THE EFFECT OF VOICE PACKET SIZE ON END-TO-END DELAY

IN 802.11b NETWORKS

by

Haritha Phalgun

B.E. in Electronics and Communication Engineering, University of Madras, 2001

M.S. in Telecommunications, University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

Information Sciences in partial fulfillment

Of the requirements for the degree of

Master of Science in Telecommunications

University of Pittsburgh

April 29,2003

UNIVERSITY OF PITTSBURGH

FACULTY OF INFORMATION SCIENCES

This thesis was presented

By

Haritha Phalgun

It was defended on

April 29, 2003

and approved by

Dr. Richard Thompson, Director and Professor

Dr. Prashant Krishnamurthy, Assistant Professor

Thesis Advisor: Dr.Joseph Kabara, Assistant Professor

The Effect of Voice Packet Size on End-to-End Delay in 802.11b Networks

Haritha Phalgun, MST

University of Pittsburgh, April 29, 2003

Voice over IP (VoIP) uses the existing data networks to support voice services. It has a broad appeal in that it is currently unregulated and calls can be placed free of cost to any part of the globe. The integration of voice traffic with data traffic opens up opportunities for new revenue stream for Internet Service Providers. However, in mixing data types the constraints on each data type must still be met and unlike regular data, voice networks are chiefly limited by end-to-end delay. In the case of packet switched networks delay becomes a determining factor in the quality of the voice call and therefore the success of VoIP. At the same time, WLANs are becoming widely adopted due to the simplicity in installation and convenience offered. Advancement in technology now enables WLANs to provide most of the facilities provided by their wired counterparts with the added benefit of mobility at a very low cost.

The benefits of combining IP telephony and WLANs can be effectively utilized if the control over end-to-end delay can be achieved. In conventional IP telephony the voice packets travel across the wired Internet. We developed a study in which the final hop on each end of the communication channel is a wireless 802.11b network. Results show that with a wireless network at the transmitting end the delay characteristics change considerably.

Acknowledgements

First and foremost, my heartfelt thanks to my advisor Dr. Joseph Kabara for his invaluable guidance at every stage of my research work. I sincerely thank Dr. Prashant Krishnamurthy for patiently assisting me throughout the thesis. I also express my gratitude to Dr. Richard Thompson, member of the thesis committee.

I convey my special thanks to Boonchai Ngawongwattana for his valuable suggestions. I wish to express my appreciation to my friends Sohail Hirani, Anirban Ghosh and Eric Castillo for their constructive comments. I also thank Jim Fausnaught, Sopan Joshi, Ramakrishnan Balakrishnan and Vishnu Rajan for their help.

Last but not the least I extend my gratitude to my mother, father, sister and uncle for their constant support and encouragement without whom this thesis work would have been impossible.

CONTENTS

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Drivers for Packet Voice

Packet voice technology uses the existing data networks for offering voice. It has a broad appeal in that it is currently unregulated and calls can be placed free of charge to any part of the globe. Since packet networks offer the capability for multiplexing, voice packets travel over the Internet just as data packets do. Thus the operating and maintenance costs may be reduced. The integration of voice traffic with Internet traffic opens up many opportunities. Incorporating value added features with normal telephony services opens up a new revenue stream for Internet service provider. Services such as web integration, collaboration and instant messaging can also be provided [5]. Additionally, in a properly engineered IP network changing the location of the telephone does not incur extra costs. The procedure for adding or removing features from the telephone is simple and can be completed remotely. Conventional circuit switched networks dedicated to voice traffic offer very good voice quality because the voice traffic has its own dedicated bandwidth. With circuit switching, data bits go directly to the receiver in an orderly fashion, one after another on a single path. With packet switching routers determine a path for each packet on the fly, directing them over any path available to get to the destination [2]. In IP telephony analog voice signals generated for transmission are first converted into a stream of bits. The digitized voice is then packetized and sent over the network. The process of packetization involves the collection of compressed voice frames into an IP packet. At the receiving end the process is reversed. The voice frame is decompressed. In the case of packet switched networks delay becomes an important issue. The perceived quality of a voice call is delay sensitive. Variation in the delay, Jitter, is another problem for interactive voice applications [4]. The time sensitive voice packets and the regular data packets share the same single network. The quality of voice delivered does not always remain the same. The delivered voice quality is an important factor in determining the success of VoIP [5].

