(D < d)	STA 2 P(D < d)	STA 13 P(D < d)	STA 14 P(D < d)		STA1 Peak	STA2 Peak	STA13 Peak	STA14 Peak
1	0.72788	0.71195	0.70964	0.65482	Delay:	1ms	1ms	1ms	1ms
2	0.87199	0.86890	0.84157	0.81365	Probability:	0.49	0.53	0.49	0.47
3	0.95554	0.94828	0.90934	0.89256					
4	0.98411	0.97856	0.93286	0.92726					
5	0.98926	0.98609	0.94120	0.93577					
10	0.99807	0.99673	0.95061	0.94912	Standard Deviation:	1.28	1.37	22.05	17.75
20	1.00000	1.00000	0.95745	0.95462					
30			0.95938	0.95729	Variance:	1.63	1.88	486.34	315.10
40			0.96322	0.96079					
50			0.96643	0.96296	Mean:	1.26	1.37	5.45	4.93
60			0.96814	0.96747					
70			0.97092	0.98031					
80			0.97413	0.98415					
90			0.97627	0.98699					
100			0.98033	0.98866					
110			0.98247	0.99016					
120			0.98589	0.99266					
130			0.98867	0.99449					
140			0.99188	0.99650					
150			0.99508	0.99800					
200			0.99872	1.00000					
250			1.00000	1.00000					

Table 5.6: Comparison of delay values for STAs 1, 2, 13 and 14 (16-STA test).

At this point it is clear that the behaviour of STA 13 and STA 14 with regard to the AP transit time differs from all other STAs. The increased delays endured by both of these STAs were experienced consistently throughout all tests involving STA 13 and STA 14. In order to determine the cause of this anomaly the hardware and software settings were thoroughly checked. The PCs and wireless network cards for both of these STAs were the same as STAs used during experimentation (see Table 4.1). When STA 13 and STA 14 were isolated and tested alone they displayed no signs of the anomalous readings. Both STAs were situated on a worktop approximately 1m above finished floor level and in close proximity to the AP (i.e. within 4m). At this point, despite exhaustive investigation, we still cannot establish why the behaviour for these two STAs differs from all other STAs although we have observed that it becomes more pronounced when the contention for access (i.e. increased number of STAs contending) and load on the medium is increased.

5.3.3 Delay Distribution Tails

In Figure 5.9 a typical Watermark plot for a 2-station test is displayed. Under those conditions we would expect relatively small delays to be incurred, as indeed is the case. We will now present the results for a watermark plot from the 16-station test scenario. In Figure 5.19 a watermark plot is displayed for each station. As mentioned in Section 2.9 such a plot should exhibit certain characteristics in order to determine the rate function, namely a “straight” part and a “wobbly” part. This is best illustrated in the case of STA 13 and STA 14 as indicated in Figure 5.19. The section before the “straight” part represents short time delays. The “wobbly” part represents the portion of the curve where there is an insufficient amount of data for reliable statistics. Once the relevant section has been identified the asymptotic slope can then be calculated.

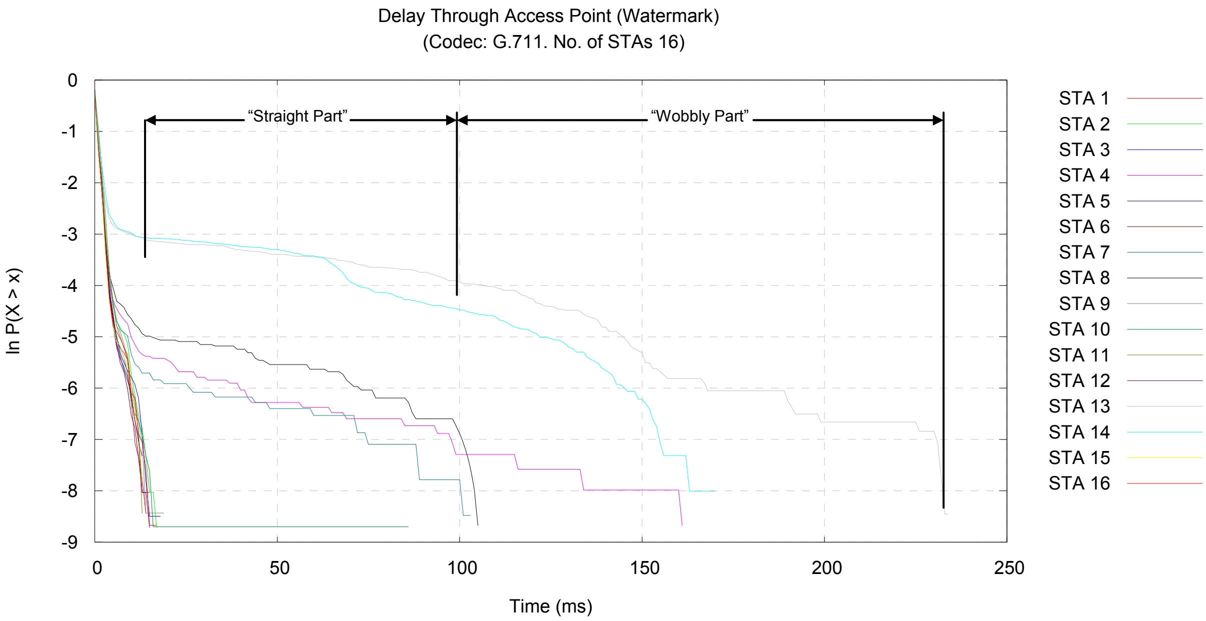


Figure 5.19: Watermark indicating “straight” and “wobbly” parts.

As we require the asymptotic slope of the “straight” part for each station it is important to note that this section will not be the same for all plots. In the case of Figure 5.19 for instance, the period for STA 13 and STA 14 is approximately between 10 ms – 100 ms. However, we can see from Figure 5.19 that this is not the case for all stations. In Figure 5.20 we take a closer look at the period from 0 – 100 ms and we can see that for STA 4, STA 7 and STA 8 the slope should be taken during the period of approximately 15 ms – 70 ms.

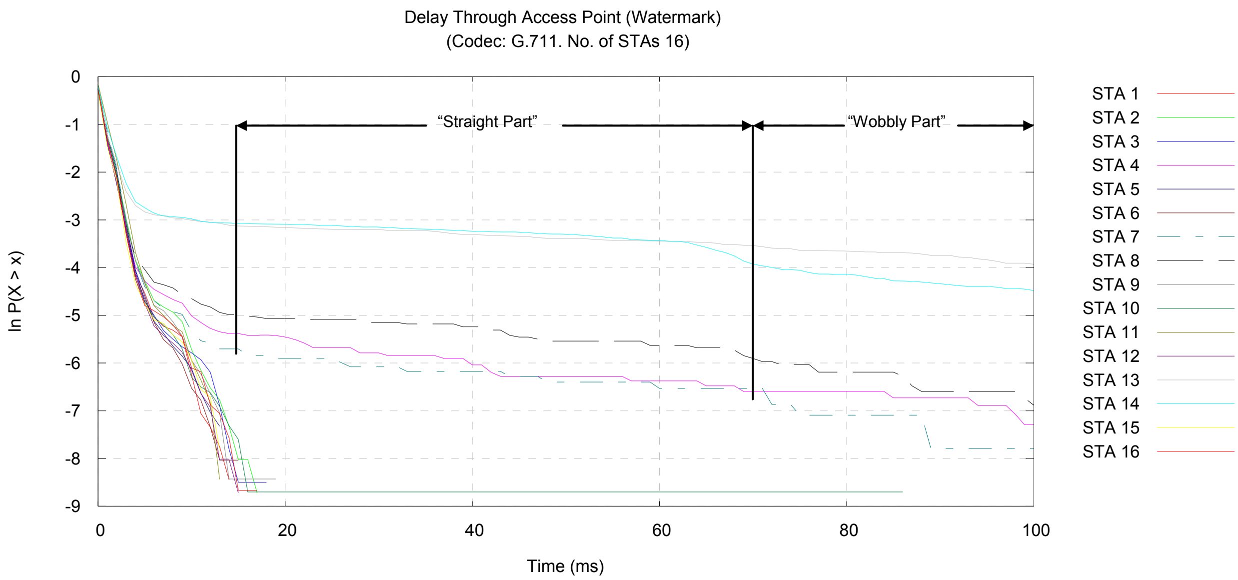


Figure 5.20: Different periods for characteristics of watermark plot.

In Section 2.9 we outlined the observations which could be made by measuring the asymptotic slope. We stated that if the magnitude of the asymptotic slope were large then there would be a small probability of finding large values of delay and consequently a small probability of buffer overflow. If the magnitude of the slope were small then there would be an increased probability of finding large delays and hence a greater possibility of buffer overflow. As the slope approaches zero, the rate of function indicates an increase in the duration to clear the AP buffer. As the magnitude of the asymptotic slope increases, the rate function indicates that delays are relatively small and can be serviced quickly. The majority of stations share a similar slope between approximately –0.20 and –0.40 (or greater), indicating that delays incurred are relatively small. On occasion few stations exhibit a much greater slope, for example –0.01 for both STA 13 and STA 14, during the 16-station test. This can be clearly seen in both Figure 5.19 and Figure 5.20 as the two curves obviously depart from the main group of stations.

	Number of Stations							
	2	4	6	8	10	12	14	16
STA1	-0.28	-0.27	-0.01	-0.23	-0.31	-0.52	-0.24	-0.39
STA2	-0.30	-0.35	-0.23	-0.34	-0.33	-0.36	-0.32	-0.31
STA3		-0.29	-0.25	-0.35	-0.38	-0.30	-0.45	-0.20
STA4		-0.28	-0.01	-0.08	-0.27	-0.03	-0.02	-0.02
STA5			-0.23	-0.20	-0.39	-0.35	-0.27	-0.32
STA6			-0.23	-0.26	-0.23	-0.22	-0.44	-0.36
STA7				-0.23	-0.31	-0.43	-0.02	-0.01
STA8				-0.20	-0.22	-0.02	-0.02	-0.01
STA9					-0.38	-0.23	-0.35	-0.32
STA10					-0.36	-0.29	-0.34	-0.29
STA11						-0.31	-0.28	-0.34
STA12						-0.33	-0.20	-0.29
STA13							-0.01	-0.01
STA14							-0.01	-0.01
STA15								-0.28
STA16								-0.29

Table 5.7: Comparison of asymptotic slopes for each station.

5.4 Results Summary

The main goals of the experiments in this section were, to determine an upper limit on the number of VoIP stations capable of being supported by an IEEE 802.11b network and also to measure the MAC bandwidth components at this point. In particular, we were interested in the BW_{Free}. Secondary objectives were to determine if the network enjoyed a statistical multiplexing gain as well as to determine the asymptotic slope of an ICDF of AP delay and hence the rate function. By allowing for a maximum AP delay time of 10 ms based on [65] and a PLR of less than 1% based on [66], [68] it is possible to determine the point at which the level of QoS begins to decline. Experiments show that when using a G.711 codec with a 30 ms frame duration, the number of STA pairs could be increased to at least 8 (i.e. 16 STAs in total). Experiments also showed that 2 STAs under performed with regards to delay times compared to all other STAs. Results indicate that when 16 STAs are in use the available BW_{Free} for each STA must be greater than 8.60 Mbps and 7.30 Mbps for the AP. The standard deviation of BW_{Busy} and $BW_{Load,}$ increase as the number of stations increase, indicating that the network does not benefit from statistical multiplexing gains.

5.5 Introducing Background Traffic

The next set of results introduces UDP background traffic on the medium by means of a traffic generator (TG) alongside 7 VoIP conversation pairs. The TG packet sizes used are 128 byte, 256 byte and 512 byte. The TG generates packets with a Poisson distribution, with packet rates ranging from 100 pps to 1000 pps in steps of 100 pps. Each test is again run over 600 seconds and can be broken down into two main categories as defined by how the TG is arranged.

- Wireless TG source with Wired TG sink.
- Wired TG source with Wireless TG sink.

The main difference between these two tests (as outlined in Section 4.1) is that when the TG operates from wired to wireless, the AP contention mechanism would be under greater stress.

5.6 VoIP with TG (WLAN-to-LAN)

We will follow a similar pattern of analysis as that of the experiments without using the TG. Each of the TG tests can be further divided depending on the packet size and rate of packets generated. We will begin with the largest payload used (i.e. 512 byte) starting at a packet rate of 100 pps. We then analyse the MAC bandwidth components as the packet rate is increased up to 1000 pps. We then compare this to results obtained for a payload of 256 bytes and 128 bytes. Results for station delay times through the AP as well as results obtained for producing Watermark plots follow a similar line of investigation.

5.6.1 MAC Bandwidth Components

TG with 512 byte Payload:

Figure 5.21 displays the MAC bandwidth components for the wireless medium and the AP as the packet rate is increased. We can quickly see that when the offered load reaches a rate of 600 pps, the various MAC bandwidth components begin to level off. At this point the TG is still producing packets at an increased rate, however they are not all being successfully transmitted and received. This is also indicated by the comparable increase in packet loss due to collisions (i.e. the AP displays an increase in collisions from approximately 2% to 8%).

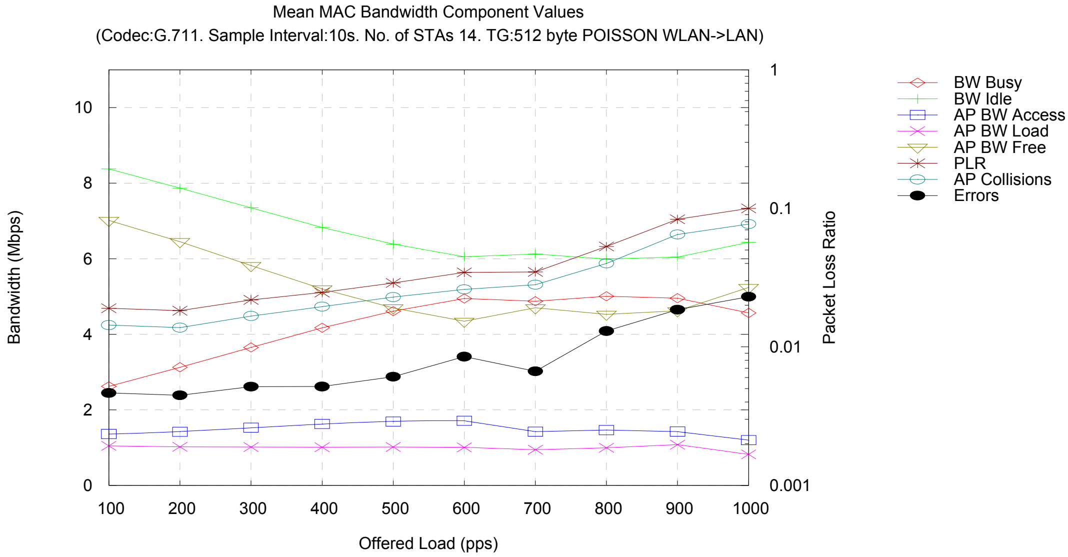


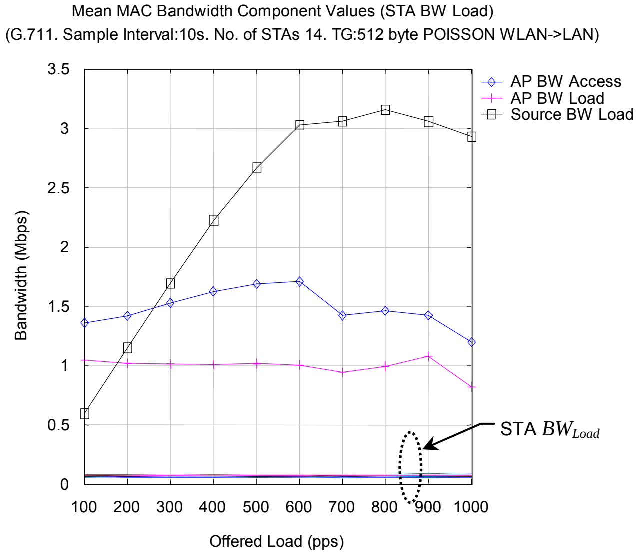
Figure 5.21: Increased TG packet rate (WLAN-to-LAN, 512byte).

Table 5.8 outlines the main MAC bandwidth components at the point where they begin to saturate (i.e. when the offered load reaches 600 pps).

Offered Load (pps)	BW_{Busy}	BW_{Idle}	AP BW_{Access}	AP BW_{Load}	AP BW_{Free}
100	2.62	8.38	1.36	1.05	7.01
600	4.95	6.05	1.71	1.01	4.34

Table 5.8: Summary of AP MAC bandwidth components in Mbps.

We now take a look at the effects of increased background traffic on the station MAC bandwidth components. Figures 5.22(a) and (b) are zoomed in sections of the lower end of the bandwidth scale. We can see that while the STA BW_{Load} remains relatively constant (as expected) and the BW_{Access} shows a marginal increase (owing to increased contention and hence increased deferrals), they are both well below 0.5 Mbps. However, the TG source BW_{Load} and TG source BW_{Access} increase linearly until the 600 pps mark (both at approximately 3 Mbps at this point with TG source BW_{Access} being slightly higher).



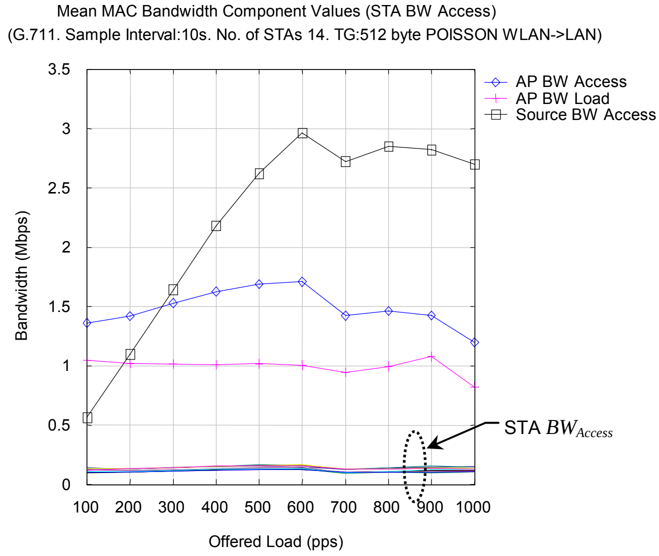


Figure 5.22(a): STA $BW_{Load} \approx$ 0.07 Mbps.

Figure 5.22(b): STA $BW_{Access} \approx$ 0.12 Mbps.

The plot of the STA BW_{Free} components shown in Figure 5.23 highlights the point at 600 pps when the BW_{Free} components begin to level off. All stations experience the same BW_{Free} characteristics as the TG increases its rate. At the 600 pps cut-off all stations show a BW_{Free} of approximately 6.00 Mbps.

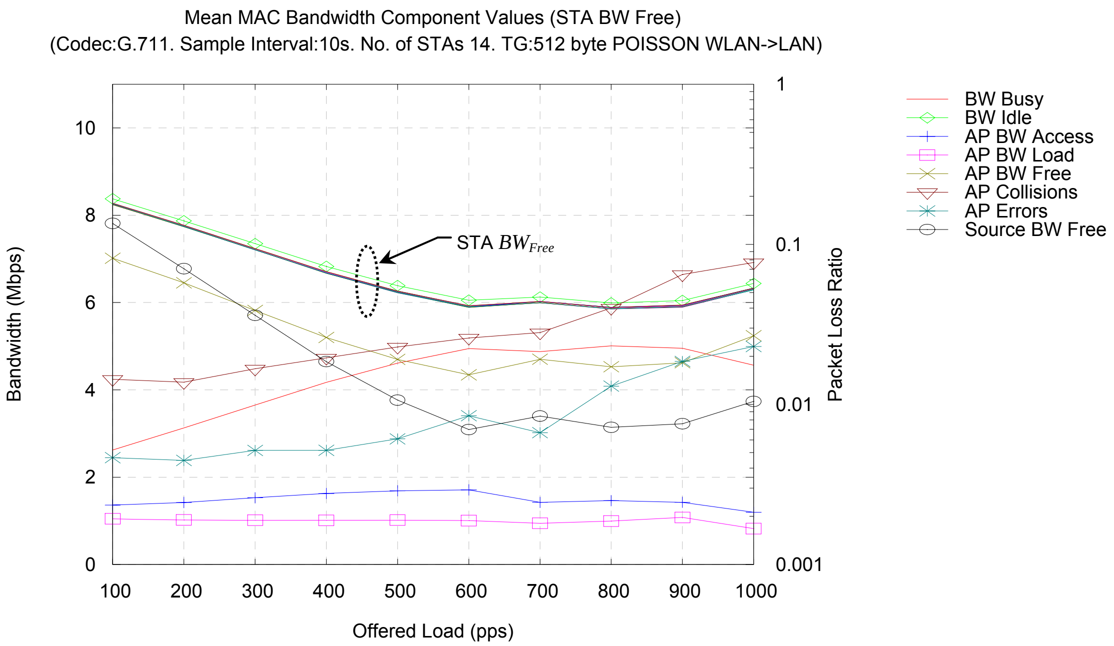


Figure 5.23: STA BW_{Free} with 512 byte TG payload (600 pps cut off).

Finally, for the tests using a payload of 512 bytes we will observe changes in the MAC bandwidth components for one station pair, STA 1 and STA 2 (all other station pairs can be shown to exhibit similar behaviour during the experiment).

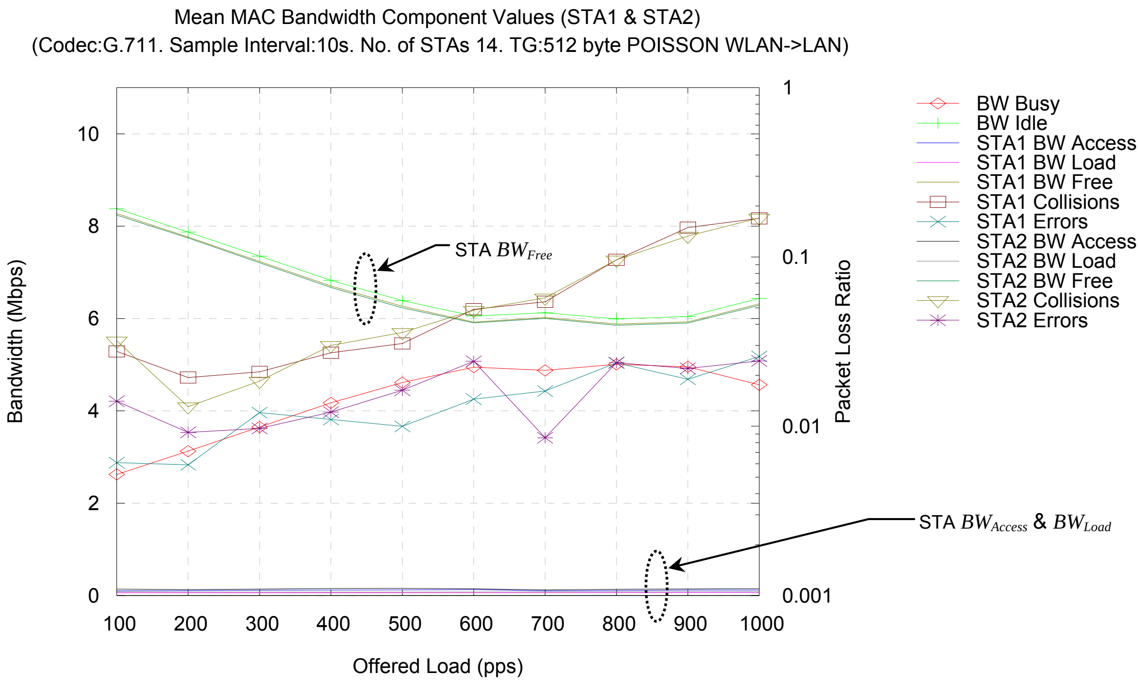


Figure 5.24: MAC BW values and PLR for a conversation pair.