1.2 Opportunities for wireless LANs

Advancements in technology now enable wireless LANs to provide most of the facilities provided by their wired counterparts with added benefits. In certain scenarios wireless LANs complement existing wired infrastructure. A wireless local area network (WLAN) uses radio waves for data transmission between devices in a license-free band. Wireless LANs allow user mobility and simple low cost installation, operation and management. Vendors are now integrating wireless capabilities into computers leading to further drop in prices. Coverage can be provided to areas that are difficult to wire by mounting an access point (AP). The aesthetics of the existing infrastructure is preserved. Wireless LAN systems can be configured in a variety of topologies to meet the needs of specific applications and installations. Configurations can support thousands of users over a broad area. Although WLANs are being rapidly adopted they have additional constraints when compared to wired networks. Interactions with typical building objects, including walls, metal, and even people, can affect the manner in which energy propagates and the range and coverage of the system. Additional steps must be taken to ensure that the networks are secure. Since wireless LANs operate in the unlicensed band products that transmit energy in the same frequency spectrum can potentially interfere with a wireless LAN system. Energy efficient devices must be used to prolong the battery life of mobile equipment without losing mobile connectivity. The low cost of wireless devices has encouraged extensive research on wireless LANs.

1.3 Thesis Overview

We propose to combine benefits of IP telephony and wireless LANs. Previous studies show that on wired networks delay increases as a function of packet size [5]. We measured delay characteristics when one VoIP node is connected to the network through a wireless link. Packet sizes are varied for different load scenarios and the end-to-end delay across the network measured. The increase in delay value on addition of the wireless hop is noted. We also measured the interaction between the wireless link and the queuing algorithm enabled on intermediate routers. The packet loss is experimentally observed. The savings in the delay budget on using data frames of different sizes is estimated. We also measured the interaction between the wireless link and the queuing algorithm enabled on intermediate routers. Two commonly used queuing methods are dealt with in this thesis, Weighted Fair Queuing (WFQ) and the First-in First-out (FIFO) queuing.

The organization of this thesis material is as follows. Chapter 2 provides an overview of the performance requirements of VOIP and highlights the network parameters in wireless LANs. Chapter 2 also discusses the RTP protocol, the IEEE 802.11and various speech codecs and queuing types. Chapter 3 presents the experimental results and Chapter 4 introduces the system model. Chapter 5 presents a conclusion to the current work and addresses issues that may be appropriate for future work.

Chapter 2

BACKGROUND

We first look at the 802.3 (Ethernet) as it helps understanding the 802.11b better. Besides the Ethernet comprises a large part of the network used in the experimental work. The Ethernet converter box converting 802.11 to 802.3 and vice versa, routers and switches are 802.3 devices used in the setup.

2.1 Ethernet

Ethernet is a network cabling and signaling specification using a logical bus topology. All the hosts on the network share the same bandwidth using a contention media access method called the Carrier Sense Multiple Access with Collision Detection (CSMA/CD). In CSMA/CD before a device puts a frame onto the wire it listens for any ongoing transmissions. If the line is clear the device sends the packet. This prevents two devices from transmitting at the same time. A transmitting host constantly monitors the wire and if another signal is detected a jamming signal is transmitted. This way all the stations get to know that there has been a collision and they stop transmitting. They try retransmitting after a random amount of time determined by a truncated exponential back-off algorithm [22]. A time out occurs if collisions continue after 15 attempts.

2.1.1 Ethernet at the Physical Layer

The first Ethernet LAN specification was a 10 Mbps network that ran on coax, twisted-pair and fiber physical media. The IEEE extended this to the 802.3U (fast Ethernet) and the 802.3Z (Gigabit Ethernet) [1]. According to the EIA/TIA (Electronics Industries Association and the Telecommunications Industry Association) Ethernet uses a registered jack (RJ) connector with a 4/5 wiring sequence on unshielded twisted-pair (UTP) cabling. The 4/5 designation specifies the pin numbering scheme. Pins 4 and 5 are not used in a 10Mbit LAN. The Ethernet standards follow the notation: <Data Rate in Mbps> <Signaling method> <Maximum Segment Length (hundreds of meters)>. The original 802.3 standard 10base5 (10B5), also known as thicknet, supports 10 Mbps to 500 meters. The coverage can be extended to 2500 meters supporting 1024 users using repeaters [2]. 10b5 employs Manchester signaling over a 50-ohm coaxial cable. A 10b2, also known as thinnet, can support upto 30 workstations on a single 185 meter segment. 10Base2 was created for small lower-cost network and is the least expensive method of cabling. 10b2 and 10b5 segments can be combined in the same network using a repeater that supports a 10Base2 on one side and a 10Base5 on the other. Unshielded Twisted Pair (UTP) refers to copper wiring used in small-to-large networks to connect host devices to hubs and switches. It may also be used to connect switch to switch or hub to hub. 10BaseT supports 10Mbps using Category 3 UTP wiring. There can be only one host per segment and each device must connect directly to a hub or switch. The length of a link is limited to 100 meters due to the limited bandwidth and high noise of UTP. The 10BaseT is the most flexible topology for LANs and the cabling is cheaper and requires minimum installation skills.