In Figure 5.24 we can see the effect on all MAC bandwidth components for a conversation pair. We can also observe the packet loss on a per-station basis (Note: this value of packet loss is not entirely accurate. It is not a ratio of packets transmitted to packets received. Instead it is based on the packet header information which can be used to detect if a packet contained errors or if collisions took place). Although this value of packet loss underestimates the amount of packets possibly lost, it does give us an indication of why packets are being lost once they have been transmitted. We can see from Figure 5.24 that as the TG packet rate is increased the majority of packet loss occurs due to collisions. Losses due to collisions at the 600 pps cut-off point are approximately 5 times greater than losses due to errors at 5% and 1% respectively. As the TG increases its rate each station begins to experience an increase in collisions up to almost 17% when the TG is at 1000 pps.

Offered Load (pps)	STA 1					STA 2				
	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors
100	0.10	0.06	8.27	0.03	0.01	0.14	0.08	8.24	0.03	0.01
600	0.13	0.06	5.92	0.05	0.01	0.15	0.07	5.91	0.05	0.02

Table 5.9: Summary of main MAC BW values in Mbps from Figure 5.24.

TG with 256byte Payload:

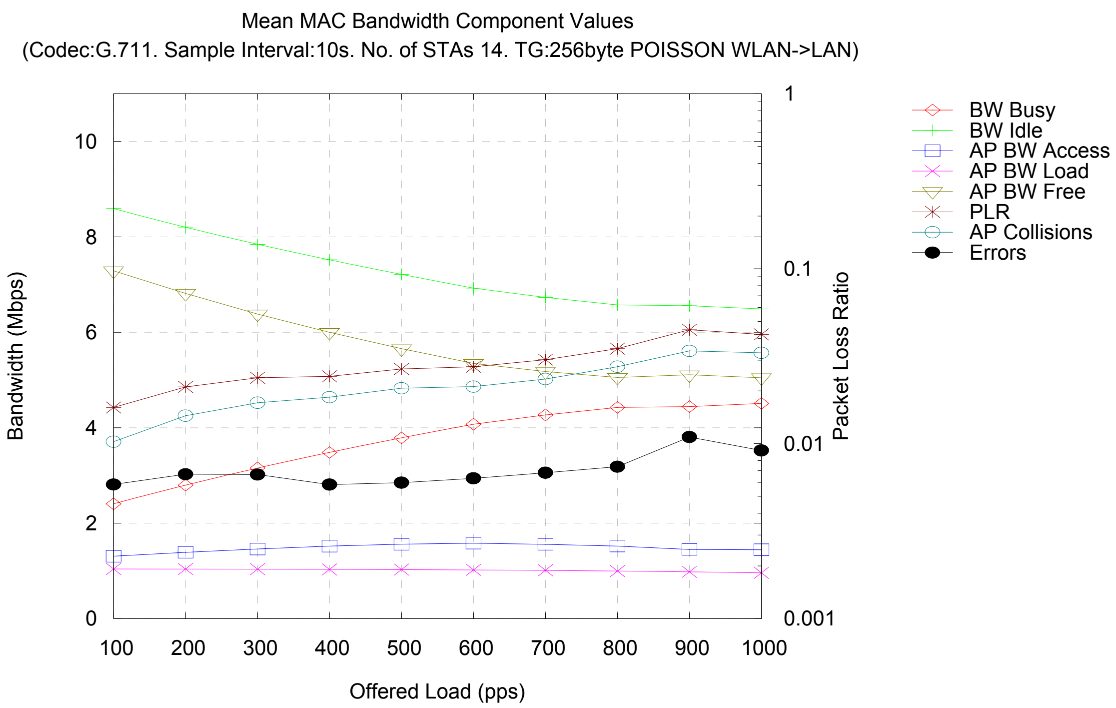


Figure 5.25: Increased TG packet rate (WLAN-to-LAN, 256 byte).

In Figure 5.25 we show the main AP MAC bandwidth components and the AP packet losses when the TG is increased using a 256 byte payload. This is similar to the graphs presented for the 512 byte payload, however the cut-off point when the MAC bandwidth components begin to level off appears to be 800 pps. Table 5.10 summarises the various MAC bandwidth components at this point. Also notice that the AP is showing a packet loss ratio of approximately 0.04 with the majority of these due to collisions (losses due to collisions account for approximately 3.5%).

Offered Load (pps)	BW_{Busy}	BW_{Idle}	AP BW_{Access}	AP BW_{Load}	AP BW_{Free}
100	2.40	8.59	1.30	1.04	7.29
800	4.42	6.57	1.52	0.99	5.05

Table 5.10: Summary of AP MAC bandwidth components in Mbps.

In Figures 5.26(a) and (b) we can see that the TG load and access components increase linearly until they reach 600 pps and then appear to begin levelling off at 800 pps. By looking at the lower half of the bandwidth scale we can observe the BW_{Access} and BW_{Load} values in closer detail. They appear to be in the same region as the values observed during the 512 byte payload experiments with the BW_{Access} showing a mean reduction of approximately 25 kbps at 800 pps (in comparison to the 600 pps cut-off for the 512 byte experiment).

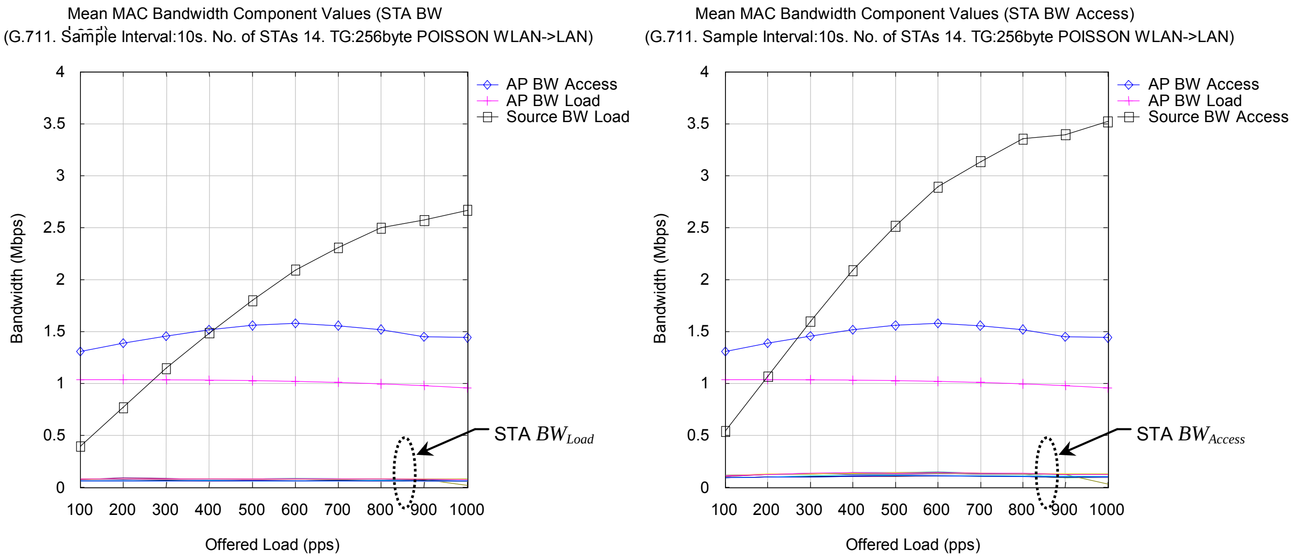


Figure 5.26(a): STA BW Load ≈ 0.07 Mbps.

Figure 5.26(b): STA BW Access ≈ 0.12 Mbps.

Finally, for this size TG payload we examine the BW_{Free} and also concentrate on all components for STA 1 and STA 2. We can see from Figure 5.27 that the TG BW_{Free} rapidly drops from 8.10 Mbps at 100 pps down to 4.05 Mbps at 600 pps and then begins to level out at 3.20 Mbps at 800 pps. A similar pattern is experienced for the STA BW_{Free}. If we look at Figure 5.28 we can see that the curve for the BW_{Free} indicates a linear decrease before an abrupt change at 600 pps and again at 800 pps. Table 5.11 summarises the main findings from this graph.

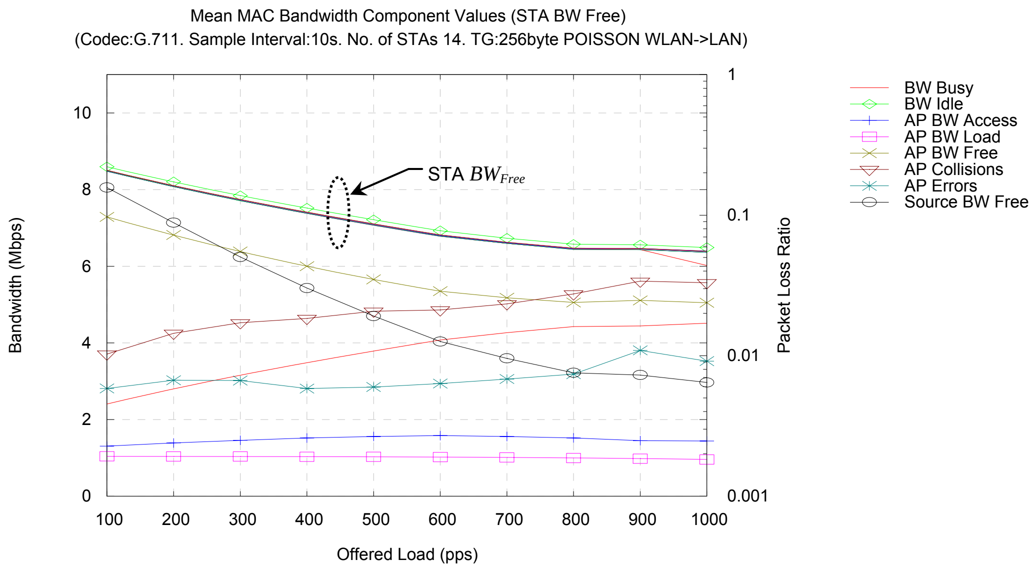


Figure 5.27: STA BW_{Free} components.

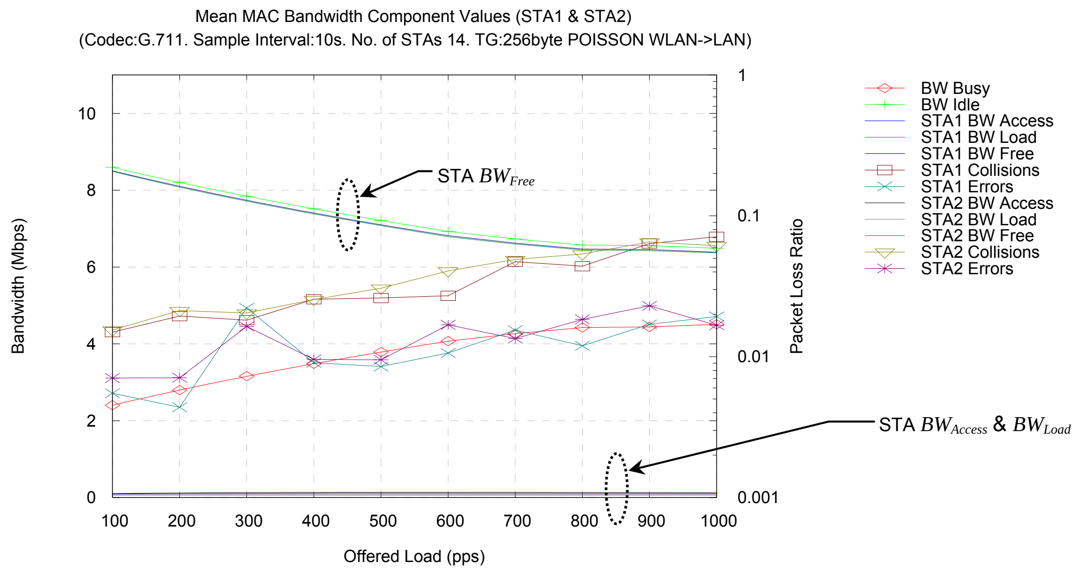


Figure 5.28: MAC BW components & PLR for conversation pair.

Offered Load (pps)	STA 1					STA 2				
	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors
100	0.09	0.06	8.50	0.01	0.00	0.10	0.07	8.49	0.01	0.01
600	0.11	0.06	6.82	0.03	0.01	0.14	0.08	6.79	0.04	0.02
800	0.10	0.06	6.47	0.04	0.01	0.13	0.08	6.44	0.05	0.02

Table 5.11: Summary of main points from Figure 5.28 (BW values in Mbps).

Again we can see that the majority of errors are due to collisions. Also, even though the traffic on the medium begins to saturate (as opposed to a continuous increase or decrease), the packet loss rate still tends to increase as the TG increases its rate.

TG with 128byte Payload:

The final TG packet size considered, for tests involving a wireless source to a wired sink, is 128 bytes. From Figure 5.29 it would appear that the BW_{Idle} and AP BW_{Free} almost decrease constantly while the BW_{Busy} and AP BW_{Access} increase constantly. When we look closer we can see a minor change in the curves when the offered load is at 800 pps. The BW_{Load} remains relatively constant, as was the case for the previous two experiments using a 256 byte and 512 byte TG payload. Table 5.12 summarises the MAC bandwidth components when the offered load is at 800 pps.

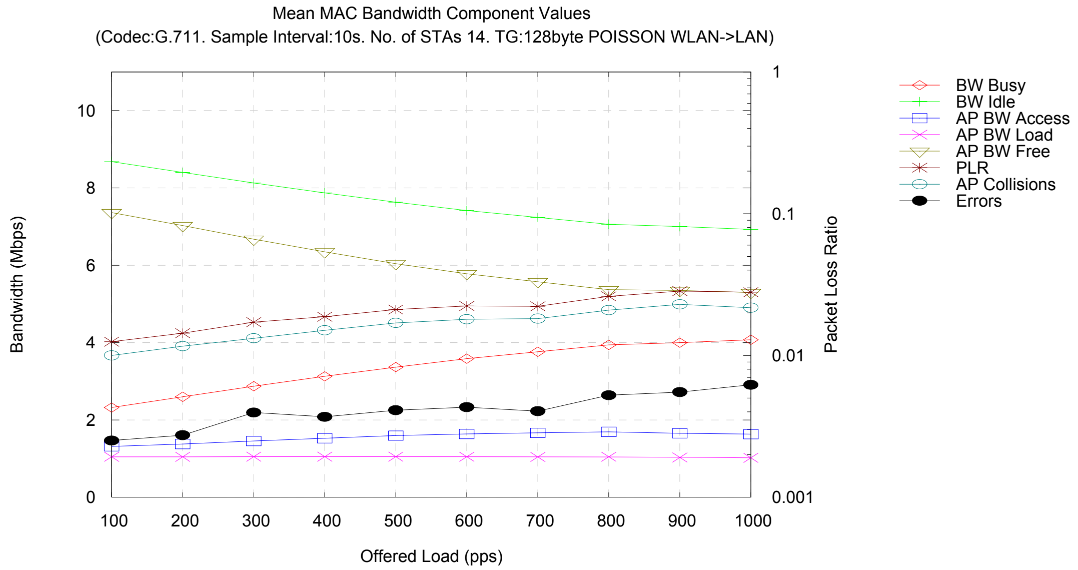


Figure 5.29: Increased TG packet rate (WLAN-to-LAN, 128 byte).

We can also note that there is a significant reduction in the AP packet loss compared to the tests using a 256 byte and 512 byte payload. AP losses due to collisions are of the order of 2% when the TG produces packets at 800 pps.

Offered Load (pps)	BW_{Busy}	BW_{Idle}	AP BW_{Access}	AP BW_{Load}	AP BW_{Free}
100	2.32	8.68	1.31	1.04	7.36
800	3.94	7.06	1.69	1.05	5.37

Table 5.12: Summary of AP MAC bandwidth components in Mbps.

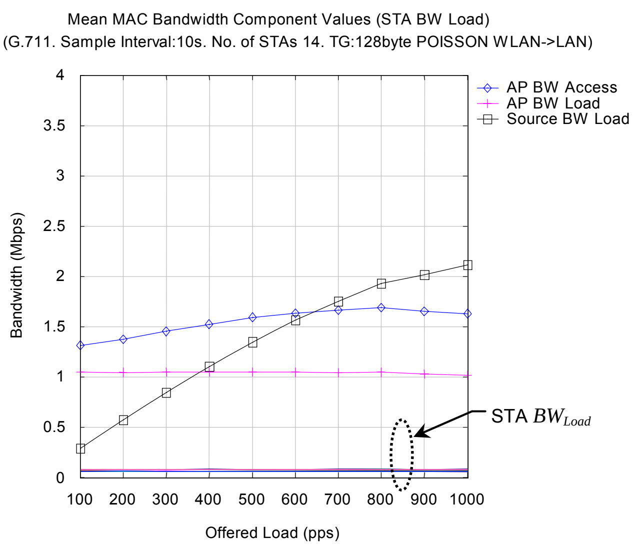


Figure 5.30(a): STA BW Load ≈ 0.07 Mbps.

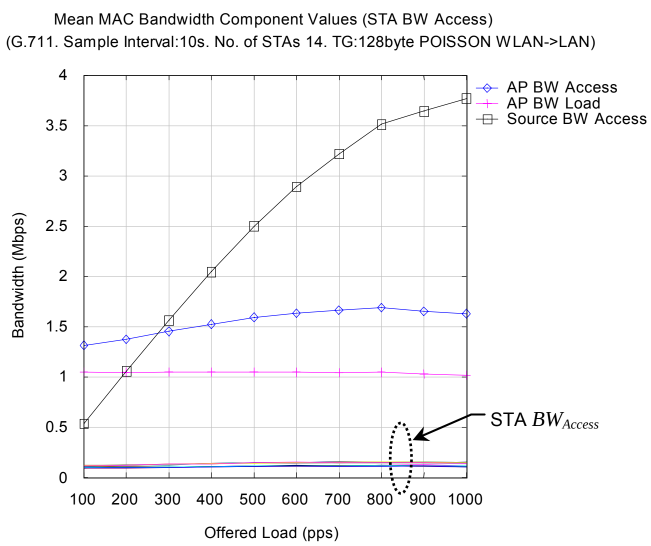


Figure 5.30(b): STA BW Access ≈ 0.13 Mbps.

The lower half of the bandwidth scale reveals that the STA BW_{Load} remains constant throughout each stage of increased TG packet rate. The STA BW_{Access} appears to remain relatively constant at approximately 0.13 Mbps. However, due to the increased contention it does show a slight increase over each subsequent test, from 0.09 Mbps at an initial offered load of 100 pps, to 0.15 Mbps at an offered load of 1000 pps. This is consistent with the increase in AP BW_{Access} but on a smaller scale.

Finally, we can see from Figure 5.31 that all stations experience the same BW_{Free} as the offered load is increased. We can notice that as the TG source BW_{Free} rapidly decreases the STA BW_{Free} parallels the characteristic of the AP BW_{Free} as well as the BW_{Idle}. In fact the STA BW_{Free} closely matches the BW_{Idle} at each increase in the offered load. We then look at the individual components for the VoIP conversation pair STA 1 and STA 2. We can see from Figure 5.32 that as the offered load increases the packet loss increases to about 5.5% for each station. A summary of values at key points is given in Table 5.13.

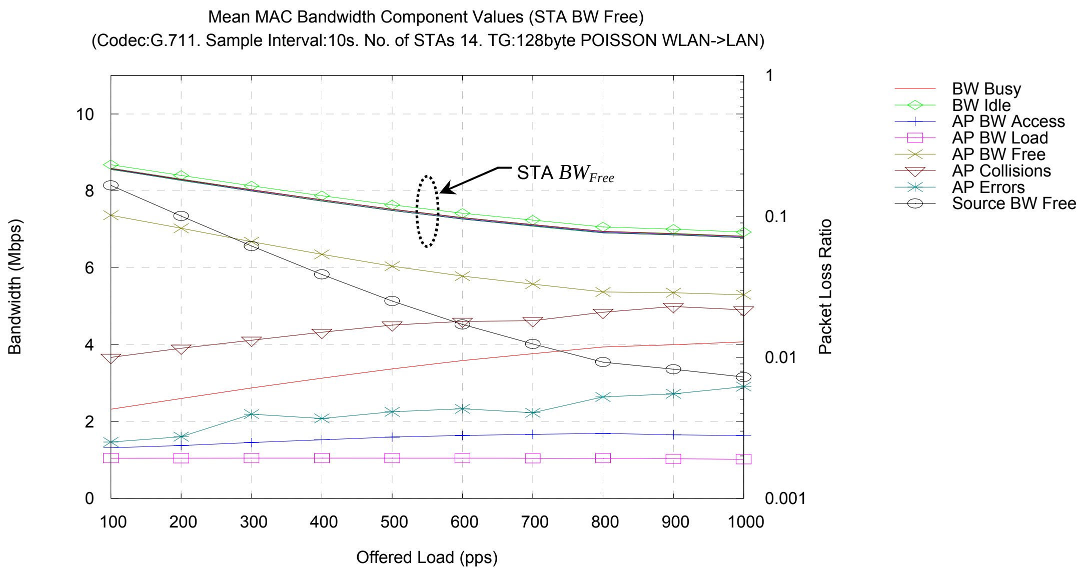


Figure 5.31: STA BW_{Free} components.

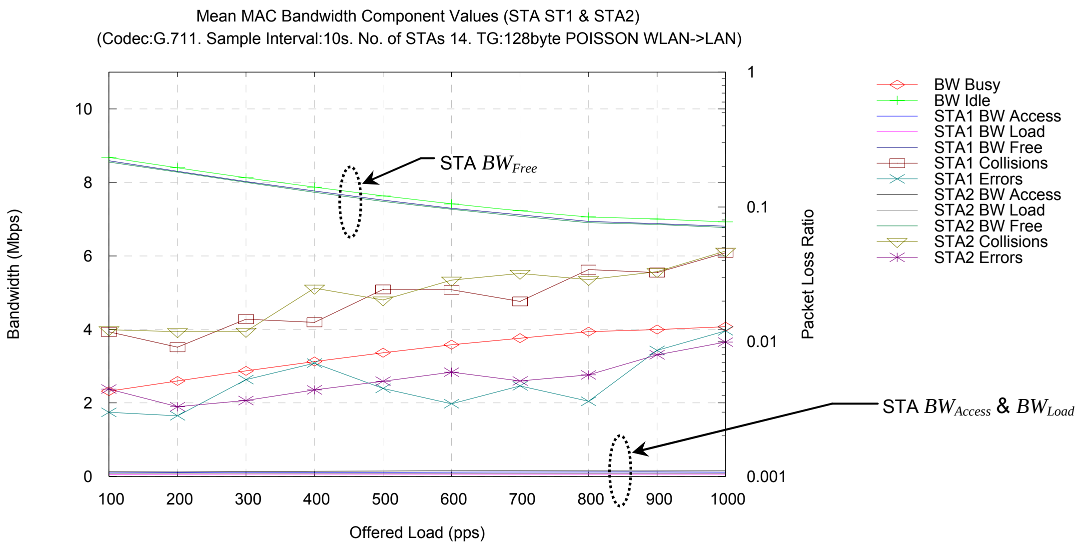


Figure 5.32: MAC BW components & PLR for conversation pair.

Offered Load (pps)	STA 1					STA 2				
	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors
100	0.09	0.06	8.59	0.01	0.00	0.12	0.08	8.55	0.01	0.00
800	0.12	0.06	6.94	0.03	0.00	0.15	0.08	6.91	0.03	0.01
1000	0.12	0.06	6.81	0.05	0.01	0.15	0.08	6.78	0.05	0.01

Table 5.13: Summary of main points from Figure 5.32 (BW values in Mbps).