The 10BaseF corresponds to 10Mbps Ethernet running over fiber-optic cable. Each transmission link consists of a pair of optical fibers. Manchester encoding is used for signaling.

The standard has three specifications - 10-BASE-FP (passive), a passive star topology for connecting stations and repeaters upto 1km on a segment, 10-BASE-FL (link), a point-to-point link for connecting stations and repeaters upto 2 km on a segment and 10-BASE-FB (backbone), a point-to-point link for connecting repeaters upto 2 km on a segment. Its installation is recommended in specialized environments that are extremely noisy or very secure since it is expensive. They may also be used in intra-building backbone networks.

In 10Broad36 10Mbps Ethernet runs through a broadband cable. This is the only 802.3 specification for broadband. The Community Antenna Television (CATV) cable is used for broadband local networks, and broadcast TV distribution. A standard 75-ohm CATV coaxial cable of dual-cable or split-cable configuration can be used. The maximum length of an individual segment, emanating from the headend, is 1800 meters; this results in a maximum end-to-end span of 3600 meters. Differential Phase Shift Keying (DPSK) is the signaling technique used on the cable.

The 802.3U (Fast Ethernet) has the same physical characteristics as the 802.3 Ethernet but offering a speed 10 times that of the 802.3 Ethernet. The data transmission speed over 802.3Z (Gigabit Ethernet) networks is 1000 megabits per second. The 100BaseTX supports 100 Mb/s transmission rate over up to 100 meters of Category 5 (CAT5) Unshielded Twisted Pair (UTP) cabling over two pairs of twisted pair cabling, one for receiving data and the other for transmission. CAT5 cable supports transmission up to 100 MHz. The 100BaseFX is basically a fiber optic cabling version of 100BaseTX supporting 100 Mb/s transmission rate over two fiber optic cables. When this mode is used, the maximum segment length increases to 2000 meters or more. The 1000BaseLX standard uses long wavelength lasers for data transmission over fiber optic cable. The motivation for 1000BaseSX, using short wavelength lasers is that short wavelength lasers are less expensive than long wavelength lasers. 1000BaseCX using shielded balanced copper jumper cables supports a maximum segment length of 25 meters and is therefore restricted to very small areas. The 1000BaseT supports Gigabit Ethernet over 100 meters of CAT5 balanced copper cabling. It achieves a data rate of 1000 Mb/s by transmitting data at a rate of 250 Mb/s over four CAT5 wire pairs.

2.1.2 Ethernet Cabling

We look at the configurations in which twisted pair copper wires are used for communication over a network. The straight through and crossover cables are widely used Ethernet cables. In a straight-through cable four wires are used to connect to Ethernet devices. This connection is used when a host or a router is connected to a switch or a hub. The cables used on twisted pair networks are usually wired straight through. A crossover cable is used for switch-to-switch, hub-to-hub or host-to-host connections. As in the case of a straight-through cable four wires are used in this cable but the way the pins are connected differs.

Coaxial cables are widely used for computer networks. It consists of a center wire surrounded by insulation and then a grounded shield of braided wire that minimizes electrical and radio frequency interference. When devices are connected to cable, electronic terminations are placed at both ends of the cable. This prevents electronic echoing of signals. A transceiver creates the interface between the device and the cable. Bits are transmitted on the cable using CSMA/CD contention technique. The transceiver clamps onto the cable using a vampire clamp. Communication with the device is done via a transceiver cable or Attachment Unit Interface (AUI) cable. The AUI cable is made up of five twisted pairs two of which are used to send data and control information to the device. For receiving data and control information two others are used. The last pair is used to connect the source of power and the ground.