5.6.2 AP Delay Characteristics

TG with 512byte Payload:

The delay characteristics for tests running the TG from a wireless source to a wired sink will allow us to determine the onset of the AP reaching its maximum service rate. Section 2.9 discussed the concept of the buffer service rate and in Section 4.1 we addressed the issue of the direction of TG packets through the AP. We saw that under no background load conditions, the medium could support 16 VoIP calls with STAs exhibiting 99% of AP delays under 10 ms. We now examine the AP delay as the service rate is stressed. We observed from the MAC bandwidth components that STA 1 and STA 2 were typical of average station behaviour. We now present the delays for these stations as the offered load is increased. From the CDF plots in Figures 5.33 (a) and (b) we can immediately observe the large delays incurred by STA 1 and STA 2 above 600 pps.

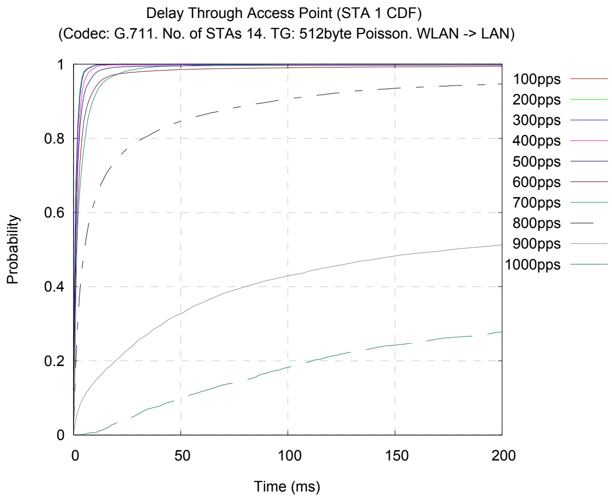


Figure 5.33(a): STA1 CDF of AP transit delay.

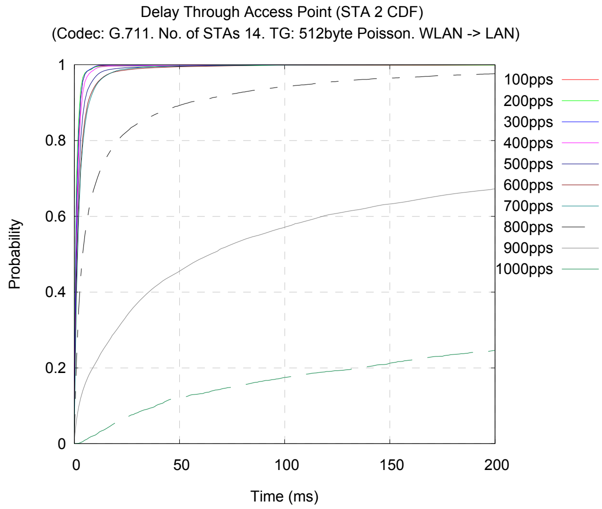


Figure 5.33(b): STA2 CDF of AP transit delay.

However, when we analyse the results for STA 13 and STA 14 we see a slightly different pattern of behaviour. In Figures 5.34(a) and (b) we again notice the low probability of delays under 5 ms for packet rates above 600 pps. We can also see that, with the exception of packets rates of 300 pps and 400 pps, there is a significant reduction in probability of delays under 10 ms for all other rates with respect to STA 1 and STA 2.

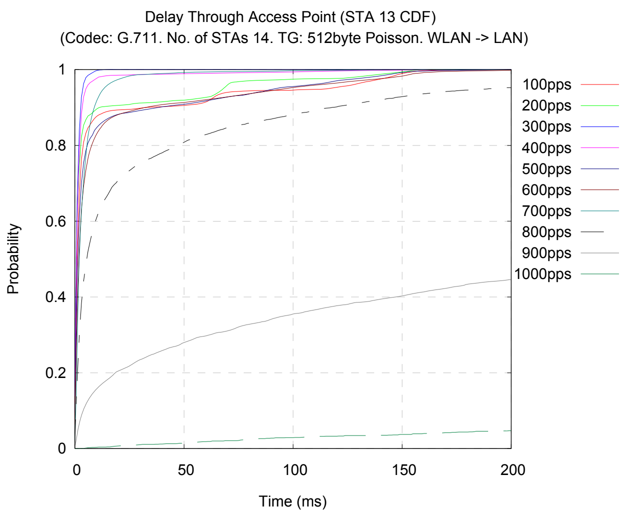


Figure 5.34(a): STA13 CDF of AP transit delay.

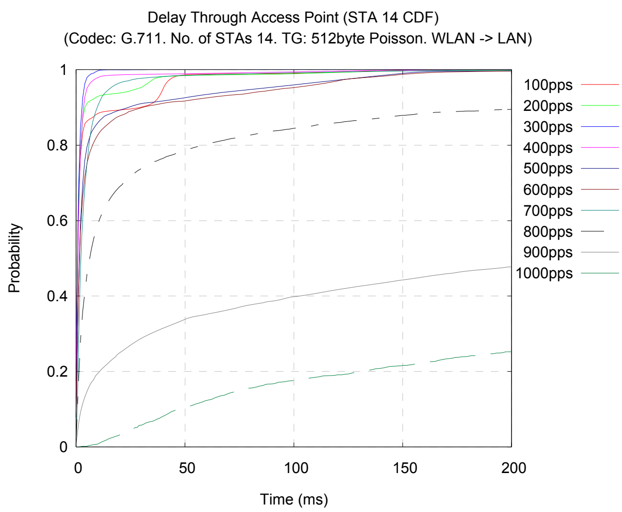


Figure 5.34(b): STA14 CDF of AP transit delay.

TG with 256 byte Payload:

For a 256 byte TG payload STA 1 and STA 2 begin to exhibit large delays when the packet rate is above 800 pps. Below this packet rate, 99% of delays are less than 10 ms. From Figures 5.35(a) and (b) we can notice than at 900 pps and 1000 pps, an appreciable amount of packets are delayed as long as 40 ms. This means that the packets are either delayed up in the AP buffer as it fills up or require retransmission.

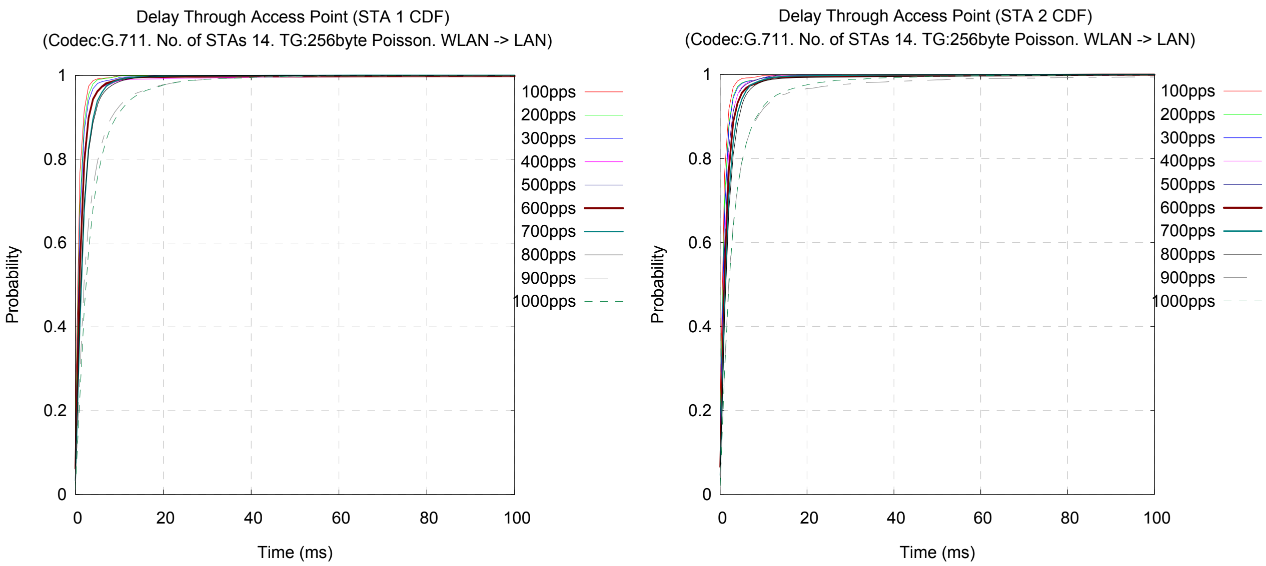


Figure 5.35(a): STA1 CDF of AP transit delay.

Figure 5.35(b): STA2 CDF of AP transit delay.

We will again compare these results to STA 13 and STA 14. In Figures 5.36(a) and (b) we can see that there are still relatively large proportion of delays greater than 50 ms. This occurs for both STA 13 and STA 14 when the packet rate is above 800 pps. Interestingly however, this pattern of behaviour also occurs most noticeably at 500 pps for each station. This means that we are not observing a monotonic increase in the delay with increased packet rate, as one would expect. This may be due to some form of temporal synchronization occurring between streams for a given TG packet rate and packet size. It would appear that AP begins to clear its buffer more effectively and is capable of minimising delays through it when the TG is generating packets at 700 pps and 800 pps.

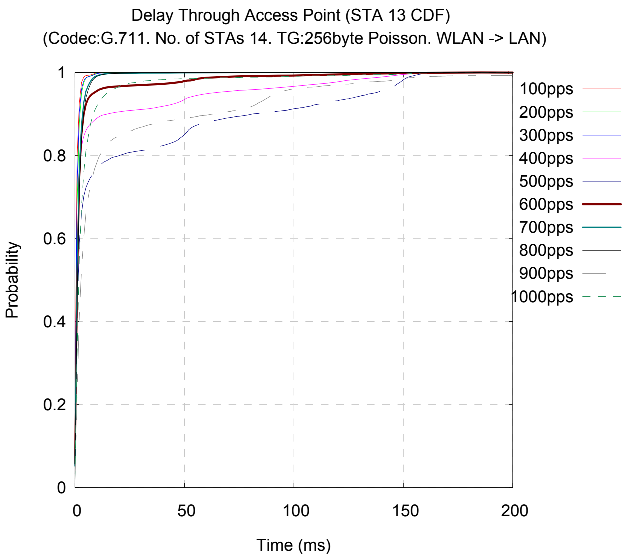


Figure 5.36(a): STA13 CDF of AP transit delay.

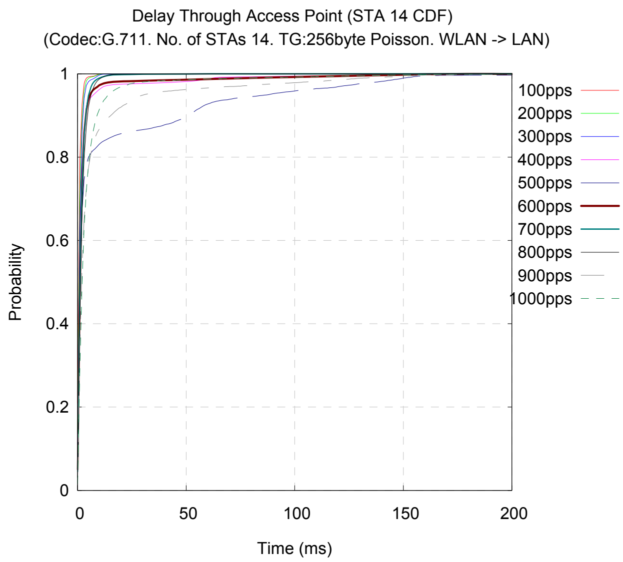


Figure 5.36(b): STA14 CDF of AP transit delay.

TG with 128byte Payload:

We would expect that as the TG payload decreases the STA delay times through the AP would also decrease. If we look at the delay performance for STA 1 and STA 2, Figures 5.37(a) and (b) respectively and compare them to Figure 5.35 and Figure 5.33, we can observe that this is indeed the case. With the exception of TG packet rates of 900 pps and 1000 pps, the probability of a delay being less than or equal to 10 ms is approximately 0.99 or greater (i.e. $P(D \leq 10\ ms) \approx 0.99$). Only when the TG rate is increased above 800 pps do the stations begin to exhibit increased delays through the AP with larger probabilities. Even at these rates, 97% of packets are cleared from the AP buffer within 10 ms and 99% of packets are cleared within 20 ms. The performance of STAs 3 to 12 can be shown to behave similarly. We look again at the performance of STA 13 and STA 14 for comparison.

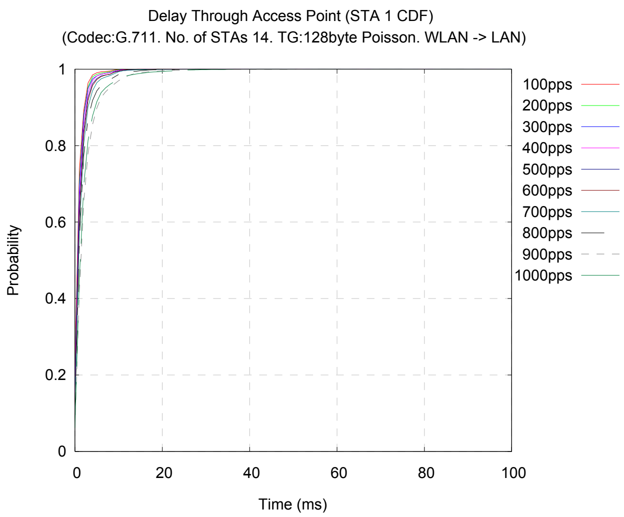


Figure 5.37(a): STA1 CDF AP transit delay.

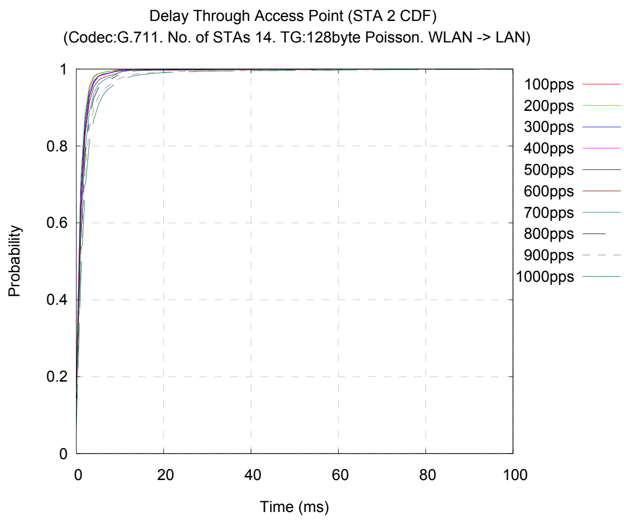


Figure 5.37(b): STA2 CDF AP transit delay.

From Figures 5.38(a) and (b) we can see a clear difference in the performance of STA 13 and STA 14 when compared to other STAs during the same experiment. Almost every TG packet rate impacts on the station delay performance. For the lowest TG packet rate used, 100 pps, the probability of delay being less than or equal to 10 ms for STA 13 is approximately 0.87. In fact, 1% (or more) of AP delays will be greater than 120 ms (in other words $P(D \le 10\ ms) \approx 0.87$ and $P(D \le 120\ ms) \approx 0.99$).

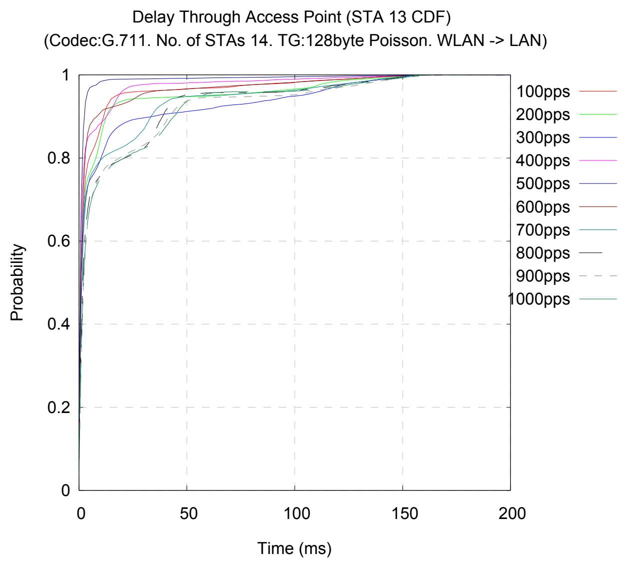


Figure 5.38(a): STA13 CDF, $P(D \leq 120\ ms) \approx 0.99$).

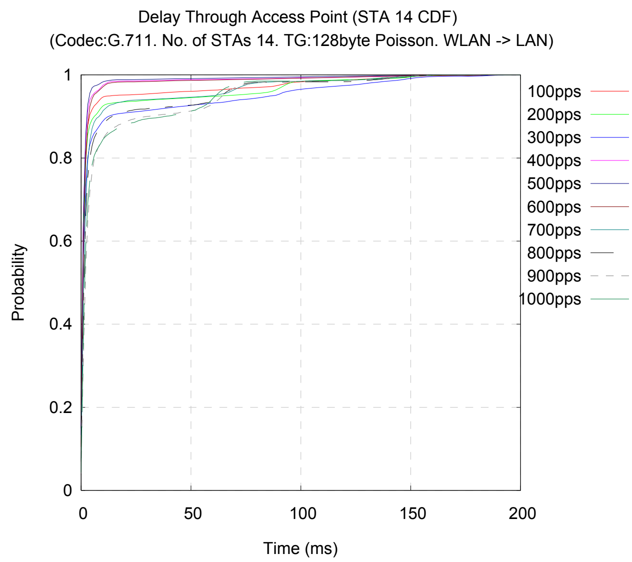


Figure 5.38(b): STA14 CDF, $P(D \leq 120\ ms) \approx 0.99$).

We can summarise the results for this section by tabulating some important values for STA 1, STA 2, STA 13 and STA 14. Tables 5.14 to 5.17 compare the probability of delays less than or equal to 10 ms. According to [65] when building a delay budget an AP delay of 10 ms should be allowed for. The tables also allow us to determine the mean delay time expected for various background traffic loads and packet rates. STA 1 and STA 2 are chosen, as they are typical of conversation pair, exhibiting a similar pattern of behaviour as STAs 3, 5, 7, 9 & 11 and STAs 4, 6, 8, 10 & 12 respectively.

	STA 1								
	P(D ≤ 10ms)			AP Delay (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.99672	0.99882	0.99921	1.59	1.72	1.20	1.47	1.34	1.21
200	0.99698	0.99825	0.99847	1.49	1.91	1.40	1.84	1.35	1.67
300	0.99669	0.99489	0.99708	1.57	1.55	1.92	9.86	1.39	1.39
400	0.99111	0.98708	0.99739	2.29	10.51	2.46	11.11	1.48	1.42
500	0.97521	0.99090	0.99450	3.30	17.45	1.88	1.94	1.61	1.88
600	0.93521	0.98986	0.99511	6.14	30.33	2.39	11.85	1.67	2.05
700	0.91497	0.98982	0.99334	4.36	6.93	2.40	2.82	1.82	2.36
800	0.63355	0.98496	0.98607	54.06	188.24	2.44	3.21	2.03	2.29
900	0.14477	0.92663	0.96983	505.34	601.79	4.30	7.03	2.75	3.71
1000	0.00650	0.91182	0.97540	902.61	681.26	4.69	6.41	2.54	3.20

Table 5.14: Comparison of delay statistics for STA 1

	STA 2								
	P(D ≤ 10ms)			AP Delay (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.99526	0.99762	0.9984	1.64	1.91	1.23	1.71	1.23	1.81
200	0.99408	0.99183	0.99818	1.84	10.21	1.69	4.75	1.31	1.25
300	0.99604	0.99159	0.99680	1.62	2.17	2.01	10.37	1.44	1.43
400	0.98961	0.99239	0.99719	2.10	4.44	1.82	3.61	1.51	1.45
500	0.96836	0.99258	0.99558	2.86	6.45	1.76	1.71	1.64	2.75
600	0.94649	0.98650	0.99142	3.73	12.236	2.20	4.45	1.82	3.75
700	0.94217	0.98785	0.99171	3.66	6.34	2.32	3.02	1.89	1.88
800	0.68546	0.98373	0.98690	25.35	79.92	2.82	6.85	2.06	2.95
900	0.21221	0.91944	0.97620	315.39	492.51	6.25	27.42	2.57	6.03
1000	0.02299	0.92517	0.96909	1003.20	692.76	4.46	7.73	2.91	6.55

Table 5.15: Comparison of delay statistics for STA 2

We also noted a marked difference in the conversation pair STA 13 and STA 14 when compared to all other stations. These could be considered as a worst possible case. There is a slight difference between the performance of odd numbered stations (using speech sample “Talker A”) and even numbered stations (using speech sample “Talker B”). This can be accounted for by the slight difference between the two audio speech samples (i.e. load and hence access time as well as differences in on-off speech due to the differences in speech).

	STA 13								
	P(D ≤ 10ms)			AP Delay (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.87669	0.99968	0.87366	12.35	32.89	1.14	1.06	5.92	20.06
200	0.89618	0.99824	0.84785	9.23	26.28	1.43	1.26	8.93	24.88
300	0.99758	0.99759	0.79718	1.51	1.47	1.35	1.68	12.81	30.52
400	0.97735	0.89041	0.88388	3.00	11.54	9.40	26.63	5.10	16.03
500	0.84709	0.76830	0.98410	12.65	31.13	19.99	40.34	2.40	10.63
600	0.82996	0.95575	0.99268	13.10	32.61	4.18	13.54	6.48	19.85
700	0.91952	0.99202	0.79555	4.82	9.45	2.05	2.31	11.90	26.81
800	0.60986	0.99102	0.75868	53.26	165.12	2.29	2.64	13.49	27.44
900	0.15833	0.78328	0.75936	573.87	615.56	17.28	43.12	14.78	29.82
1000	0.00346	0.91984	0.75123	1267.02	545.40	5.43	14.94	14.41	28.39

Table 5.16: Comparison of delay statistics for STA 13

	STA 14								
	P(D ≤ 10ms)			AP Delay (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.88414	0.99920	0.94706	6.43	17.04	1.14	1.16	1.34	21.69
200	0.92914	0.99871	0.92548	4.75	16.22	1.36	1.20	7.39	23.76
300	0.99726	0.99684	0.88946	1.49	1.51	1.55	1.46	10.24	29.19
400	0.97933	0.96257	0.97534	2.98	11.46	3.70	12.47	3.04	13.13
500	0.85941	0.83098	0.98367	11.53	30.25	13.36	32.19	2.516	10.48
600	0.82742	0.97259	0.90949	3.15	5.80	3.69	13.49	3.42	13.30
700	0.90506	0.99022	0.91387	6.36	21.30	2.20	2.50	7.32	21.60
800	0.59092	0.98734	0.88080	104.63	295.69	2.57	7.72	8.84	23.20
900	0.19523	0.86760	0.84629	576.89	627.80	8.47	22.69	10.44	23.39
1000	0.00786	0.91452	0.84496	965.72	712.44	5.59	14.66	10.15	22.86

Table 5.17: Comparison of delay statistics for STA 14

5.6.3 Delay Distribution Tails

TG with 512byte Payload:

While the CDF of delay times allow us to quickly determine which stations will have a relatively high probability of short delays, the extent of large delays are not always clear. We now present the findings of the ICDF for the same group of stations (STA 1, 2, 13, and 14) for the WLAN-to-LAN experiments and measure their asymptotic slopes to determine the rate function. Again it can be shown that all other stations compare similarly to STA 1 and STA 2.

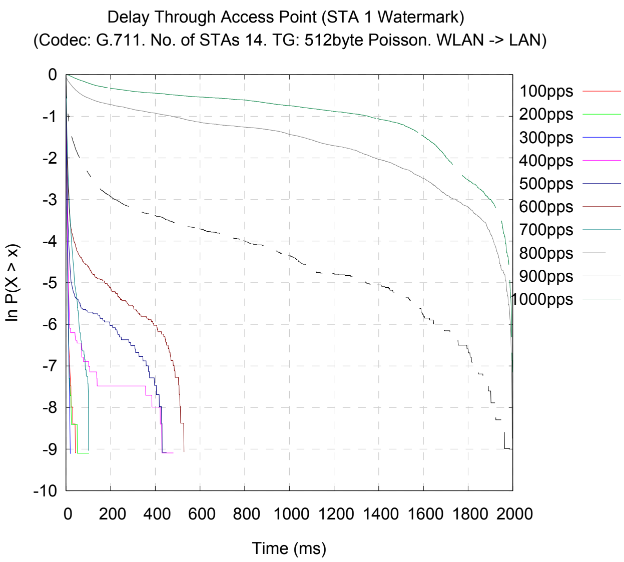


Figure 5.39(a): STA 1 Watermark plots.