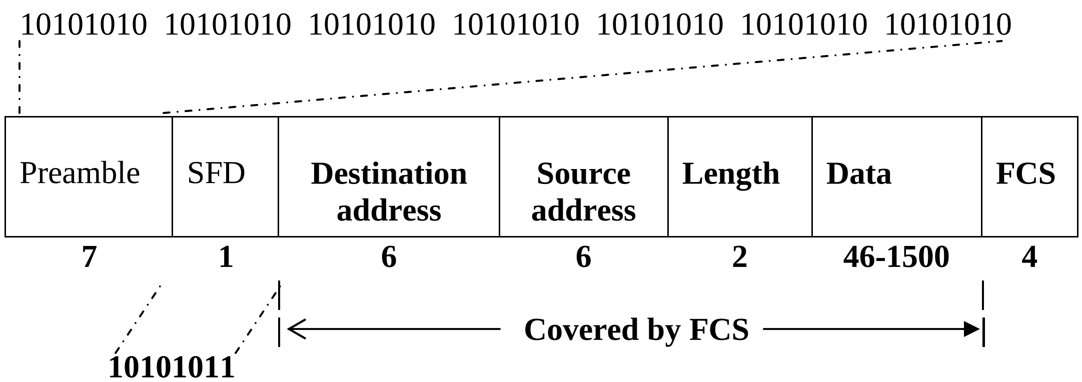
The transceiver cable connects to a PC using the network interface card (NIC) installed in the PC. The NIC performs functions appropriate for the MAC layer protocol including error checking, creating frames, checking for retransmissions and identifying frames destined for the particular PC. The logic on the NIC buffers data and moves it between the transceiver cable and the memory of the PC. The processor of the PC is thus freed from the overhead of these tasks.

2.1.3 Ethernet Frame Format

The Ethernet frame format encapsulates data and delivers it across a link. Figure 2-1 shows the frame format for Ethernet [3]. The Preamble is a seven-octet pattern of alternating zeros and ones used for synchronization of the receiver clock. The start of the frame delimiter (SFD) is indicated by the special pattern 10101011. If the first bit in the destination address is a zero then it is goes to a particular station. If it starts with a one the frame is sent to a group of stations with the specified group address. In case all the bits are ones the frame is broadcast, to be received by all nodes on the LAN.

The Data Length or Type Field specifies the number of octets (bytes) in the combined data and pad field. There is an upper and lower limit placed on the number of data and pad octets corresponding to 46 and 1500 bytes respectively. If the length of the packet is less than the lower limit extra octets are padded to make up the difference. The lower limit is to ascertain that the collision technique works properly. This minimum value of the frame size depends on the maximum time it takes a sending station to detect a collision before sending the frame’s last bit.

In the worst case, the time to detect a collision is twice the time it takes a signal to span the longest distance covered by the medium.



SFD: Start frame delimiter

FCS: Frame check sequence (32-bit CRC value)

Figure 2-1 The Ethernet Frame Format

The distance a frame travels, a bit's propagation speed, the data rate, and the delays caused by any repeaters determine the length. Considering all these factors, the 802.3 standard defines a minimum frame length as 64 bytes (octets). The minimum Ethernet header size is 18 Bytes [Destination Mac (6) + Source Mac (6) + Frame Type (2) +CRC (4)] and the minimum data portion is 46 Bytes. The upper limit prevents a transmission from monopolizing the medium. Cyclic Redundancy Check (CRC) is an error detecting code in which the code is the remainder resulting from dividing the bits to be checked by a predetermined binary number. A 32-bit CRC is used for error checking. A Frame Check Sequence (FCS) is an error-detecting code inserted as a field in a block of data to be transmitted. The code serves to check for errors upon reception of the data. A (FCS) is produced from a 32-bit cyclic redundancy check (CRC-32), for error checking.

2.2 An Overview of IEEE 802.11

The IEEE 802.11 standard describes a WLAN that can deliver services previously provided by wired networks with the added benefit of mobility. The architecture encompasses protocols for efficient utilization of the battery life of the mobile device. Power-saving modes of operation are also included in the architecture. This flexible architecture supports small and large networks, transient or permanent in nature. The IEEE 802.11 is inserted under the IEEE 802.2 logical link control (LLC) and has the same interface as the IEEE 802.3. Thus the concept of mobility is introduced. The protocols above the Medium Access Control (MAC) layer are unaware of this. The IEEE 802.11 MAC supplies the functionality required to provide a reliable delivery mechanism for user data over noisy, unreliable wireless media.