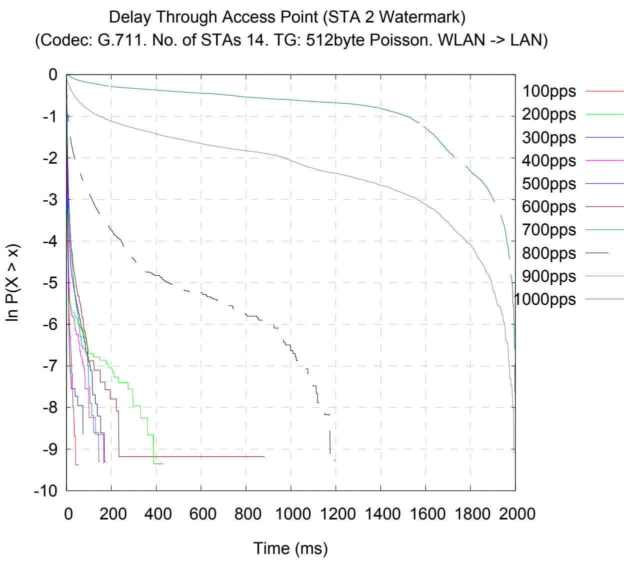


Figure 5.39(b): STA 2 Watermark plots.

Figure 5.39 shows that for both STA 1 and STA 2, VoIP packets are held up for increasingly long periods during tests using a 512 byte TG payload. It is clear from the plots that the thick tails occur for packet rates of 800 pps, 900 pps and 1000 pps. As such, we should notice that the asymptotic slope approaches zero as the offered load increases. Figure 5.40 shows the watermark plots for STA 13 and STA 14. Their characteristics are similar to STA 1 and STA 2 in that the longest AP delays occur during TG packet rates of 800 pps, 900 pps and 1000 pps.

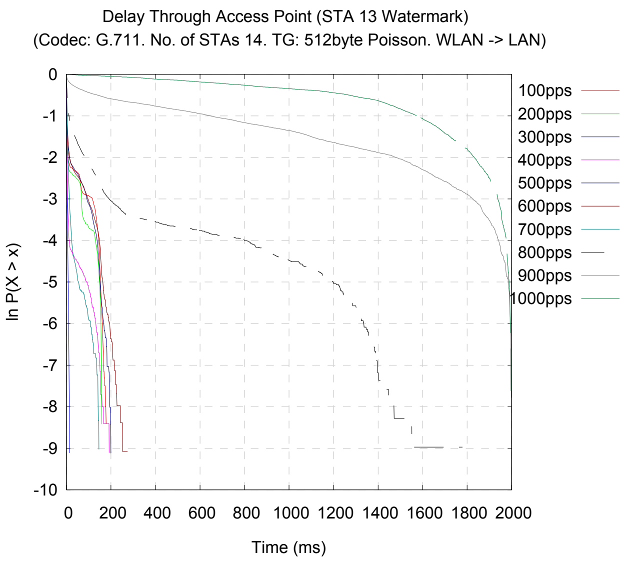


Figure 5.40(a): STA 13 Watermark plots.

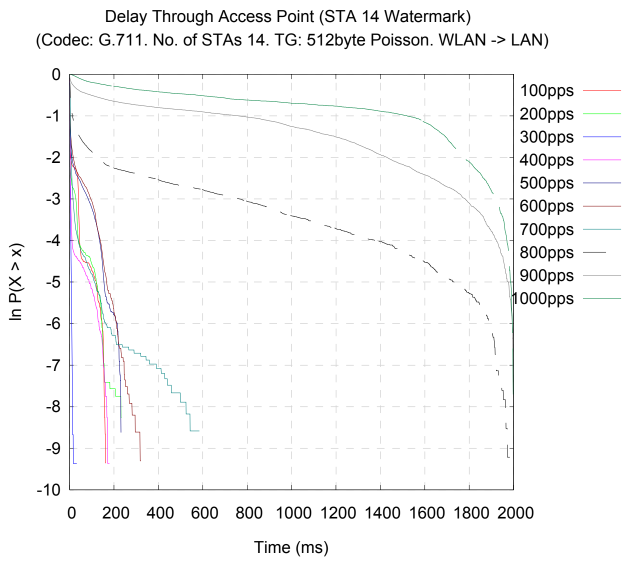


Figure 5.40(b): STA 14 Watermark plots.

Table 5.18 compares the asymptotic slopes of each of the four stations previously mentioned. We can see that the value of the asymptotic slope for each station approaches zero as the TG packet rate is increased.

Offered Load (pps)	Asymptotic Slope STA 1	STA 2	STA 13	STA 14
100	-0.42	-0.08	-0.01	-0.00
200	-0.44	-0.01	-0.02	-0.01
300	-0.42	-0.11	-0.42	-0.37
400	-0.36	-0.02	-0.01	-0.01
500	-0.00	-0.03	-0.01	-0.01
600	-0.00	-0.04	-0.01	-0.01
700	-0.06	-0.05	-0.02	-0.01
800	-0.00	-0.01	-0.00	-0.00
900	-0.00	-0.00	-0.00	-0.00
1000	-0.00	-0.00	-0.00	-0.00

Table 5.18: The asymptotic slope as the TG packet rate increases.

TG with 256byte Payload:

As the TG packet size is decreased, the average slope for each station is increasing (negatively). However, slopes begin to tend toward zero again as the packet rate is increased. This can be seen in both Figure 5.41 and Figure 5.42. Table 5.19 compares the slopes for each station.

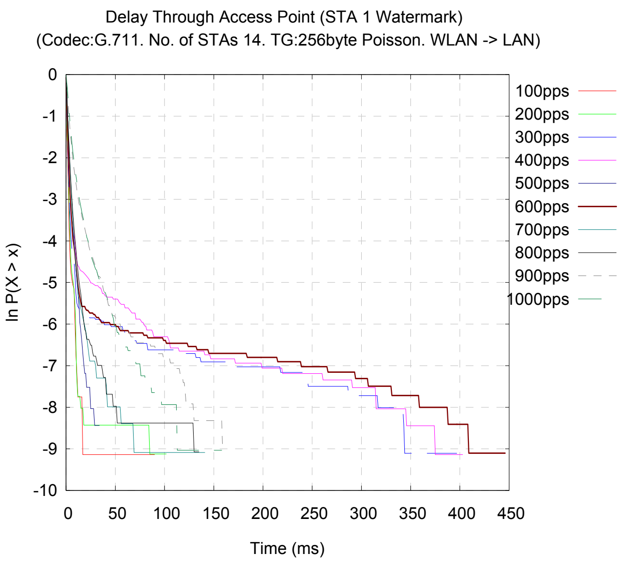


Figure 5.41(a): STA 1 Watermark plots.

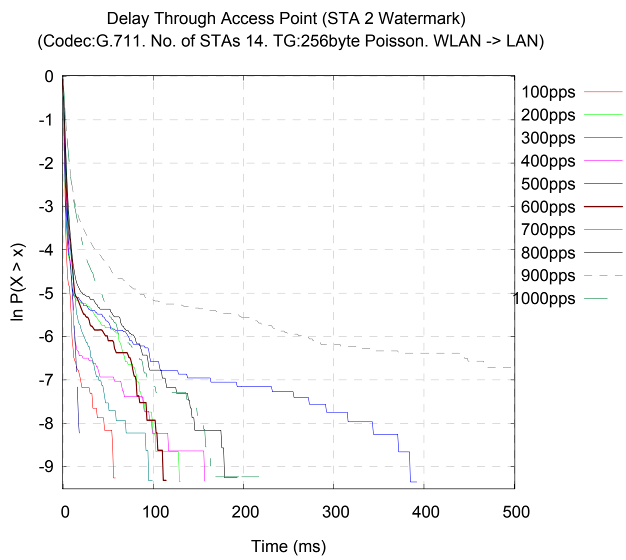


Figure 5.41(b): STA 2 Watermark plots.

In Figure 5.42 we can notice that the rate of decay is not consistent as the TG packet rate increases. What we mean by this is that, one would expect the slope to tend toward zero for each subsequent increase of background traffic. We can clearly see from the figure that five curves stand out (at 400 pps, 500 pps, 600 pps, 900 pps and 1000 pps). We can contrast this irregularity with other STAs in Table 5.19.

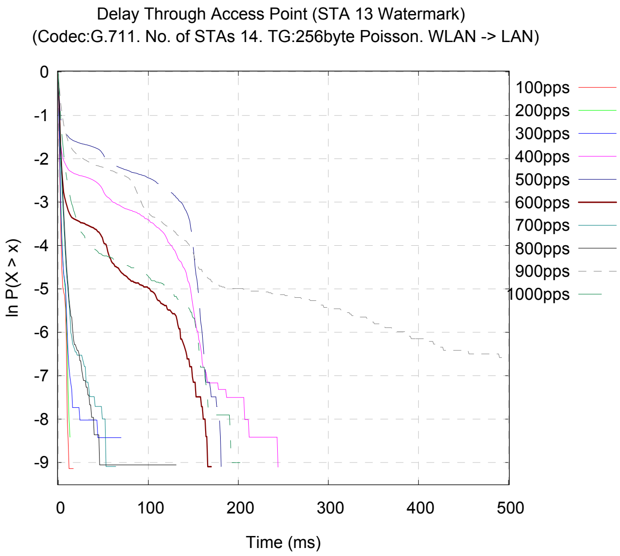


Figure 5.42(a): STA 13 Watermark plots.

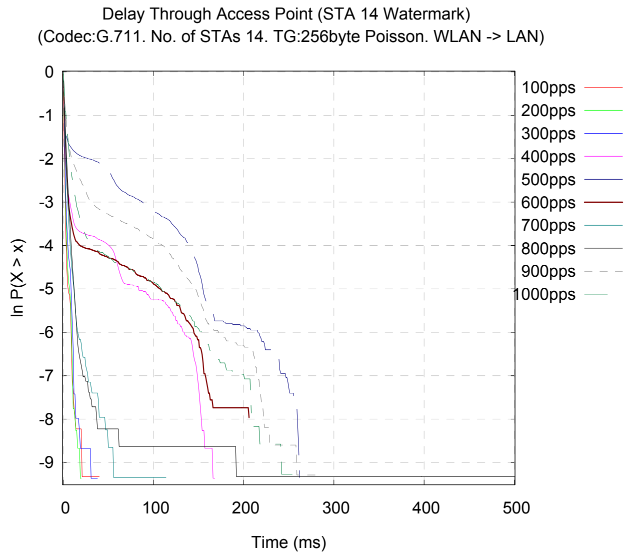


Figure 5.42(b): STA 14 Watermark plots.

Offered Load (pps)	Asymptotic Slope			
	STA 1	STA 2	STA 13	STA 14
100	-0.79	-0.29	-0.52	-0.87
200	-0.59	-0.17	-0.43	-0.70
300	-0.25	-0.18	-0.35	-0.58
400	-0.19	-0.31	-0.01	-0.02
500	-0.29	-0.30	-0.01	-0.02
600	-0.22	-0.23	-0.02	-0.01
700	-0.29	-0.29	-0.27	-0.28
800	-0.26	-0.26	-0.20	-0.34
900	-0.08	-0.00	-0.00	-0.01
1000	-0.09	-0.05	-0.01	-0.01

Table 5.19: The asymptotic slope as the TG packet rate increases.

TG with 128byte Payload:

Finally, in this section we look at the decay rate when a 128 byte TG payload is used. We can see from Figure 5.43 that the delays for STA 1 are relatively small compared to previous results (nearly all are under 50 ms). As a result the slope is quite large as packets are cleared from the buffer effectively. STA

2 exhibits some long tails at higher packet rates but for the best part the slope values are similar to STA 1.

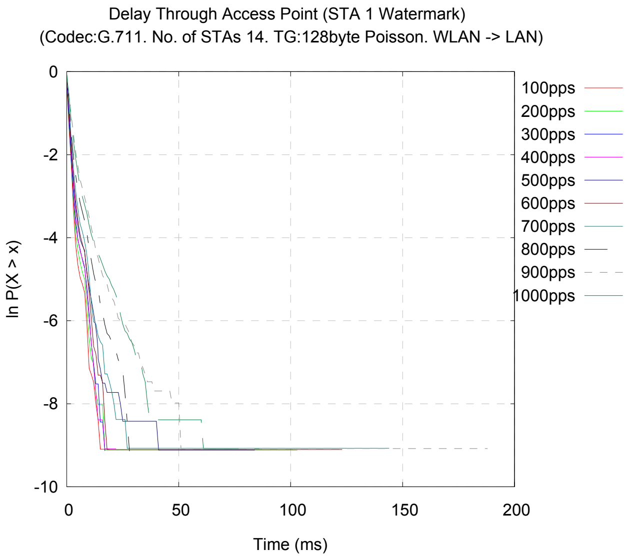


Figure 5.43(a): STA 1 Watermark plots.

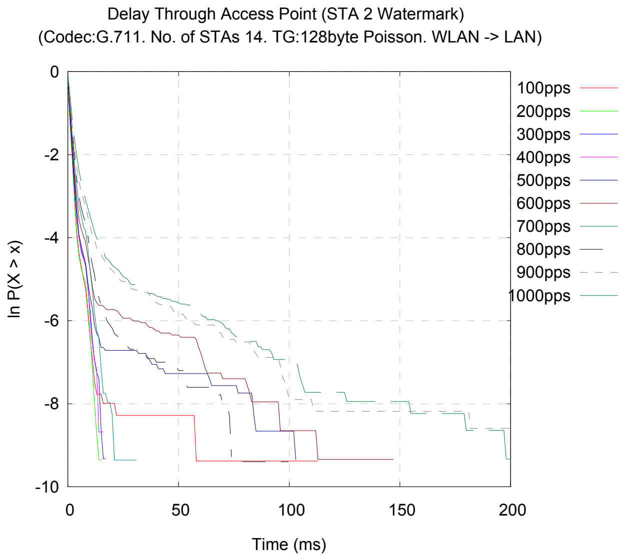


Figure 5.43(b): STA 2 Watermark plots.

In Figure 5.44 we observe some interesting results for STA 13 and STA 14. We would have expected these graphs to be comparable to Figure 5.40 and Figure 5.42 in the sense that these figures would indicate that as the TG payload decreases, the average slope tends to increase (negatively) also. However, we can clearly see that this is not the case. For each of these tests both STA 13 and STA 14 performed poorly compared to all other stations. It should be pointed out that each of these tests (i.e. at each packet rate) is independent of one another. The experimental set-up is reset for each test.

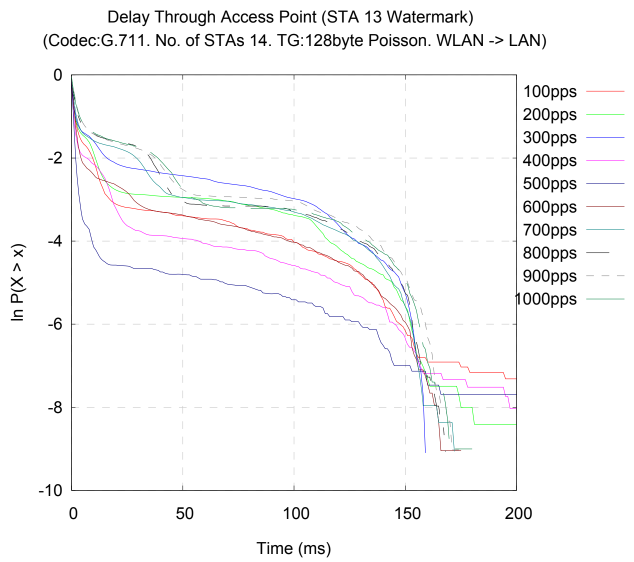


Figure 5.44(a): STA 13 Watermark plots.

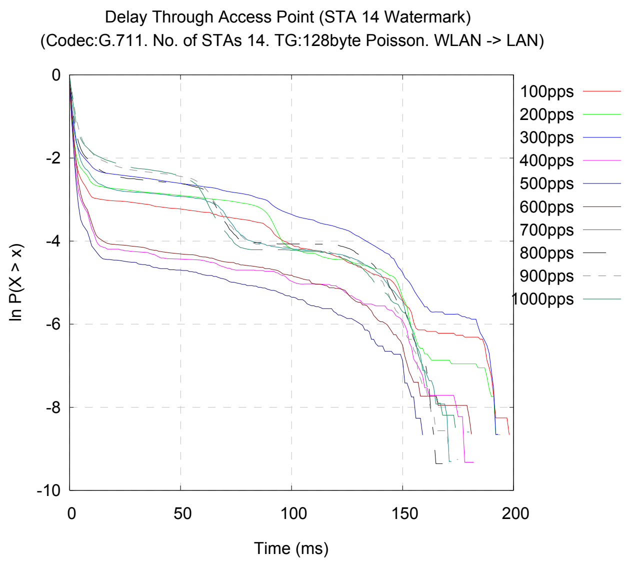


Figure 5.44(b): STA 14 Watermark plots.

Offered Load (pps)	Asymptotic Slope			
	STA 1	STA 2	STA 13	STA 14
100	-0.69	-0.44	-0.01	-0.01
200	-0.65	-0.47	-0.01	-0.01
300	-0.61	-0.41	-0.01	-0.01
400	-0.63	-0.42	-0.01	-0.01
500	-0.40	-0.32	-0.01	-0.01
600	-0.54	-0.02	-0.01	-0.01
700	-0.50	-0.46	-0.02	-0.02
800	-0.23	-0.03	-0.00	-0.01
900	-0.19	-0.03	-0.00	-0.01
1000	-0.14	-0.03	-0.01	-0.01

Table 5.20: The asymptotic slopes as the TG packet rate increases.

5.6.4 Results Summary

We found from the MAC bandwidth results that as we increase the offered load the MAC bandwidth components all tend to either increase or decrease linearly up until a point. The TG packet rate and packet size determines the point at which this linear behaviour ceases. This allows us to determine the approximate point for the onset of saturation. As the packet size is decreased the cut-off point increases (i.e. it occurs at a higher TG packet rate). A reduced TG packet size also results in lower percentage of errors due to collisions as well as an increased available BW_{Free}.

Analysis of delays show that as we increase the rate of background traffic, in general stations will exhibit longer delays. The TG is operating with a wireless source and transmitting to a wired sink via the AP. An increased packet rate essentially means that there will be more contention and as a result the AP will be under increasing pressure to service packets at the top of the buffer queue. The results show that delays do not always increase as the TG packet rate increases. They also show that in general, decreasing the TG packet size will result in an overall reduction in delays, however this is not consistent, particularly in the case of a TG payload of 128 bytes.

Comparison of station delays using the watermark plot method shows that two stations (i.e. STA 13 and STA 14 which are part of a conversation pair) consistently perform significantly worse than all other stations. This observation is particularly well defined at higher packet rates.

5.7 VoIP with TG (LAN-to-WLAN)

The following results will demonstrate the effects on various VoIP STAs as the background traffic is increased. This time however, the TG will operate by generating UDP packets from a wired source (i.e. by use of an Ethernet NIC) and transmitting them via an AP to a wireless sink. By doing this the AP buffer will contain both VoIP and TG packets. This should have an impact on the buffer occupancy and as a result it will fill up the AP buffer relatively quickly, hence increasing the probability of packets being lost (due to buffer overflow). The additional packets in the AP will also affect the service rate, as there will be an increase in packets on the AP downlink.

5.7.1 MAC Bandwidth Components

TG with 512byte Payload:

We first present the results for the AP MAC bandwidth components as the offered load is increased. This tends to give us an overall view of the network behaviour. In Figure 5.45 our attention is quickly drawn to sudden change in the MAC bandwidth components. While the TG is producing packets between 100 pps and 600 pps, the BW_{Idle} and BW_{Free} appear to decrease linearly from a starting point of 8.42 Mbps and 6.74 Mbps respectively. The BW_{Busy}, BW_{Access} and BW_{Load} all increase in a similar linear fashion starting at 2.58 Mbps, 1.68 Mbps and 1.62 Mbps respectively. All MAC bandwidth values then begin to saturate. This occurs when the TG is producing packets at a rate of 600 pps. We can clearly see that at this point, the network is not capable of accepting any more increases in load. The packet loss rate for the AP is relatively constant at approximately 1%. It would appear that the majority of packet loss is due to collisions.

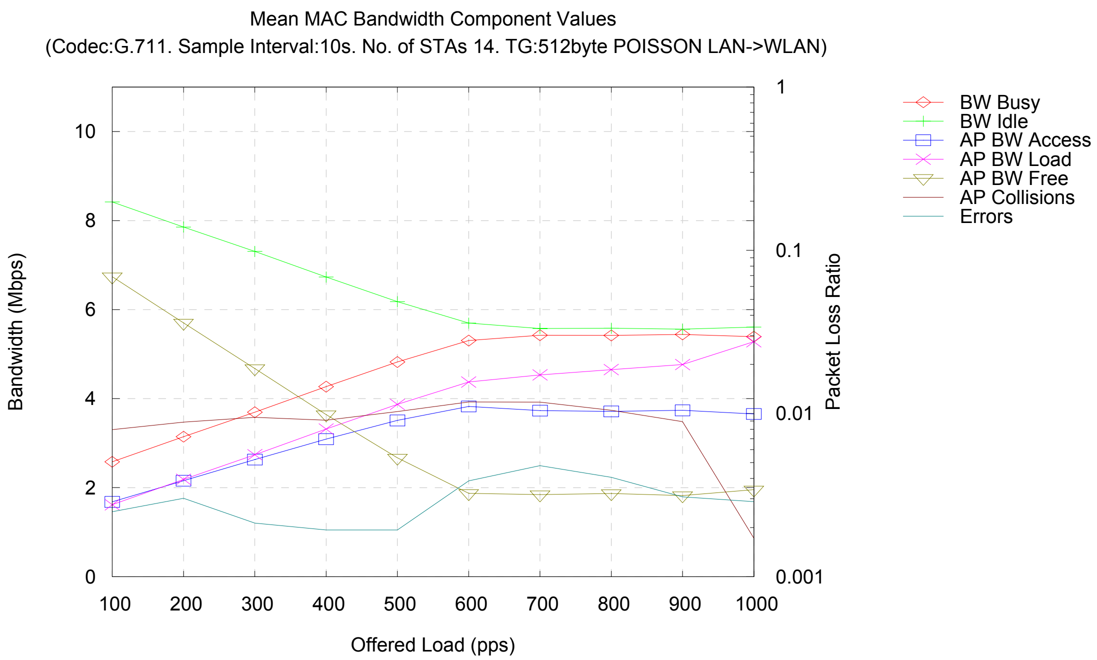


Figure 5.45: AP MAC BW components show distinct cut-off at 600 pps.

Another point to note is the increase in BW_{Access} and BW_{Load}. As the TG packets are destined for a wireless sink, the AP should indicate BW_{Access} and BW_{Load} values comparable to the source generating the packets. This is indeed the case, however we can also notice that the BW_{Load} increases at a greater rate than the BW_{Access} before they both level off at a rate of 600 pps. This would indicate that the medium is being used more efficiently as the offered load increases. Table 5.21 summarises the MAC bandwidth components at the point where their values begin to saturate.

Offered Load (pps)	BW_{Busy}	BW_{Idle}	AP BW_{Access}	AP BW_{Load}	AP BW_{Free}
100	2.58	8.42	1.68	1.61	6.74
600	5.30	5.70	3.82	4.37	1.87

Table 5.21: Summary of MAC BW values in Mbps.

We now concentrate on the station BW_{Access}. When we take a close look at the BW_{Access} for each station we can make a number of interesting observations. First of all in Figure 5.46 we notice two main groups. This is due to small differences in the voice samples used (i.e. Talker A and Talker B). The seven curves sitting on the lower half of the graph correspond to the odd numbered stations (or Talker A) and the upper seven correspond to the even numbered stations (or Talker B). The next point of interest is the behaviour of each station attempting to gain access to the medium. As the TG increases its rate, each station requires more BW_{Idle} to compete in winning more access opportunities for its load. This results in each station increasing its BW_{Access}, until a cut-off at 600 pps has been reached. However, instead of maintaining a relatively constant BW as the TG increases, the stations begin to suffer a dramatic decrease. This continues to occur until the stations gain virtually no access to the medium.

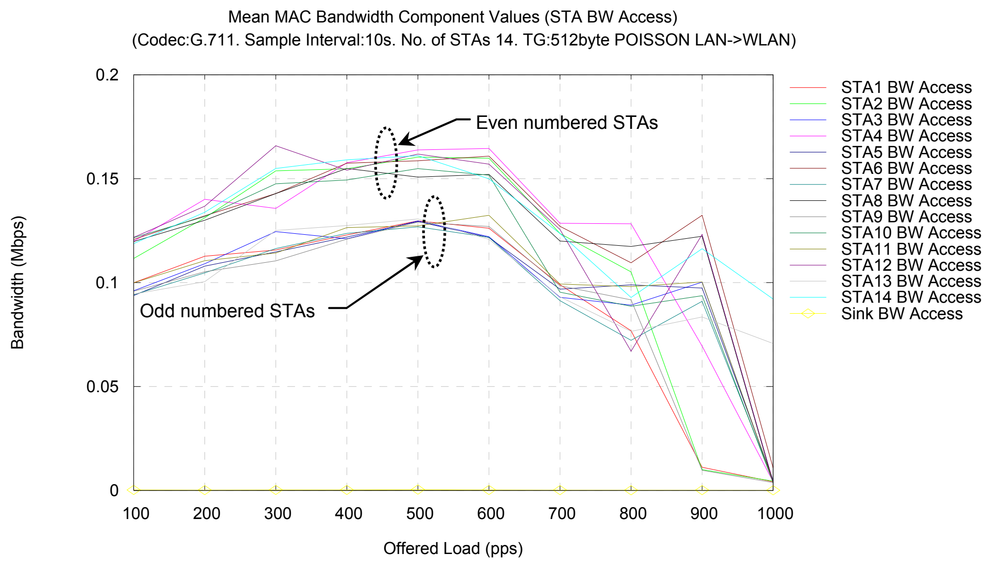
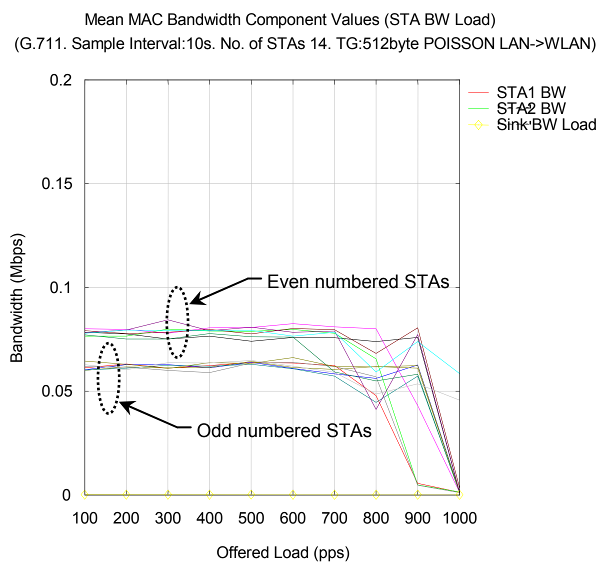


Figure 5.46: STA BW_{Access} increases until 600 pps cut-off is reached.

The BW_{Load} for each station remains relatively constant as the voice codec used has a fixed payload. Again we observe the two groups of curves based on the voice sample used. As the TG rate increases beyond 600 pps stations begin to drop off. This can be seen in Figure 5.47(a).



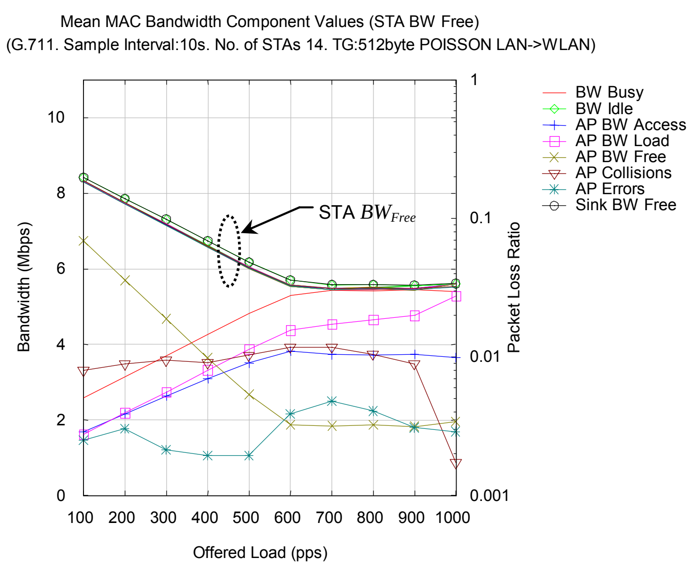


Figure 5.47(a): STAs dropped after 600 pps.

Figure 5.47(b): STA $BW_{Free} \approx$ 5.55 Mbps at 600 pps.

In Figure 5.47(b) we can see that all STAs exhibit the same values of BW_{Free}. As the TG packet increases the BW_{Free} available to each station starts off at approximately 8.31 Mbps, before dropping at a relatively constant rate until it levels off at roughly 5.55 Mbps when the TG reaches 600 pps. The MAC bandwidth components of STA 1 and STA 2 can be shown to be indicative of the remaining stations. We can see from Figure 5.48 that the BW_{Free} closely matches the BW_{Idle}.

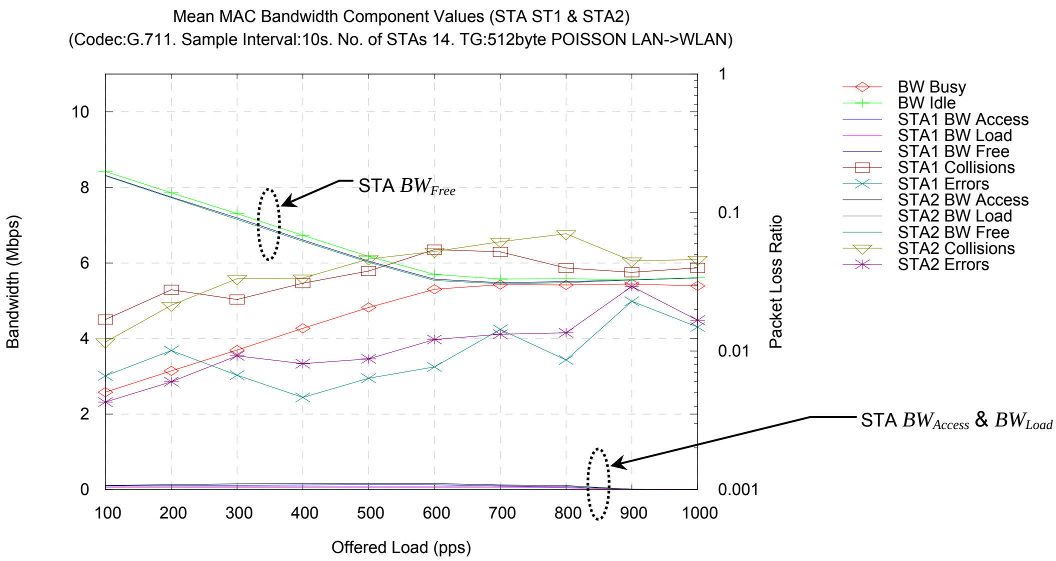


Figure 5.48: Comparison of STA 1 and STA 2 MAC BW components.

Also note that the majority of errors are due to collisions and are increasing as the background traffic increases. A summary of the STA MAC bandwidth components during the 512 byte payload tests can be seen in Table 5.22.

Offered Load (pps)	STA 1					STA 2				
	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors
100	0.10	0.06	8.32	0.02	0.01	0.11	0.08	8.31	0.01	0.00
600	0.13	0.06	5.57	0.05	0.01	0.16	0.08	5.53	0.05	0.01

Table 5.22: Summary of STA MAC BW values in Mbps.

TG with 256byte Payload:

The following sets of results are based around a TG payload of 256 bytes. From Figure 5.49 we can see that the general characteristic of the AP MAC bandwidth components are similar to the previous test. There is however a very noticeable difference, this time the MAC bandwidth components continue to change at a constant rate up until the TG is generating packets at 800 pps.

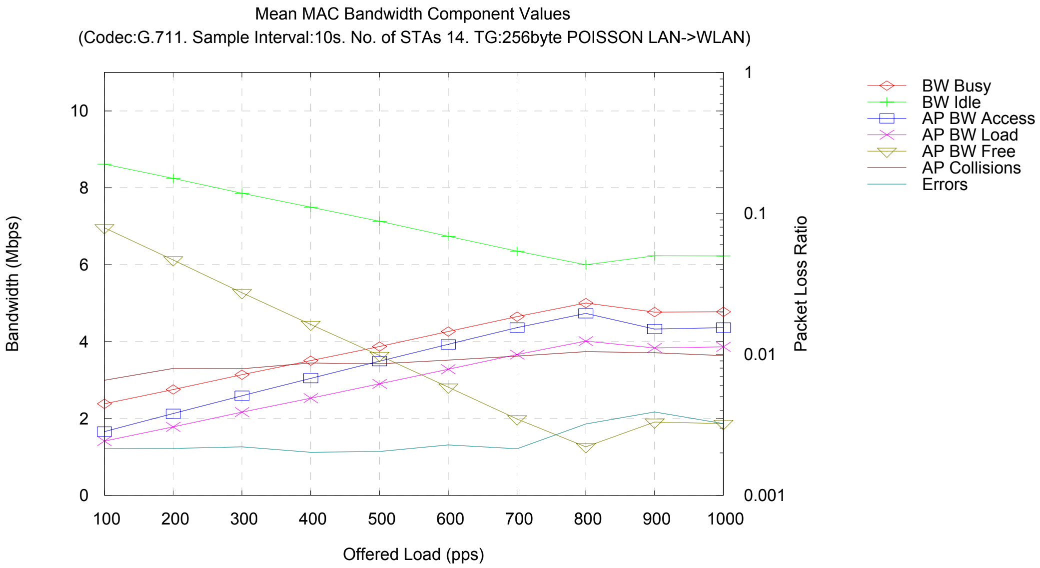


Figure 5.49: AP MAC BW components show distinct cut-off at 800 pps.

The load has peaked at a higher TG packet rate, however the BW_{Load} at this point is less than the peak value during the 512 byte experiment. As a result the AP BW_{Free} is higher at this point. The values at the 800 pps cut-off are summarised in Table 5.23. Again the majority of packet loss is due to collisions (at 800 pps percentage collisions ≈ 1% compared to percentage errors ≈ 0.3%)

Offered Load (pps)	BW_{Busy}	BW_{Idle}	AP BW_{Access}	AP BW_{Load}	AP BW_{Free}
100	2.381	8.619	1.658	1.413	6.961
800	5.003	5.997	4.735	4.011	1.263
900	4.768	6.232	4.326	3.836	1.906

Table 5.23: Summary of AP MAC BW values in Mbps.

Once again, from Figures 5.50(a) and (b) we can distinguish between "Talker A" and "Talker B". From Figure 5.50(a) it can be seen that the STA BW_{Access} increases as the offered load increases. After this point the values for each STA suddenly decrease rather than level out as might have been expected.

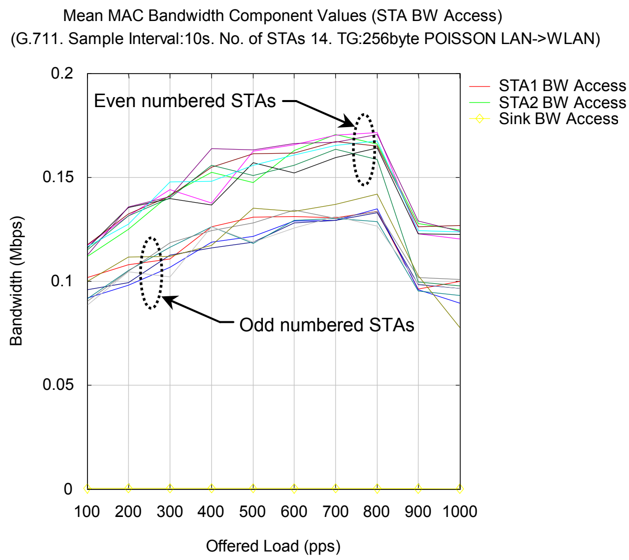


Figure 5.50(a): BW_{Access} decreases after 800 pps.

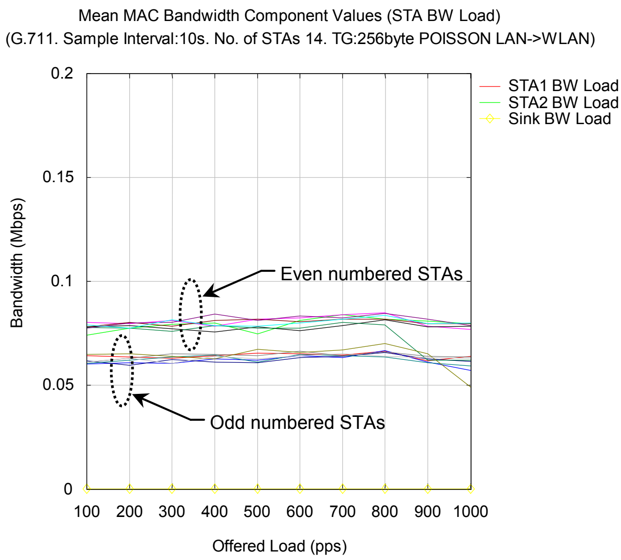


Figure 5.50(b): Talker "A" and "B" $BW_{Free.}$

Figure 5.51(a) compares the STA BW_{Free} as the offered load is increased. We can see from this graph that as the TG increases its packet rate, the STA BW_{Free} continues to decrease until it reaches a minimum of approximately 5.85 Mbps. The BW_{Free} closely matches the BW_{Idle}.

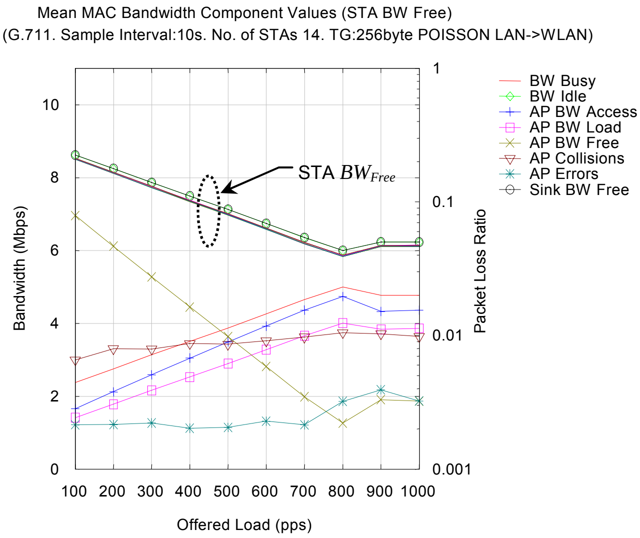


Figure 5.51(a): STA $BW_{Free} \approx$ 5.85 Mbps at 800 pps.

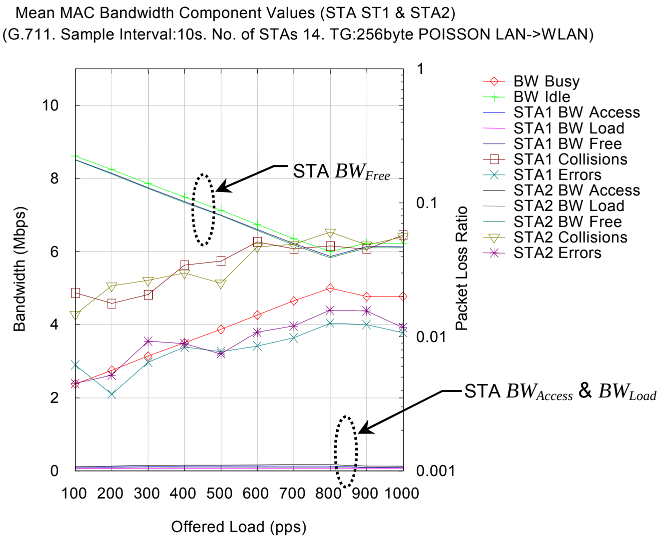


Figure 5.51(b): Comparison of STA1 & STA2.

From Figure 5.51(b) we can compare the various results for STA 1 and STA 2. From the graph we can see that when we compare packet loss, the majority of losses are due to collisions (approximately 5.4%). Table 5.24 compares a selection of results for STA 1 and STA 2.

Offered Load (pps)	STA 1					STA 2				
	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors
100	0.10	0.06	8.52	0.02	0.01	0.11	0.07	8.51	0.01	0.00
800	0.13	0.07	5.86	0.05	0.01	0.17	0.08	5.83	0.06	0.02
900	0.10	0.06	6.14	0.04	0.01	0.13	0.08	6.10	0.05	0.02

Table 5.24: Summary of STA MAC BW values in Mbps.

TG with 128byte Payload:

The final sets of results we will look at for this section are those with a TG payload of 128 bytes. Figure 5.52 shows the AP MAC bandwidth component results as the offered load is increased. The MAC bandwidth components either increase or decrease accordingly as the packet rate increases as before. However, instead of the values sharply levelling off at some point, this time the values tend to roll off more smoothly. It would appear that each of the

MAC bandwidth values begin to level off at a constant value when the TG is generating packets at 1000 pps. The transition would seem to have started at the 800 pps mark. We can see from Figure 5.52 and Table 5.25, that for the last three experiments (i.e. at 800 pps, 900 pps and 1000 pps), there is only a slight change in each subsequent MAC bandwidth values.

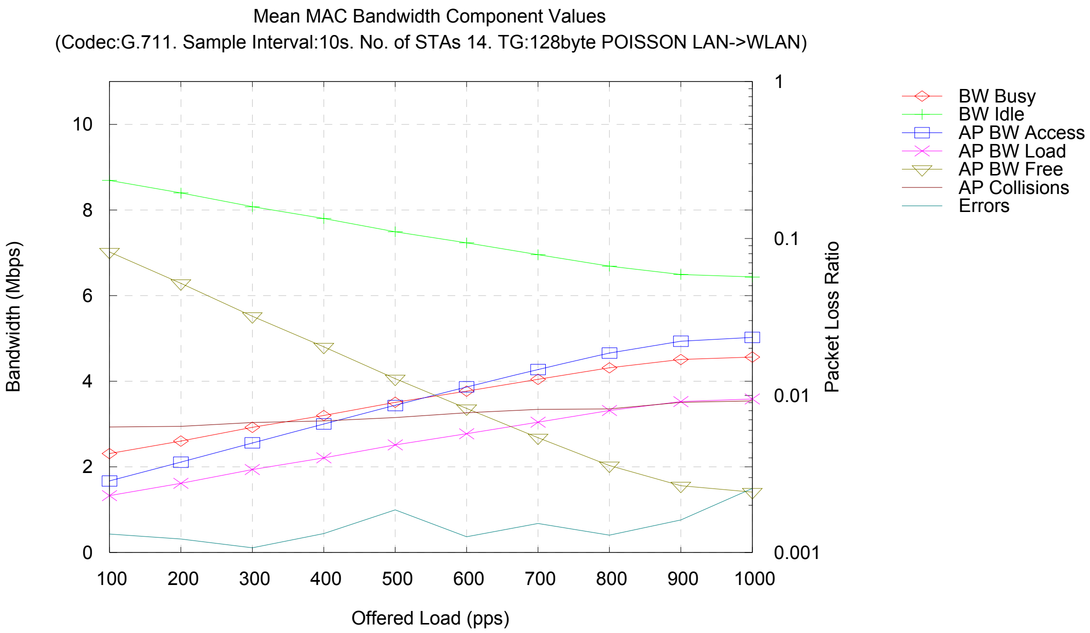


Figure 5.52: AP MAC BW components. Values roll off at 800 pps.

Once again, packet loss is mainly due to collisions. The percentage packet loss at 900 pps due to collisions is approximately 0.8% compared to 0.1% for errors. The packet loss ratios show a decrease in comparison to the previous two experiments using a payload of 512 bytes and 256 bytes.

Offered Load (pps)	BW_{Busy}	BW_{Idle}	AP BW_{Access}	AP BW_{Load}	AP BW_{Free}
100	2.31	8.69	1.67	1.33	7.02
800	4.31	6.69	4.66	3.32	2.03
900	4.51	6.49	4.93	3.52	1.56
1000	4.56	6.43	5.02	3.58	1.41

Table 5.25: Summary of AP MAC BW values in Mbps.

The STA BW_{Access} and BW_{Load}, as shown in Figures 5.53(a) and (b), show similar but less pronounced characteristics compared to the previous two experiments. The BW_{Access} would appear to reach a maximum when the offered load is at 800 pps, before dropping slightly at 900 pps and 1000 pps. The station BW_{Load} appears constant over all TG packet rates.

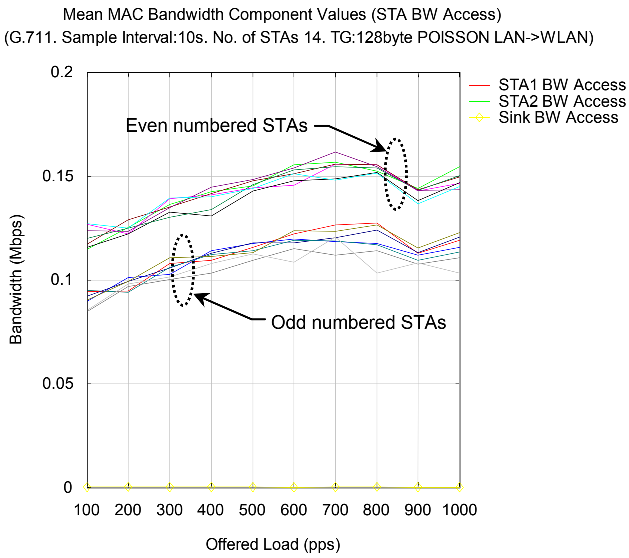


Figure 5.53(a): STA BW_{Access}. Turning point at 800 pps.

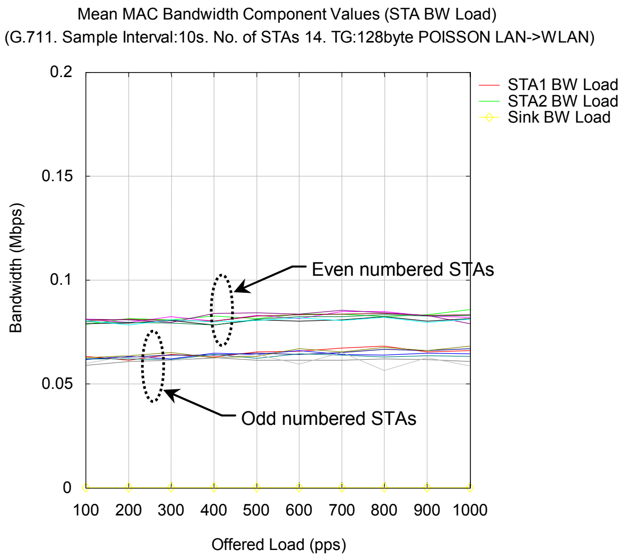


Figure 5.53(b): STA BW_{Load} values

Figure 5.54(a) compares the STA BW_{Free} over all offered loads and displays it in relation to the AP MAC bandwidth components. We can notice a significant difference between the BW_{Free} as seen by a station compared to the BW_{Free} as seen by the AP. Finally, Figure 5.54(b) compares the results for STA 1 and STA 2 as the offered load is increased. When we compare these results to the previous two experiments (Figures 5.48 and 5.51(b)) we notice that the average BW_{Free} at the cut-off point is increasing as the payload decreases. A quick comparison of these figures will show us that for payloads of 512 bytes the STA BW_{Free} < 6 Mbps, for payloads of 256 bytes the STA $BW_{Free} \approx$ 6 Mbps and for payloads of 128 bytes the STA BW_{Free} > 6 Mbps.