2.2.1 The IEEE 802.11 Architecture

The components of the IEEE 802.11 architecture are as follows. The station may be stationary, portable or mobile. Connection to the wireless medium is established through it. It has a MAC and a PHY. The services supported by the station include delivery of data, authentication, deauthentication and privacy. As in wired networks, the station may also be referred to as the network adapter or the NIC. An access point is also a station. This station provides distribution services.

The Basic Service Set refers to a set of stations communicating with each other. A BSS is termed as an Independent BSS when all the stations are mobile and they are not connected to a wired network. In order to communicate with one another they should be within the direct communication range. A BSS becomes an infrastructure BSS when it includes an access point. Unless explicitly stated the term BSS refers to an infrastructure BSS. In an infrastructure BSS all communication between the mobile stations go through the access point. The access point may in turn be connected to a wired backbone.

The Extended Service Set extends the range of mobility over BSSs. Thus mobile devices can move from one BSS to another and still stay connected. An ESS is a collection of infrastructure BSS. Data traffic is forwarded from one BSS to another via the access point using the distribution system. The ESS provides a level of indirection to network equipment outside it. To these equipments all the mobile stations in the ESS appear physically stationary [8].

The Distribution System (DS) is a layer in the access point. It determines if the communication must be transmitted within the same BSS, relayed to a different BSS or sent to a destination not in the ESS over a wired network. It enables seamless communication with the mobile station moving from one BSS to the other [8]. The DS may consist of wired or wireless networks.

The radio communication link in Wireless LANs is difficult to control because the signals are propagated over a dynamic media whose characteristics are constantly changing. The movement of the objects in the environment can produce significant changes in the power at the radio receiver. Multipath and path loss must be considered in the design of an IEEE 802.11 wireless system. When the reflected signal path combines with the direct signal path, the signal at the receiver is corrupted, termed as multipath. In indoor environments multipath is a performance concern. The delay of the reflected signal is called the delay spread. This parameter signifies the multipath. A RAKE receiver is well-known architecture used to remove delay spreads on the order of 100 nsec. The RAKE is structured as a bank of correlators (fingers) with weighed delays and a combiner. RAKE processing and equalization are used to resolve delay spread. Path loss represents signal attenuation determining the operating range. The radio signal attenuates as a square of the distance between the mobile node and the access point.

IEEE 802.11 services are classified into station and distribution services. The station services are data delivery, authentication, deauthentication and privacy [8]. Data frames are reliably delivered using the data delivery service. The authentication service proves the identity of the station. Thus only authorized users are allowed access to the network. After using the network the user is deauthenticated. The privacy service protects data as it travels over the wireless medium.

The distribution services included in the architecture are association, reassociation, disassociation and integration [8]. These services facilitate roaming of the mobile station in an ESS. Connection to the wired infrastructure is provided by the distribution services comprising a thin layer between the MAC and the LLC sublayer. When the mobile station enters the WLAN for the first time or when rediscovering the WLAN after being out of touch for some time the association service is invoked. This service makes a logical connection between the mobile station and the access point. The connection established lets the DS know how to deliver data to the mobile station and lets the access point accept data from the mobile station and allocate resources to it. When a mobile node moves reassociation services provide information to the access point with which the mobile station will be associated. The new access point contacts the access point with which the station was previously associated gets the remaining data frames for delivery to the mobile station. A mobile station uses the disassociation service to inform the access point that it no longer requires the services of the WLAN. Alternately this service forces a mobile station to associate. An access point may use the disassociation service when it is shutting down or if the demand is more than the resources available. The distribution service determines if a frame received for delivery from a mobile station is to be sent to another access point or to a station linked to the same access point. The frame is sent to the DS for delivery to a mobile station associated with another access point. The IEEE 802.11 WLAN is connected to other LANS using the integration service.

According to the IEEE 802.11 standard each station must maintain two variables. One variable represents the state of the authentication or deauthentication services and the other variable corresponds to the state of the association, reassociation or disassociation services. While a station may be authenticated with many stations simultaneously it may be associated with only one station at a time.