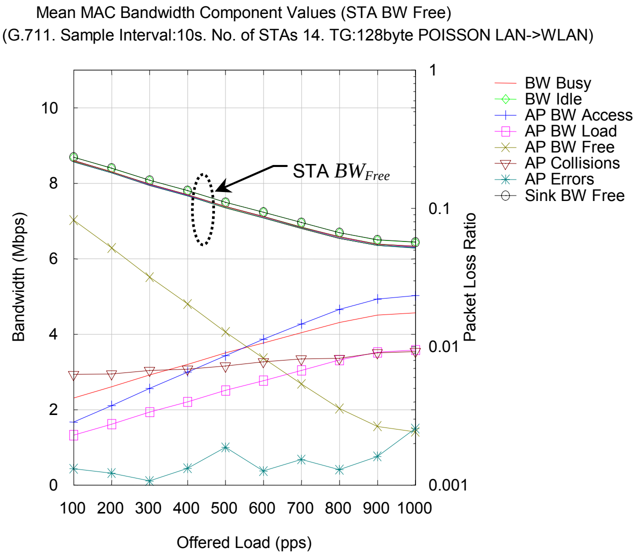


Figure 5.54(a): STA $BW_{Free} \approx$ 6.55 Mbps at 800 pps.

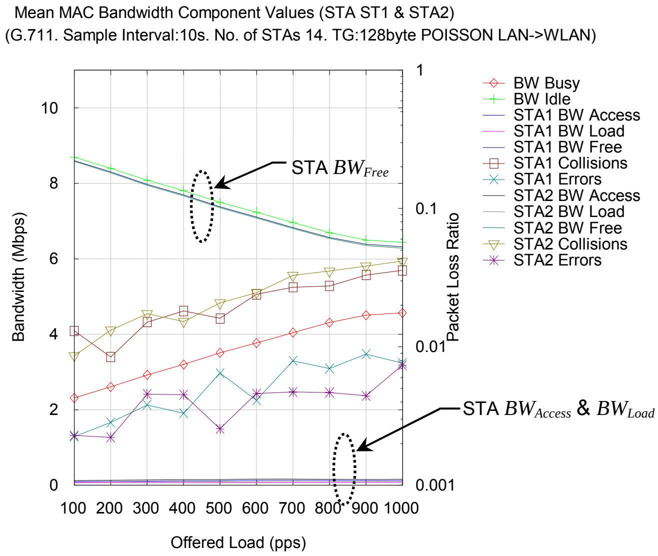


Figure 5.54(b): Comparison of STA1 & STA2.

Table 5.26 compares the results for STA 1 and STA 2. The table shows the results for the last three tests as there is no clear cut-off point as seen in previous experiments with larger payloads. The results indicate the packet loss rate increases as the TG packet rate is increased. Again collisions account for the bulk of losses.

Offered Load (pps)	STA 1					STA 2				
	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors	BW_{Access}	BW_{Load}	BW_{Free}	Collisions	Errors
100	0.09	0.06	8.60	0.01	0.00	0.11	0.08	8.58	0.01	0.00
800	0.13	0.07	6.56	0.03	0.01	0.15	0.08	6.54	0.03	0.00
900	0.11	0.06	6.38	0.03	0.01	0.14	0.08	6.35	0.04	0.00
1000	0.12	0.07	6.32	0.04	0.01	0.15	0.09	6.28	0.04	0.01

Table 5.26: Summary of STA MAC BW values in Mbps.

5.7.2 AP Delay Characteristics

In the previous experiment (Section 5.6) we tested the effects on the AP delay times by stressing the servicing mechanism. This time when we introduce background traffic we attempt to overload the buffer. We do this by generating

UDP packets from a wired source and directing them to a wireless sink via the AP. By doing this the AP downlink buffer contains both VoIP packets as well as packets from the TG. Once again the payload and rate of the traffic is changed in order to compare results.

TG with 512 byte Payload:

Figure 5.55 shows a PDF of delay times for STA 1 when a TG payload of 512 bytes was used. The offered load was increased in steps of 100 pps starting at 100 pps. We can see from the figure that as the offered load increases the peak value decreases and values become more spread out (i.e. the variance and standard deviation increase). We can also see that STA 1 only shows results from 100 pps to 900 pps. When the experiment was conducted at 1000 pps no data was captured for stations 1 to 12. Only stations 13 and 14 yielded results at 1000 pps and indicated delays greater than those found at 900 pps.

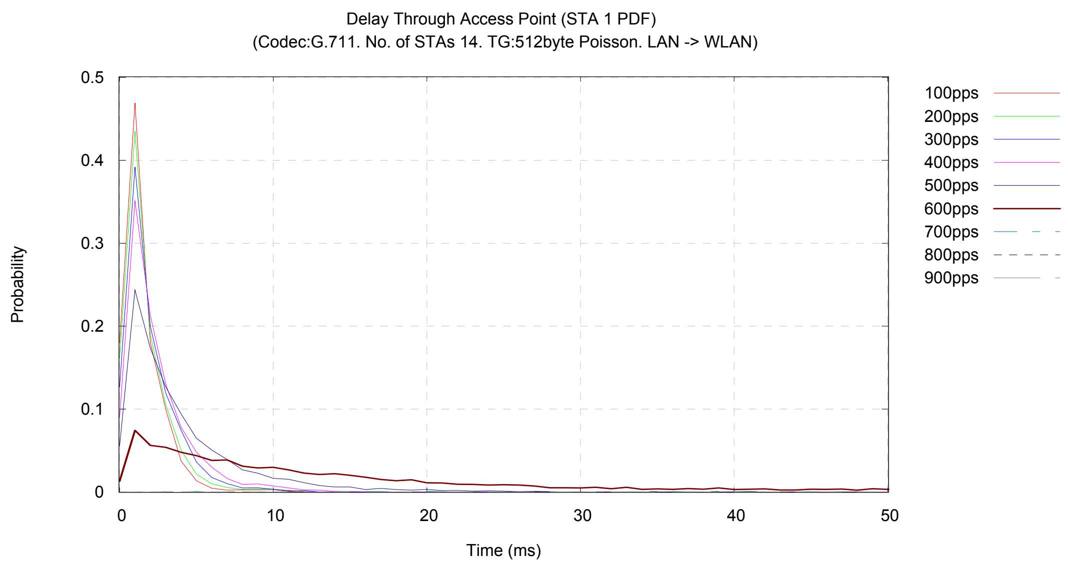


Figure 5.55: STA 1 PDF of AP delay times. Increase in variance as offered load is increased.

The CDF results of packet delays through the AP for STA 1 are shown in Figure 5.56. We can clearly see that at 600 pps and above, the station shows a marked increase in delay times. Results for packet rates below 600 pps all appear to exhibit relatively smaller delays in comparison. This can be observed for all stations using a 512 byte payload. As such we will concentrate on the lower end of the delay time scale. We will also be analysing the same four stations (i.e. STA 1, 2, 13 and 14) in detail as we did for the WLAN-to-LAN experiments.

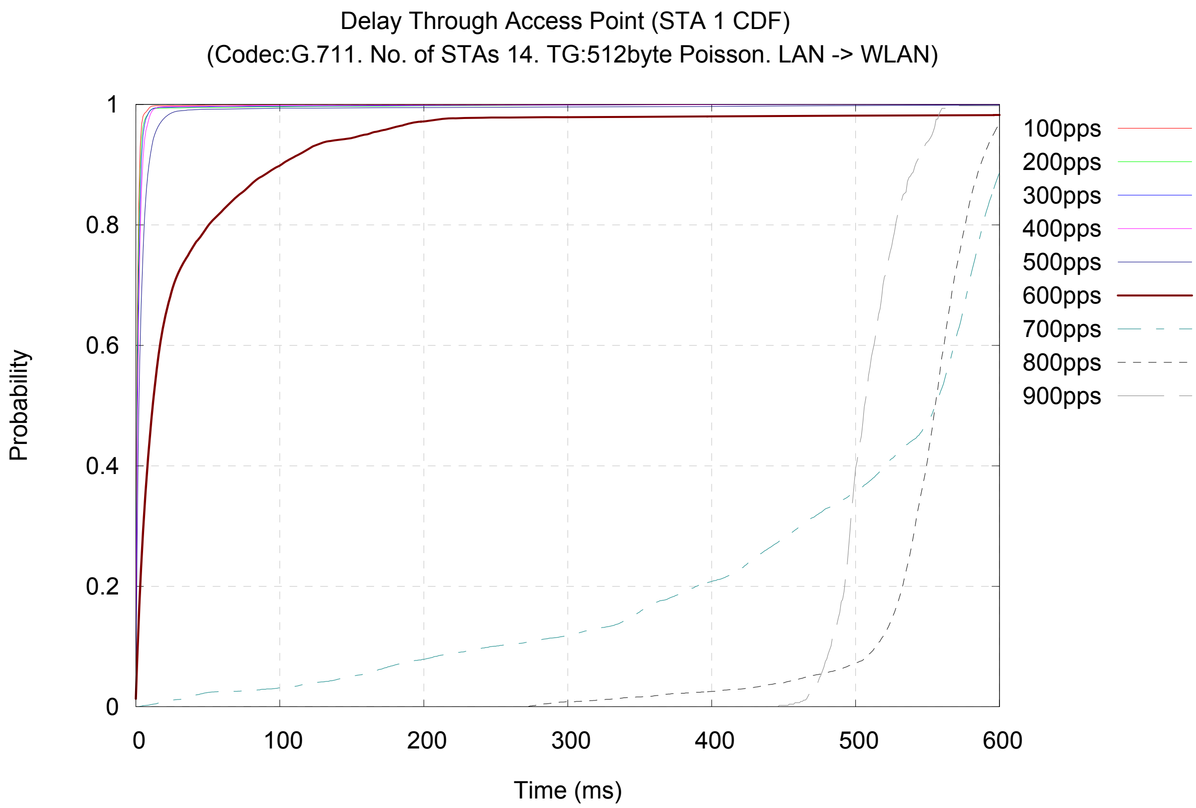


Figure 5.56: STA 1 CDF for increased TG rate (LAN-to-WLAN, 512 byte).

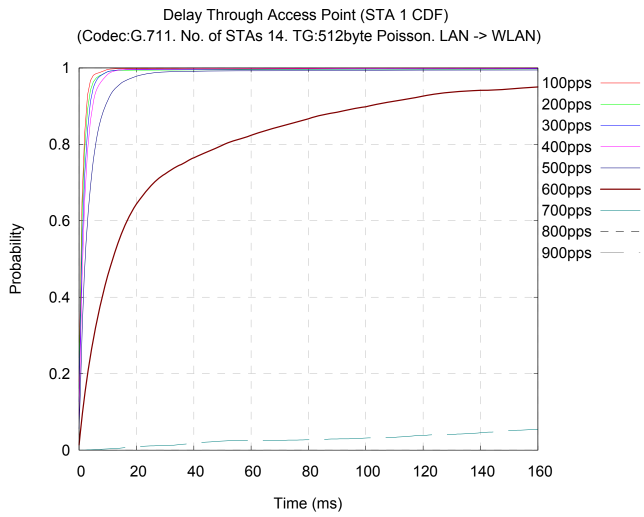


Figure 5.57(a): STA 1 CDF for increased TG rate.

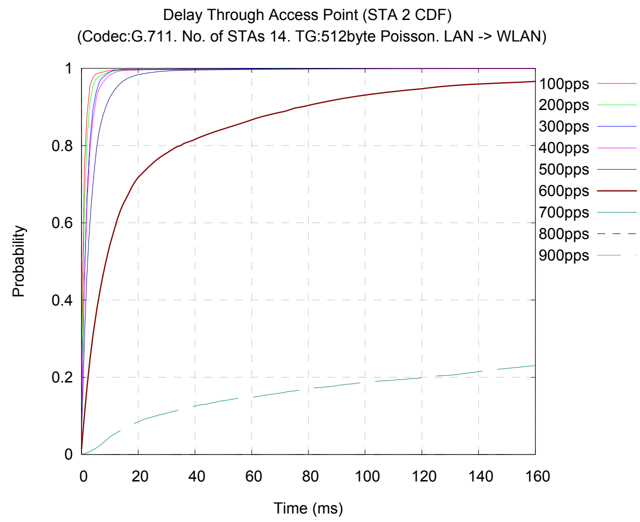


Figure 5.57(b): STA 2 CDF for increased TG rate.

We can see from Figures 5.56(a) and (b) that when the TG is producing packets at rates below 600 pps, nearly all of the VoIP packets for these stations are held up in the AP buffer for less than 40 ms. STA 13 and STA 14 behave similarly. However, for both stations (i.e. STAs 13 and 14) the probability of delay being less than 10 ms at 100 pps and 200 pps is greatly reduced. This is also true for STA 13 when a TG rate of 300 pps is utilized. This can be seen in Figures 5.57(a) and (b) by the kink in the curves at these rates.

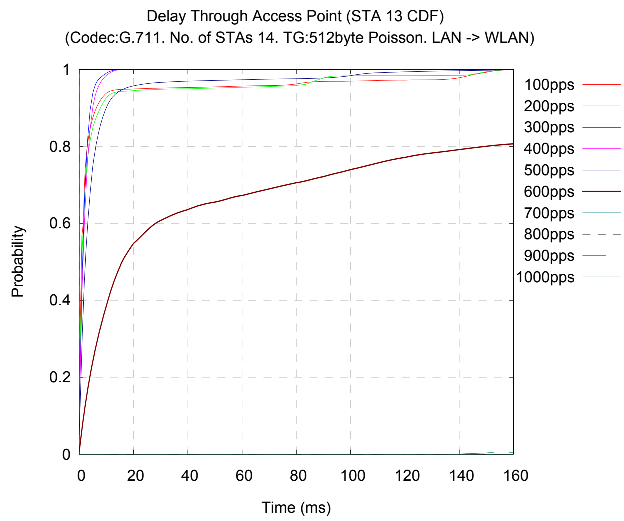


Figure 5.58(a): STA 13 CDF of AP transit times.

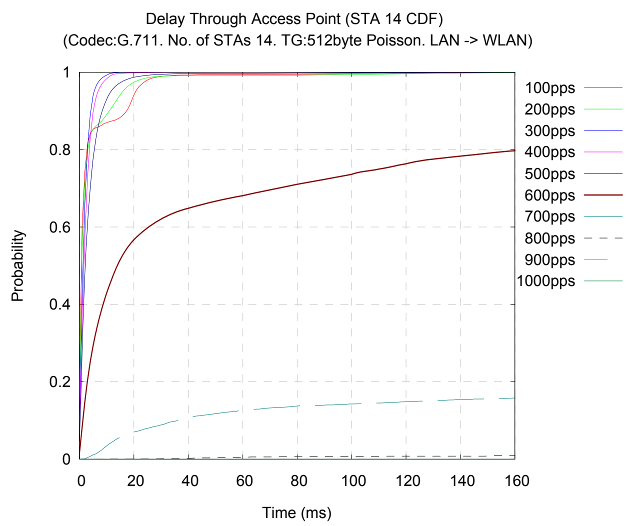


Figure 5.58(b): STA 14 CDF of AP transit times.

TG with 256 byte Payload:

When a payload of 256 bytes is used there is an increase in the station delay performance. We can clearly see from Figures 5.58(a) and (b) that the probability of delay at 600, 700 and 800 pps has greatly increased compared to using a 512 byte payload. It would also appear that the probability of stations having delays less than 25 ms has greatly increased over the previous experiment.

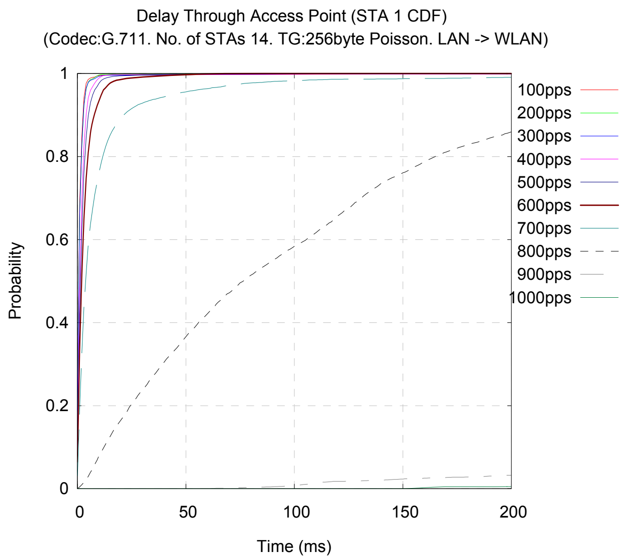


Figure 5.59(a): STA 1 CDF of AP delay.

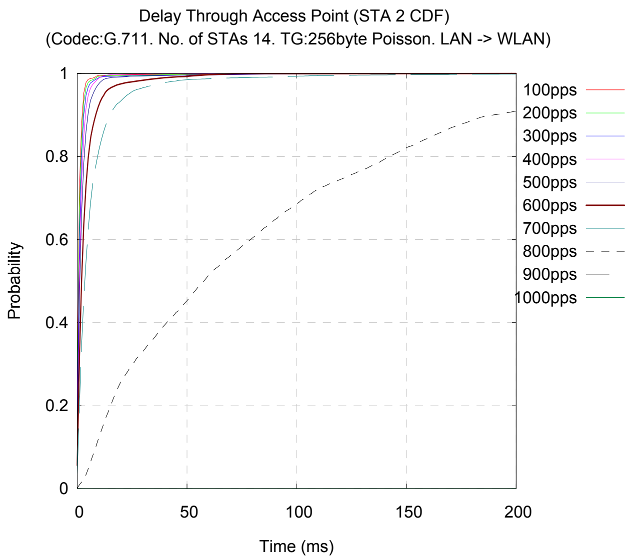


Figure 5.59(b): STA 2 CDF of AP delay.

Stations 13 and 14 continue to exhibit longer delays compared to all other stations. This is particularly evident at larger packet rates and payloads. In the case of our experiment STA 14 suffers larger delays than STA 13. This is most likely due to minor differences in the characteristics of each speech sample. For STA 13 there is a very obvious kink in the graph shown in Figure 5.59(a). This occurs at rates of 400, 500 and 600 pps. It would appear that packets are being cleared relatively efficiently before encountering larger delays. At some point the AP then clears these packets.

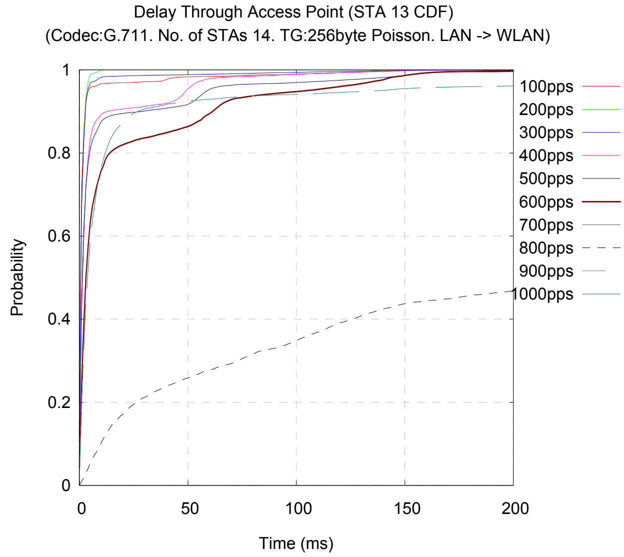


Figure 5.60(a): STA 13 CDF of AP delay.

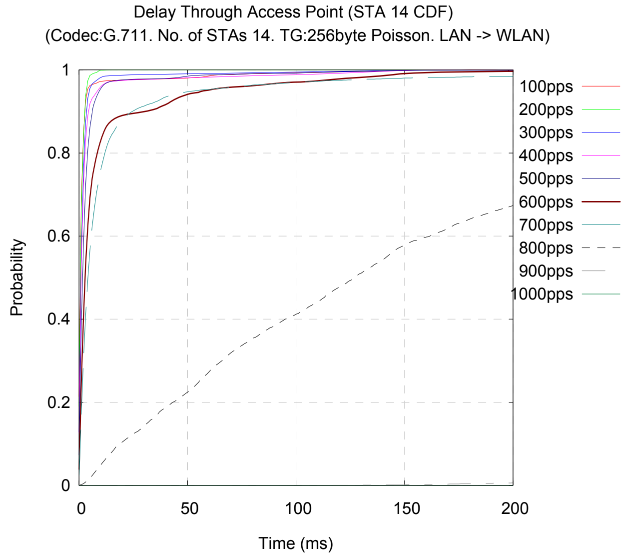


Figure 5.60(b): STA 14 CDF of AP delay.

TG with 128 byte Payload:

Finally, when we reduce the TG payload to 128 bytes we can notice a significant difference in station delay times through the AP. It is very obvious from Figures 5.60(a) and (b) that when the TG generates packets above 800 pps, VoIP packets from a station will begin to suffer increasingly larger delays. It is also apparent from these figures (and indeed all of the CDF of delay time graphs presented) that the main cluster of curves, with a knee point usually positioned in the top left corner which indicates that there is a high probability of station delay times through the AP being relatively small (typically under 10 ms).

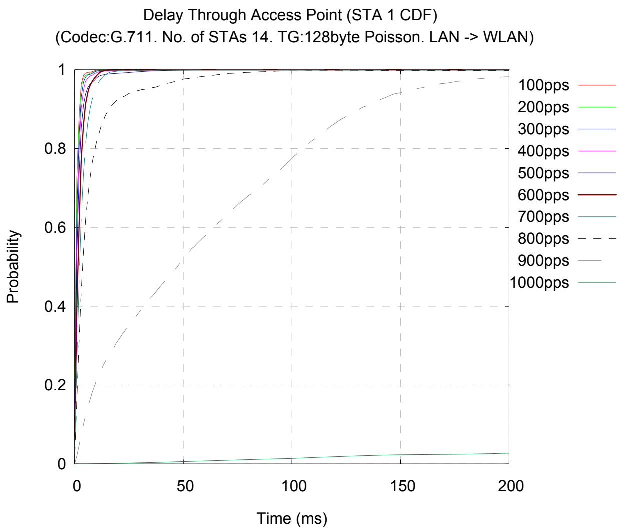


Figure 5.61(a): STA 1 CDF of AP delay.

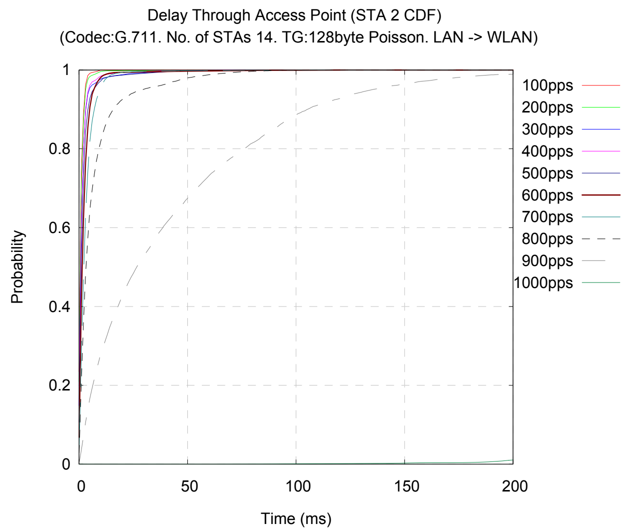


Figure 5.61(b): STA 2 CDF of AP delay.

It would appear that even at a lower payload of 128 bytes, STA 13 and STA 14 continue to underperform when compared to all other stations. This can clearly be seen from Figures 5.61(a) and (b). However, there does not seem to be a consistency in the performance due to the TG packet rate used. One might expect the performance to decline as the TG packet rate is increased. This is clearly not the case as we can see that the probability of delays under 25 ms is greater at 600, 700 and 800 pps than it is at lower packet rates.

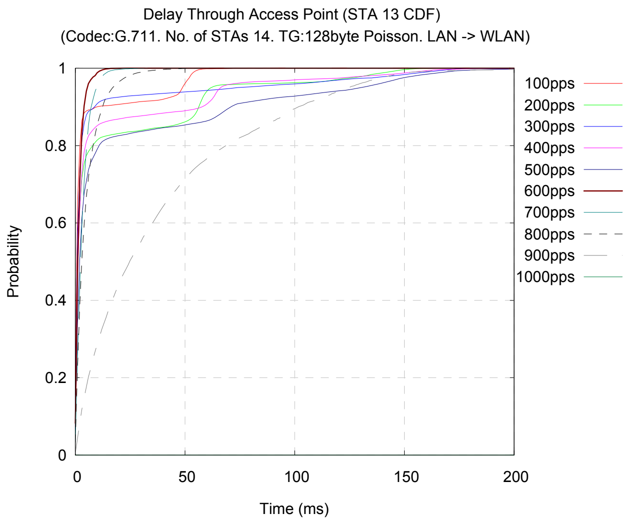


Figure 5.62(a): STA 13 CDF of AP delay.

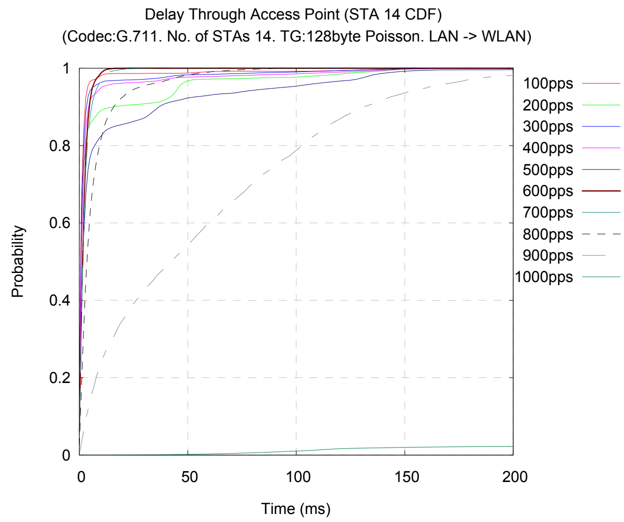


Figure 5.62(b): STA 14 CDF of AP delay.

To summarise this section we compare the statistics for each of the four stations previously discussed. We do this by analysing the results for probabilities under 10 ms for each payload as the packet rate is increased. We can also compare the mean delay values as well as the standard deviation of delay. This allows us to quickly determine the point when the TG began to make a marked impact on the AP. We can see when large delays begin to manifest themselves and compare these to previous results

	STA 1								
	P(D ≤ 10ms)			AP Delays (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.99620	0.99507	0.99858	1.62	3.35	1.88	10.68	1.25	1.46
200	0.98976	0.99656	0.99799	2.73	16.26	1.57	2.79	1.31	1.56
300	0.99017	0.99361	0.99622	2.78	14.13	2.04	11.82	1.55	1.57
400	0.98311	0.98944	0.97058	3.03	13.14	2.69	12.48	1.80	2.58
500	0.91455	0.97712	0.97898	6.67	33.89	2.80	4.16	2.32	3.91
600	0.45694	0.93566	0.98483	52.46	176.69	3.87	5.37	2.57	2.30
700	0.00320	0.76262	0.95629	491.51	167.95	14.47	60.72	3.63	3.60
800	0.00000	0.08003	0.81610	548.23	48.18	117.57	187.71	8.58	20.23
900	0.00000	0.00000	0.21211	509.86	22.61	485.13	61.67	60.43	54.44
1000	NA	0.00000	0.00049	NA	NA	469.57	42.90	449.47	96.97

Table 5.27: Comparison of delay statistics for STA 1.

	STA 2								
	P(D ≤ 10ms)			AP Delays (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.99493	0.99549	0.99921	1.40	3.54	1.46	4.03	1.21	1.18
200	0.99115	0.99342	0.99611	1.80	4.26	1.70	4.78	1.43	2.78
300	0.99050	0.99024	0.97720	2.35	3.63	2.06	5.88	2.18	4.95
400	0.98214	0.98952	0.96178	2.94	11.01	2.08	2.80	1.93	3.62
500	0.92473	0.97717	0.97234	4.32	6.35	2.79	5.30	2.58	5.05
600	0.54226	0.92622	0.98103	29.11	61.17	4.38	7.27	2.85	4.97
700	0.04527	0.81841	0.95541	379.88	214.75	7.76	15.68	3.73	5.35
800	0.00000	0.12569	0.81918	558.54	90.65	107.89	210.99	7.55	11.78
900	0.00000	0.00000	0.28099	509.71	25.68	478.33	76.51	46.00	78.38
1000	NA	0.00000	0.00000	NA	NA	473.53	55.28	452.77	103.93

Table 5.28: Comparison of delay statistics for STA 2.

	STA 13									
	P(D ≤ 10ms)			AP Delays (ms)						
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte		
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.93830	0.96647	0.90051	7.59	25.96	3.63	15.89	5.68	13.83
200	0.92629	0.99913	0.81462	7.53	23.60	1.44	1.33	13.71	29.76
300	0.99017	0.98233	0.90885	2.28	2.03	2.74	11.84	9.46	29.79
400	0.98449	0.89240	0.81016	2.59	2.39	7.70	18.77	12.16	29.57
500	0.90048	0.87538	0.79092	7.08	18.47	10.64	28.51	19.00	40.21
600	0.38585	0.75731	0.98874	140.27	307.06	18.34	42.05	2.19	2.06
700	0.00000	0.76308	0.95410	750.79	389.94	46.10	207.67	3.75	3.25
800	0.00000	0.10189	0.85414	601.29	173.77	550.21	601.07	5.57	6.17
900	0.00000	0.00000	0.28643	529.14	26.34	530.53	181.07	39.52	40.05
1000	0.00000	0.00000	0.00000	549.62	51.88	484.66	90.64	486.32	156.86

Table 5.29: Comparison of delay statistics for STA 13.

	STA 14								
	P(D ≤ 10ms)			AP Delays (ms)					
Offered Load (pps)	TG Payload (bytes)			512 byte		256 byte		128 byte	
	512	256	128	Mean	S.D.	Mean	S.D.	Mean	S.D.
100	0.86988	0.97244	0.98091	4.40	10.70	3.25	13.50	2.80	12.35
200	0.89085	0.99913	0.89052	4.21	12.56	1.43	1.35	7.90	22.25
300	0.99270	0.98201	0.96381	2.30	1.95	2.96	11.57	3.93	15.39
400	0.98314	0.95985	0.91773	2.78	2.49	4.29	15.58	5.07	18.55
500	0.92985	0.95505	0.82542	4.27	7.21	4.60	14.00	13.26	31.91
600	0.42914	0.83089	0.98792	174.80	365.04	11.88	28.87	2.54	2.17
700	0.03431	0.76010	0.97013	550.36	349.94	21.92	107.43	2.93	2.98
800	0.00021	0.04919	0.81678	574.52	161.51	402.31	577.23	7.55	11.09
900	0.00000	0.00000	0.23859	533.39	72.20	473.44	75.24	57.67	54.60
1000	0.00000	0.00000	0.00000	548.45	43.31	477.04	57.32	440.35	73.23

Table 5.30: Comparison of delay statistics for STA 14.

5.7.3 Delay Distribution Tails

TG with 512 byte Payload:

We now analyse the tails of the distribution of delay times. When we look at the results for STA 1 using a 512 byte payload we can observe some interesting results. In Figures 5.62(a) and (b) we can see that for an offered load below 600 pps, packets are serviced relatively quickly and result in a graph displaying an almost vertical line for the "straight" part of the curve. This can be backed up by the probability of delays as discussed in the previous section. We can also observe a very different set of events taking place above 600 pps. The graphs show that at increased offered loads, the packets are serviced at different rates.

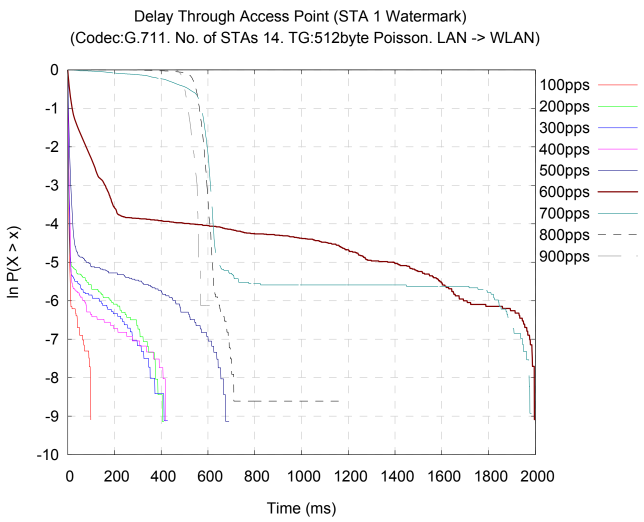


Figure 5.63(a): STA 1 Watermark plots.

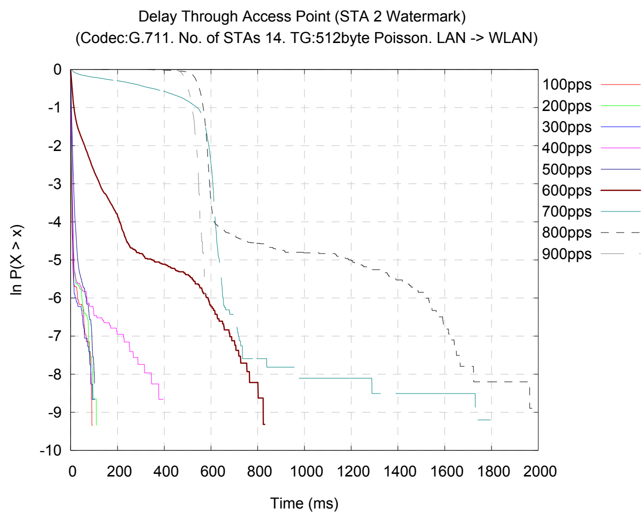


Figure 5.63(b): STA 2 Watermark plots.

As seen in the MAC bandwidth and delay plots it would appear that 600 pps is the cut-off point again. At 600 pps the packets are beginning to experience longer delays as well as decreased rate functions (i.e. the magnitude of the slope is increasing). At rates above 600 pps, VoIP packets are virtually un-serviced, or at least very few packets are serviced before a burst of packets is processed. This can be seen by a very small slope (almost zero) at the beginning of the curve before a dramatic increase to a large slope. Figures 5.63(a) and (b) display similar results for STA 13 and STA 14. As before these two stations experience longer delays. This is most obvious at higher TG packet rates. Notably, at 600 pps, the stations do not experience the dual service rate as seen in Figures 5.62(a) and (b). Instead however, the slope for these two stations is very well defined.

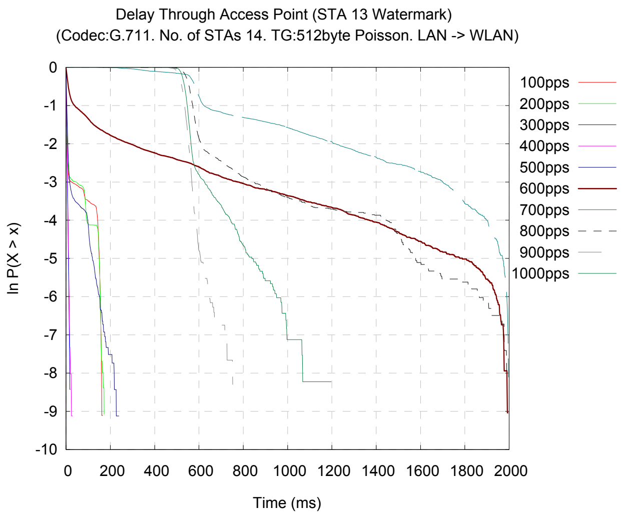


Figure 5.64(a): STA 13 Watermark plots.

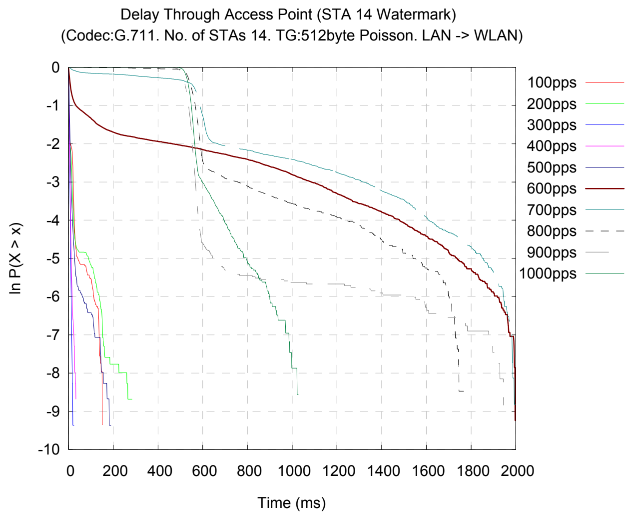


Figure 5.64(b): STA 14 Watermark plots.

Offered Load (pps)	Asymptotic Slope STA 1	STA 2	STA 13	STA 14
100	-0.12	-0.19	-0.01	-0.13
200	-0.15	-0.01	-0.02	-0.15
300	-0.37	-0.01	-0.60	-0.68
400	-0.28	-0.01	-0.45	-0.47
500	-0.07	-0.00	-0.01	-0.14
600	-0.01	-0.01	-0.00	-0.00
700	-0.00	-0.00	0.00	0.00
800	0.00	0.00	0.00	0.00
900	0.00	0.00	0.00	0.00
1000	NA	NA	0.00	0.00

Table 5.31: Comparison of asymptotic slopes (TG payload of 512 bytes).

TG with 256 byte Payload:

With a TG payload of 256 bytes we can again see an improvement in the performance of the stations, with regards to delay times. Most notably we can observe a difference in asymptotic slope for the plots when a rate of 600 pps is utilized. Figures 5.64(a) and (b) show an improvement over the previous experiment using a 512 byte payload. In Figures 5.65(a) and (b) we can see that the slopes for STA 13 and STA 14 are comparable to those in Figures 5.64(a) and (b). However, there is a notable variation in slopes when an offered load of 700 pps is used. The graphs would indicate that STA 13 and

STA 14 experience much longer delays during this experiment compared to all other stations. It can also be seen that these two stations also experience much longer delays at 600 pps.

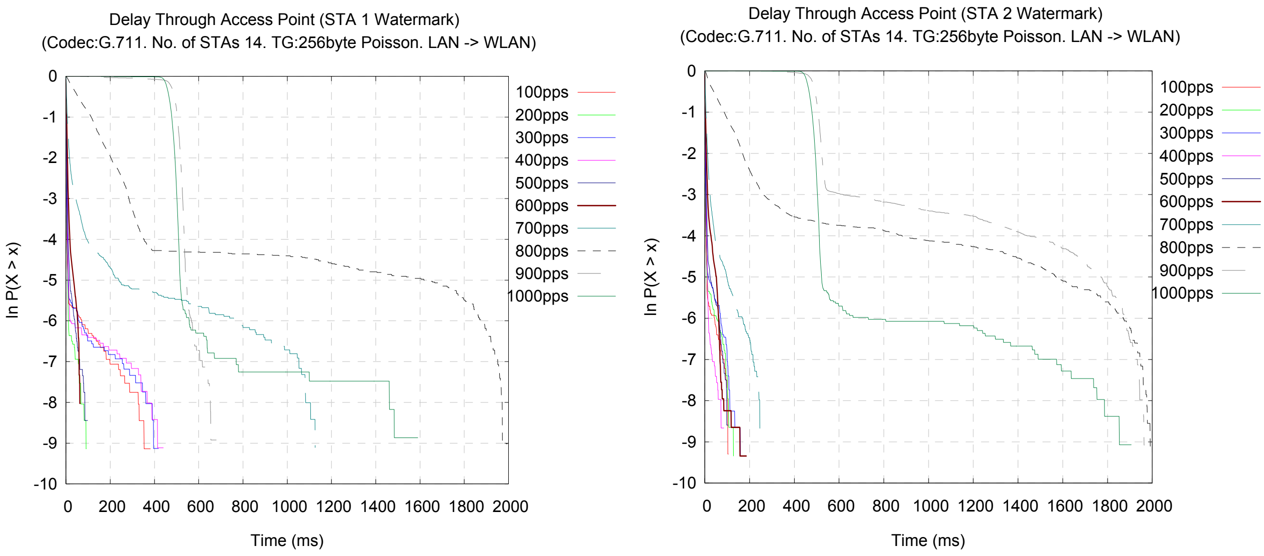


Figure 5.65(a): STA 1 Watermark plots.

Figure 5.65(b): STA 2 Watermark plots.

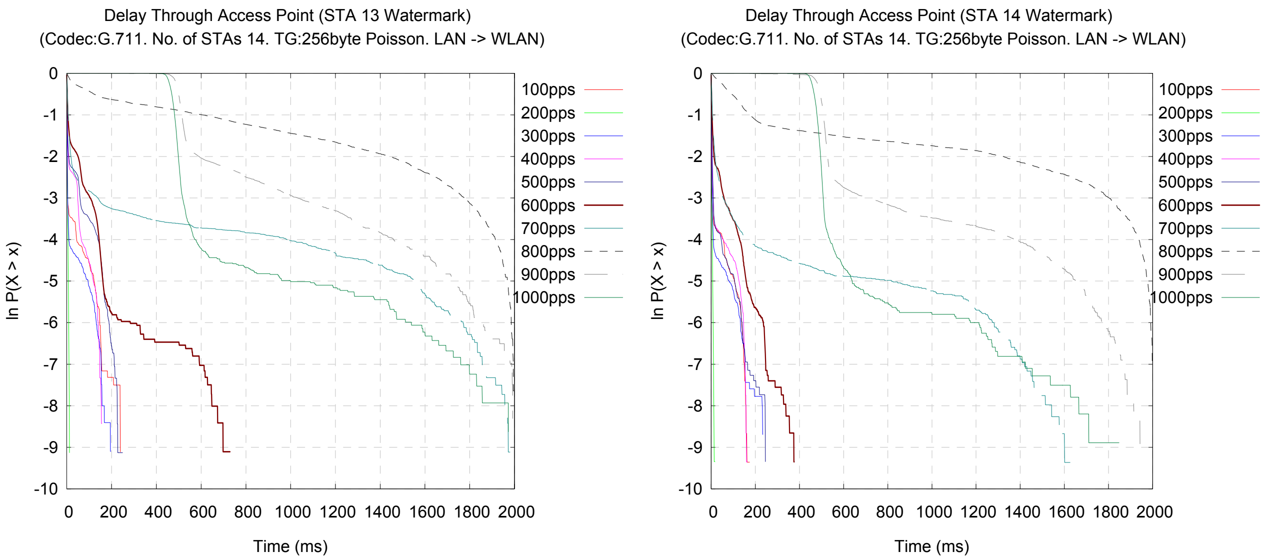


Figure 5.66(a): STA 13 Watermark plots.

Figure 5.66(b): STA 14 Watermark plots.

Offered Load (pps)	Asymptotic Slope STA 1	STA 2	STA 13	STA 14
100	-0.04	-0.12	-0.01	-0.02
200	-0.13	-0.26	-1.00	-0.60
300	-0.11	-0.18	-0.00	-0.01
400	-0.26	-0.38	-0.03	-0.01
500	-0.17	-0.26	-0.02	-0.02
600	-0.04	-0.05	-0.02	-0.02
700	-0.02	-0.02	-0.00	-0.01
800	-0.01	-0.01	-0.00	-0.01
900	0.00	0.00	0.00	0.00
1000	0.00	0.00	0.00	0.00

Table 5.32: Comparison of asymptotic slopes (TG payload of 256 bytes).

TG with 128 byte Payload:

Finally for this section, we now look at the tails of the PDF distribution of delay times for experiments using a payload of 128 bytes. We would expect the delay performance of each station to be greater than the previous experiments at this smaller payload. This is indeed the case. We can see an obvious difference in the characteristics of the curves in Figures 5.66(a) and (b) at higher TG packet rates when compared to the watermark plots in the previous two sections experiments (namely for 256 byte and 512 byte payloads).

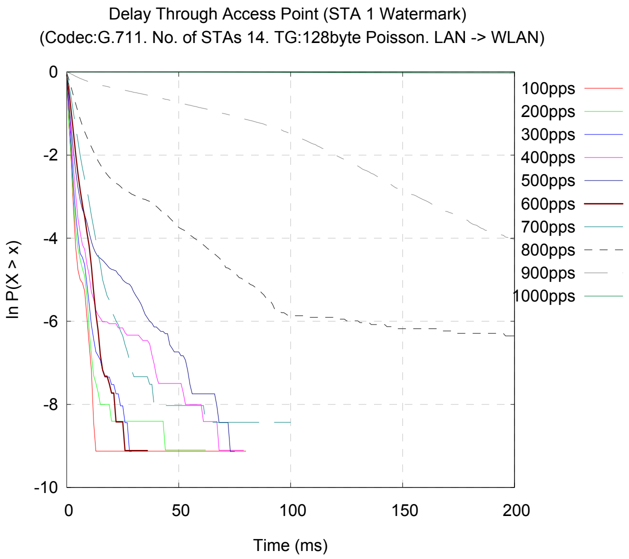


Figure 5.67(a): STA 1 Watermark plots.

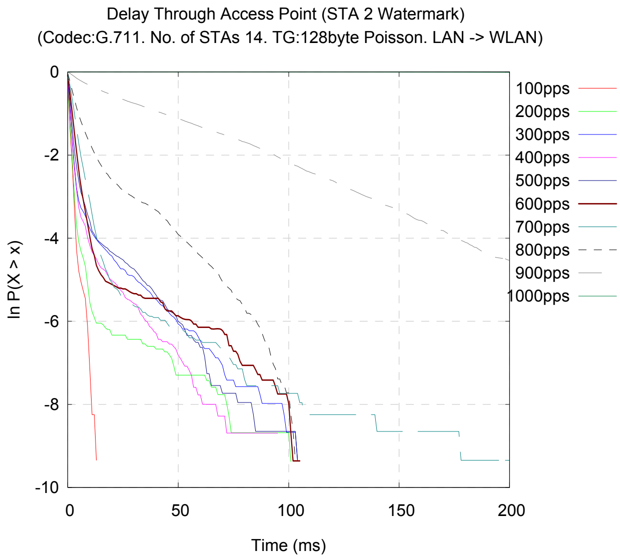


Figure 5.67(b): STA 2 Watermark plots.

For STA 1 and STA 2 we can see that the delay slopes increase toward zero as the TG packet rate increases. However, in Figures 5.67(a) and (b) we can see that this is not always the case. From the figures we can see over the long term, at packet rates of 600, 700 and 800 pps, the slope is comparable to slopes at lower packet rates. In some instances it is even larger (negatively). This is clearly seen for an offered load of 600 pps and 700 pps. In Table 5.33 we present the slopes for each station previously mentioned in order to determine the rate function.

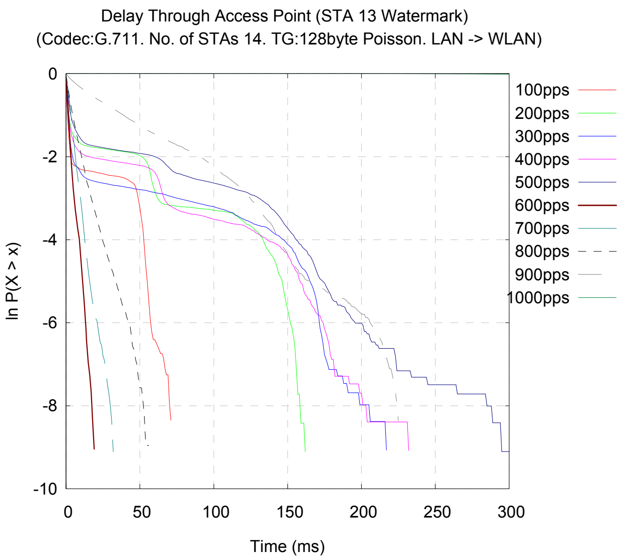


Figure 5.68(a): STA 13 Watermark plots.

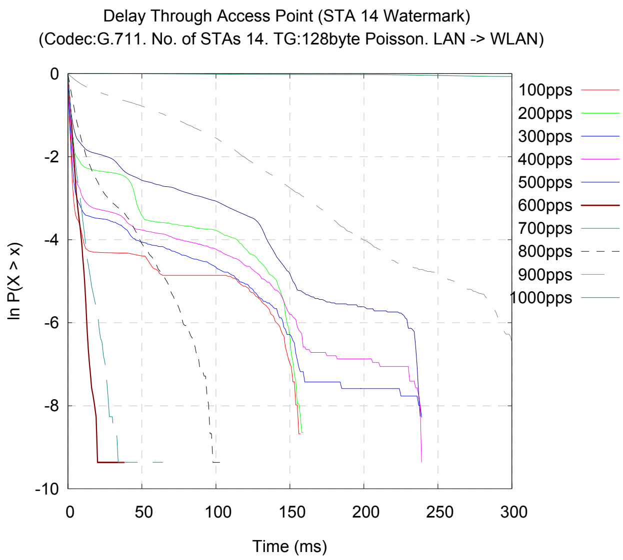


Figure 5.68(b): STA 14 Watermark plots.

Offered Load (pps)	Asymptotic Slope STA 1	STA 2	STA 13	STA 14
100	-0.55	-0.94	-0.01	-0.01
200	-0.03	-0.55	-0.02	-0.02
300	-0.05	-0.27	-0.01	-0.01
400	-0.07	-0.05	-0.02	-0.01
500	-0.05	-0.06	-0.01	-0.01
600	-0.04	-0.52	-0.50	-0.57
700	-0.04	-0.26	-0.24	-0.22
800	-0.05	-0.04	-0.12	-0.06
900	-0.02	-0.02	-0.02	-0.02
1000	0.00	0.00	0.00	0.00

Table 5.33: Comparison of asymptotic slopes (TG payload of 128 bytes).

5.7.4 Results Summary

In this section our results show the effects on an AP and contending VoIP users of increasing the background traffic on a wireless medium. This was achieved by varying both the rate of the background traffic as well as the size of the payload. By doing this, we cause the AP buffer to contain both VoIP packets as well as UDP packets from the TG. This has two main consequences. The AP buffer has limited capacity and therefore will fill up more rapidly than experiments without using a TG in this way. Another consequence is that the increased occupancy will increase the overall delay time of a VoIP packet by putting the AP service rate under stress.

Results showed that in general the MAC bandwidth components increase or decrease accordingly as the TG packet rate increases. For a given TG payload the change in MAC bandwidth component was greater than that of experiments with a TG transmitting in the opposite direction (i.e. wireless source and wired sink). For a given payload we could determine an approximate upper limit on the maximum packet rate (defined as the cut-off point when the BW values cease to increase or decrease linearly). This rate increases as the payload decreases. When comparing the same MAC bandwidth components for different TG payloads at the relevant cut-off point, the results show that a reduced payload allows for a larger available BW_{Free} for each STA. The STA BW_{Access} however exhibits a very interesting result. As the TG increases its load past the cut-off point the BW_{Access} begins to decrease rather than level off. This is particularly pronounced when a larger TG payload is used.

Analysis of the AP delays yields some interesting results. In general all stations maintain a delay below 10 ms when the TG packet rate is below 400 pps with a payload of 512 bytes and 256 bytes. However, when we use a payload of 128 bytes this is not the case. There is a lower probability of a delay being under 10 ms for the same packet rate than those with a larger payload. Also, as the packet rate increases the probability of small delays begins to increase before decreasing again.

The Watermark plots revealed that not all stations exhibit similar delays. Once again the same two stations as seen in the previous experiments (VoIP with TG operating from WLAN-to-LAN), experience much longer delays than all other stations. This becomes more exaggerated at larger background traffic rates. The plots also show that all stations experience large delays when a larger payload is used. However, when a 128 byte TG payload is used stations begin to experience large delays in different stages.

5.8 Chapter Summary

Results from tests without background traffic showed that the network was capable of supporting up to 16 VoIP STAs. As the number of STAs increased the BW_{Busy} increased while the BW_{Idle} decreased. As the number of STAs increased from 2 to 16 the BW_{Access} increased owing to the increased contention for access to the medium. This also meant that as the number of STAs increased the access efficiency is reduced due to the fact that the BW_{Load} for each station remains relatively constant. The BW_{Free} decreased as expected as the number of STAs increased. When the number of STAs was at 16 the mean BW_{Free} available for each STA was approximately 8.60 Mbps (or in other words approximately 78% of the line rate). Further analysis of the MAC bandwidth components showed that peak loads occurred during intervals of double-talk. This meant that the mean BW_{Free} available to the AP was greatly reduced. A comparison of the variance of AP MAC bandwidth components as the number of STAs increased showed the variance also increases which would indicate that the network does not enjoy statistical multiplexing. Investigation of AP transit times indicated that while the number of STAs increased the probability of delay also increased. However, results showed that 99% of delays where under 10 ms when 16 STAs connected to the network.

Background traffic was introduced with the number of VoIP STAs kept fixed at 14 while the offered load from a UDP traffic generator was varied using different payloads and packet rates. When background traffic was transmitted from a wireless STA to the distribution system via the AP, the access to the medium was affected owing to the increased contention from the traffic generator. This had a knock on effect for all STAs contending including the AP. As a result, when the packet rate was increased, a saturation point was reached where the network load could not be further increased. The point at which the network became saturated depended on the size of the traffic generator payload. Although a larger payload resulted in network saturation at a lower packet rate, the AP throughput was higher compared to saturation at higher packet rates when using smaller payloads. The upshot of this is that increased throughput can be achieved by using a larger payload due to the reduced protocol overhead incurred. By allowing for a delay budget of 10 ms for the AP transit time a comparison of STA performance was carried out. Analysis showed that the probability of delay being less than 10 ms decreased as background traffic was increased. Investigation of delays also showed that the standard deviation of mean delay times had increased significantly at packet rates below the saturation point. This was also reflected in the probability of delay. Finally, we also noted that as the traffic generator increased its offered load the AP service rate was decreased. This could be seen by a decrease in the magnitude of the asymptotic slope of the tails of the probability of delay distribution.

When the traffic generator was transmitting in the opposite direction (i.e. wired source to a wireless sink via the AP) we observed that the maximum throughput at saturation was increased. However, this also had the effect of having larger delays and increased PLR. As the AP buffer contained both VoIP and background traffic the AP service rate was stressed. This resulted in a marked increase in PLR and delays well below the saturation point. This implies that even though there is an increased throughput at saturation on the downlink, the delay and PLR have increased significantly in order to prevent QoS. Therefore even though the BW_{Free} for a station is approximately 5.60 Mbps at saturation (with an offered load of 512 bytes) results would suggest that a value somewhere between this and 8.60 Mbps (i.e. BW_{Free} when traffic generator is not present) is required in order to maintain QoS. Consequently it can be seen that the downlink from the AP acts as a bottleneck and will ultimately determine the performance of the network.

Over the course of experimentation some anomalous readings were discovered. Throughout the entire testing phase two STAs continually underperformed compared to all other stations. Extensive testing was carried out in order to determine the cause of this behaviour. The anomalous readings were discovered when analysis of delays where performed. Furthermore, the behaviour appeared to become more pronounced when the network traffic and offered load was increased. Another irregularity was found when analysing the delay characteristics. It would appear that the delays do not increase monotonically with increased offered load. A possible reason for this may be that at certain packet rates and packet sizes some form of temporal synchronization occurs between the traffic generator and the contention mechanism. As a result delays at higher packet rates may match the performance at lower packet rates.

6 Summary and Conclusions

In this thesis a set of experiments were conducted in order to investigate the effects on the capacity of an IEEE 802.11b WLAN network when the number of VoIP users is increased and also when background traffic is introduced. The main objectives were to measure the MAC bandwidth components up to the point of network saturation and also to analyse the buffer dynamics with respect to the AP transit times. In particular we were concerned with the service rate of the AP buffer which may result in large delays and also the possibility of packets being lost due to buffer overflow. Each set of experiments used recorded voice samples based on guidelines given in [43]. The codec used for transmission was the G.711 codec with a 30 ms frame duration.

The experimental testing scenarios can be divided up into two main categories; tests without background traffic and tests with background traffic. Tests involving background traffic can then be further divided depending on the origin of the background load. When conducting tests without background traffic the number of VoIP stations was increased from 2 to 16. At each stage MAC bandwidth components and AP transit times were analysed. A similar line of investigation was followed for tests involving background traffic. However, the number of VoIP stations was kept fixed at 14 while the background traffic was varied. This was achieved by using a traffic generator (TG) that allowed both the packet rate and the packet size to be independently adjusted. The TG was initially set-up with a wireless source and a wired sink while data was recorded. The set-up was then reversed (i.e. wired source and a wireless sink) and the tests were repeated.

Over the course of experimentation the following main observations were made.

- An IEEE 802.11b network is capable of supporting at least 16 VoIP stations (using the G.711 codec with 30 ms frame duration).
- The AP downlink will dictate the network capacity. This is due to the limitations of the 802.11 MAC mechanism i.e. the CSMA/CA.
- The network does not experience a statistical multiplexing gain. This is indicated by the variance of the MAC bandwidth components increasing as the number of stations is increased.
- Use of the Watermark plot method indicates that the AP service rate is decreased as the network load increases which may result in large delays and the potential for buffer overflow.
- Protocol overhead limits the capacity of the network. For VoIP packets the protocol overhead is almost equal to the transmitted payload.
- Large peaks in the aggregate load occur during periods of station double-talk. This results in reduced BW_{Free} and hence reduced capacity.
- Higher throughputs can be achieved by using larger payloads.

In addition to the main findings we also observed some anomalous results. These included the increased packet delays of two particular VoIP stations i.e. STA 13 and STA 14. Despite extensive testing and investigation we could not determine the cause of this unusual behaviour. Further research may establish the source of this potentially problematic behaviour. We also found that when analysing the AP transit times we were not observing a monotonic increase in the delay with increased TG packets rate. This may be due to some form of temporal synchronization occurring between streams for a given TG packet rate and packet size.

There are also some limitations that should be pointed out concerning the experimental set up. Gathering and sorting data in order to extract the timing information proved to be a time consuming procedure. This was due to very large files (approximately 100 MB) requiring processing. The large files resulted from the data from all 16 stations being gathered into just two files (i.e. an uplink and a downlink file). Processing these files could take in excess of 10 hours for a single 10 minute capture on a 2.7 GHz Pentium 4 PC. As a result of this the tests were limited to using only one codec, namely the G.711 codec with a 30 ms frame duration. Further research may include the use of various other audio codecs including low bit rate codecs and ideally to identify the optimal codec for VoIP over WLAN. Tests may also be carried out by setting up wireless VoIP stations connected to wired VoIP stations via an AP. This may be a more realistic set-up, as it is unlikely that a VoIP call will take place between two users connected to the same AP (due to their relatively close proximity).

Further research may also include the use of an 802.11a and 802.11g networks. Owing to their increased popularity and availability it would seem obvious to conduct VoIP experiments across these networks. The higher line rates of these networks (54 Mbps or 108 Mbps in turbo mode) would enable a larger number of stations to access the medium. In order to conduct similar experiments as presented in this thesis greater resources would be required.

As such it may be advantageous to conduct experiments using computer simulations of an 802.11a/b/g network. Such simulations may be carried out using the network simulator *ns2* [69] or alternatively *Opnet* [70]. This would enable us to quickly configure the network architecture to meet our requirements. As well as this, it would be worthwhile to compare these results to tests using artificial voice and utilising an RTP packet generator such as *RTP Tools* [71]. RTP tools would enable us to automate much of the experimental procedure. A feature of using RTP tools is that the data for each station is gathered on the PC it is operating on. This has the advantage of reducing the size of the file that requires processing as well as separating each stations data automatically.

Finally, a combination of the above factors may be used in future work for studying the behaviour of VoIP when using IEEE 802.11e. IEEE 802.11e is a MAC enhancement for supporting QoS provisioning and is intended for real-time services such as VoIP telephony and video streaming. The 802.11e MAC enhancements provide the ability to classify data into prioritised classes, i.e. to allow VoIP streams access to the medium with a higher priority compared to best-effort background traffic.

Bibliography.

[1] Forfás, *Wireless Communications: An Area of Opportunity for Ireland: A report by Forfás to the Minister of Enterprise, Trade and Employment,* Ireland, April 2004.

[2] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band", *IEEE Std 802.11b,* 1999 Edition (R2003).

[3] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz Band", *IEEE Std 802.11a,* 1999 Edition (R2003).

[4] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Further Higher Data Rate Extension in the 2.4 GHz Band", *IEEE Std 802.11g,* 1999 Edition (R2003).

[5] D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b Wireless LAN Supporting VoIP", in *Proc. IEEE Int. Conference on Communications (ICC),* 2004.

[6] M. Coupechoux, V. Kumar, and L. Brignol, "Voice over IEEE 802.11b Capacity", in *Proceedings of the 16th ITC Specialist Seminar*, 2004, pp 1-8.

[7] K. Medepalli, P. Gopalakrishnan, D. Famolari and T. Kodama, "Voice capacity of IEEE 802.11b and 802.11a WLAN systems in the presence of channel errors and different user data rates", *Proceedings of IEEE VTC*, 2004.

[8] A. Lindgren, A. Almquist, and O. Schelén, "Evaluation of Quality of Service Schemes for IEEE 802.11Wireless LANs", in *Proceedings of the 26th Annual IEEE Conference on Local Computer Networks (LCN 2001)*. November 2001.

[9] D. Deng, and H. Yen, “Quality-of-Service Provisioning System for Multimedia Transmission in IEEE 802.11 Wireless LANs”, *IEEE Journal on Selected Areas in Communications (special issue on Mobile Computing and Networking)*, Vol. 23, No. 6, pp. 1240-1252, June 2005.

[10] IEEE, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, *IEEE Std 802.11,* 1997 Edition (R2003).

[11] IEEE, “IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture”, *IEEE Std 802*, 2001.

[12] ISO/IEC, “Information Technology – Open Systems Interconnection - Basic Reference Model: The Basic Model”, *ISO/IEC 7498-1*, 1994.

[13] InteropNet Labs White Paper, “*802.11 Alphabet Soup”*, InteropNet Labs Full Spectrum Security Initiative, May 2005.

[14] IEEE, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)”, *IEEE Std 802.11e/D8.0,* February 2004.

[15] IEEE, “IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation”, *IEEE Std 802.11f*, 2003.

[16] IEEE, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Security Enhancements”, *IEEE Std 802.11i,* 2004.

[17] 3Com White Paper, “Deploying 802.11 Wireless LANs”, 2003.

[18] J. Rosenberg et al, “SIP: Session Initiation Protocol”, RFC 3261, June 2002.

[19] ITU-T Recommendation H.323, “Packet based multimedia communications systems”, February 1998.

[20] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications”, RFC 3550 July 2003.

[21] ITU-T Recommendation G.711, “Pulse code modulation (PCM) of voice frequencies”, 1988.

[22] ITU-T Recommendation G.721, “32kbit/s adaptive differential pulse code modulation (ADPCM)”, December 1990.

[23] ITU-T Recommendation G.726, “40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)”, December 1991.

[24] ITU-T Recommendation G.723.1, “Dual rate speech coder for multimedia communications transmitting at 5.3 & 6.3 kbit/s”, March 1996.

[25] ITU-T Recommendation G.729 Annex A, “Reduced complexity 8 kbit/s CS-ACELP speech codec”, November 1996.

[26] ITU-T Recommendation G.729, “Coding of speech at 8 kbit/s using conjunctive-structure algebraic-code-excited linear-prediction (CS-ACELP)”, March 1996.

[27] ITU-T Recommendation G.113, “Transmission impairments due to speech processing”, February 2001.

[28] H. Oouchi, T. Takenaga, H. Sugawara, and M. Masugi, "Study on appropriate voice data length of IP packets for VoIP network adjustment", in *Proceedings of GLOBECOM 2002 - IEEE Global Telecommunications Conference*, November 2002.

[29] R. Cox and P. Kroon, “Low Bit-Rate Speech Coders for Multimedia Communication”, Appeared in *IEEE Communications Magazine*, December 1996.

[30] ITUT-T Recommendation G.1010, “End user multimedia QoS categories”, November 2001.

[31] P. Crow, I. Widjaja, J. G. Kim, and P. T. Sakai, “IEEE 802.11 Wireless Local Area Networks”, *IEEE Communications Magazine*, September 1997, pp 116-126.

[32] H. Zhu et al., ”A Survey of Quality of Service in IEEE 802.11 Networks,” *IEEE Wireless Communications*, August 2004, pp. 6-14.

[33] Cisco White Paper, “An Introduction to Corvil Bandwidth Technology”, 2005.

[34] P. Rabinovitch, “Statistical Estimation of Effective Bandwidth”, M.Sc. thesis, Carleton University, Ottawa, ON, Canada, 2000.

[35] R. Vesilo, V.Solo, “Techniques for adaptive estimation of effective bandwidth in ATM networks”, in *Proceedings of GlobeCom97 - IEEE Global Telecommunications Conference*, 1997.

[36] C.S. Chang, and J.A. Thomas, “Effective bandwidth in high speed digital networks”, Appeared in *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pages 1091-1100, August 1995.

[37] J. Walrand and P. Varaiya, “Datagram Networks: Statistical Procedures”, in *High Performance Communications Networks”* 2nd ed. San Francisco, CA: Morgan Kaufmann, 2000.

[38] J. Gil, “Data Packet Loss in a Queue with Limited Buffer Space”, presented at the *49th Study Group with Industry*, University of Oxford, 2004.

[39] J. T. Lewis and R. Russell, “An Introduction to Large Deviations for Teletraffic Engineers”, in *Proceedings of PERFORMANCE '96*, Lausanne, October 1996.

[40] N. G. Duffield, J. T. Lewis, N. O'Connell, R. Russell, F. Toomey, “Entropy of ATM Traffic Streams: A Tool for Estimating QoS Parameters”, in *IEEE JSAC, Vol. 13, No. 6*, pp. 981 - 990, August 1995.

[41] WildPackets Inc, *AiroPeek NX*. [CD-ROM]. Walnut Creek, CA.

[42] G. Combs *et al*, *Ethereal*. Available: http://www.ethereal.com/. [Accessed June 2005].

[43] ETSI Guide (EG) 201 377-1, "Speech processing, Transmission and Quality aspects (STQ); Specification and measurement of speech transmission quality; Part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks", October 2002.

[44] ITU-T Recommendation P.59, "Artificial conversational speech", March 1993.

[45] ITU-T Recommendation P.50, "Artificial voices", 1999.

[46] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", August 1996.

[47] ITU-T Recommendation P.82, "Method for evaluation of service from the standpoint of speech transmission quality, 1988.

[48] T. A. Hall, "Objective Speech Quality Measurement of IP Telephony", Available from *National Institute of Standards and Technology*, Gaithersburg, MD, 2002.

[49] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs", February 2001.

[50] ITU-T Recommendation G.107, " The E-Model, a computational model for use in transmission planning", March 2003.

[51] Telecommunications Industry Association, "Voice Quality Recommendations for IP Telephony – TIA/EIA/TSB116", 2001.

[52] M. Davis, "A Wireless Traffic Probe for Radio Resource Management and QoS Provisioning in IEEE 802.11 WLANs," *Proc. 7th ACM MSWiM 2004*, Venezia, Italy, Oct. 2004.

[53] M. Davis and T. Raimondi, “A Novel Framework for Radio Resource Management in IEEE 802.11 Wireless LANs”, *Proc 3rd International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt’05)*, April 3-7, 2005, Riva del Garda, Trentino, Italy.

[54] F. Anjum, et al., “Voice Capacity in WLAN Networks - An Experimental Study”, in *Proceedings of GLOBECOM 2003 - IEEE Global Telecommunications Conference*, 2003 pp 3504-3508.

[55] C. Casetti, C-F. Chiasserini, “Improving Fairness and Throughput for Voice Traffic in 802.11e EDCA”, *IEEE PIRMC'04*, Barcellona, Spain, 5-8 September 2004

[56] M. Elaoud and P. Agrawal, “Voice capacity in IEEE 802.11 networks”, Appears in *Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium*, September 2004 Volume: 1, pp 78- 82 Vol.1.

[57] M. Veeraraghavan, N. Cocker, T. Moors, “Support of Voice Services in IEEE 802.11 Wireless LANs”, *IEEE INFOCOM’01*, Vol. 1, Apr. 2001, pp.488-497.

[58] A. Koepsel and A. Wolisz "Voice transmission in an IEEE 802.11 WLAN based access network" *Proc. of ACM Workshop on Wireless Mobile Multimedia*, July 2001.

[59] ETSI Technical Report ETR 101 329-6 V2.1.1, “Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 3; End-to-end Quality of Service in TIPHON systems; Part 1: General aspects of Quality of Service (QoS)”, February 2002.

[60] Nortel Networks, “Packet Loss and Packet Loss Concealment: A summary of how lost or late packets affect speech quality and how concealment is achieved”, Technical Brief.

[61] K. Medepalli, P. Gopalakrishnan, D. Famolari and T. Kodama, “Voice Capacity of IEEE 802.11b, 802.11a and 802.11g Wireless LANs”, GlobeCom ’04, 2004.

[62] S. Garg and M. Kappes, “Can I add a VoIP call?”, Appears in Communications 2003 ICC ’03.IEEE International Conference, May 2003 Volume: 2, pp 779- 783 vol.2.

[63] Spectralink Inc, “Netlink Wireless telephones FAQ”, Available at http://www.spectralink.com/products/pdfs/Netlink\%20FAQ.pdf. December 2004.

[64] ITU-T Recommendation G.114, “One-way transmission time, May 2003.

[65] Texas Instruments White Paper, “Including VoIP over WLAN in a Seamless Next-Generation Wireless Environment”, 2003.

[66] Cisco White Paper, “Understanding delay in packet voice networks”, 2005.

[67] Naval Research Laboratory (NRL) PROTocol Engineering Advanced Networking (PROTEAN) Research Group, MGEN Version 4.0, Available from http://downloads.pf.itd.nrl.navy.mil/mgen.

[68] Cisco White Paper, “Service Provider Quality-of-Service”, 2002.

[69] UCB, LLNL, Xerox PARC, and USC/ISI, *The Network Simulator (ns) v. 2*. Available: http://www.isi.edu/nsnam/ns/index.html. [Accessed January, 2005].

[70] OPNET Technologies, Inc, *IT Guru*. Available: http://www.opnet.com/products/home.html. [Accessed July, 2005].

[71] Internet Real-Time Lab (IRT), *RTP Tools (version 1.18)*, Computer Science Department at Columbia University. Available: http://www.cs.columbia.edu/IRT/software/rtptools/. [Accessed February 2005].