2.2.2 802.11 Protocol Layers

The layered protocol architecture of the IEEE 802.11 is shown in Figure 2-2. The lowest layer of the OSI is the physical layer (PHY). This layer addresses the details involved in the actual radio transmission. The PHY layer consists of the Physical Layer Convergence Procedure sublayer (PLCP) and the Physical Medium Dependent sublayer (PMD). The PLCP sublayer controls frame exchange between the PHY and the medium access control (MAC) layer above it. The PMD sublayer exercises control over the carrier and spread spectrum modulation techniques used to transmit the data frames over the media. When there is activity on the media the PHY sends a carrier indication to the MAC.

<table>
<tr><td colspan="3">802.2</td><td>Data Link Layer</td></tr>
<tr><td colspan="3">802.11 MAC</td><td></td></tr>
<tr><td>FH</td><td>DS</td><td>IR</td><td>PHY Layer</td></tr>
</table>

Figure 2-2 The Protocol Architecture of the IEEE 802.11

The three physical layer specifications described below are unique in terms of the type of modulation are designed to coexist with each other. Direct Sequence Spread Spectrum PHY (DSSS) uses the 2.4 GHz frequency band for transmission. In this scheme each data bit is mapped into a string of bits with one string representing binary 1 and another representing binary 0. Since each bit is spread out over time the effects of interference is reduced. The DSSS PMD sublayer takes the bits handed over by the PLCP sublayer and transforms it into RF signals to be transmitted over the wireless media by using carrier modulation and DSSS techniques. Frequency Hopping Spread Spectrum PHY (FHSS) uses the 2.4GHz spectrum. In the FHSS scheme frequent jumps are made from one carrier frequency to the other. Hopping sequences for IEEE 802.11 are grouped in hopping sets for worldwide operation: Set 1, Set 2, and Set 3. The hopping sets are designed to minimize interference between neighboring FHSS radios in a set. Performance degradation at a particular frequency is minimized as only a small fraction of the transmission is affected. FHSS techniques are used to transform the bits of information handed over to the FHSS PMD. The Infrared PHY (IR) uses Pulse Position modulation (PPM) is to transmit data bits. This modulation scheme encodes information in the position of the pulse keeping the amplitude and pulse width a constant. PPM is used in IR systems to reduce the optical power needed of the LED infrared source. Since infrared requires line-of-sight propagation it is not widely used.

The Medium Access Layer (MAC) layer uses a frame exchange protocol to provide reliable data delivery over the wireless media. In addition the IEEE 802.11 MAC provides the wired equivalent privacy service (WEP). This encrypts the data sent over the wireless medium.

Access to the shared wireless medium is performed through two basic access mechanisms.

The Distributed Coordination Function (DCF) is based on Carrier sense multiple access with collision avoidance (CSMA/CA). Physical carrier sensing is performed at the air interface and virtual carrier sensing is performed at the MAC sublayer. The virtual carrier sensing mechanism uses a Network Allocation Vector (NAV). The NAV is a value, which indicates to each station how long it needs to wait before the wireless medium becomes available. The NAV is updated by duration values transmitted in all frames. If the NAV value is not zero, the station shall avoid transmitting, even when the medium appears to be idle by the physical carrier sensing mechanism. Inter-frame Space (IFS) time intervals between the transmission of frames control priority access to the wireless medium. Three IFS intervals specified in the standard are DCF IFS (DIFS), point coordination function IFS (PIFS), and short IFS (SIFS). The DIFS interval is the longest and the SIFS interval is the shortest. The SIFS interval also has the highest priority. When a request to transmit a frame is made to the MAC the physical and virtual carrier sense mechanisms are checked. Transmission of the frame begins if both the mechanisms indicate that the medium is not used for a DIFS interval. Otherwise the MAC selects a back-off interval using the binary exponential back-off mechanism and a retry counter is incremented. Each time the carrier senses the medium to be idle for one time slot the back-off value is decremented. Transmission begins once the back-off interval has expired. After a frame is transmitted if an acknowledgement (ACK) is returned by the recipient, if an ACK is not received then a collision is assumed to have occurred and the transmission is considered unsuccessful. The contention window is doubled, a new back-off interval selected and the countdown restarted. This process repeats until a transmission is successful or cancelled.

The Point Coordination Function (PCF) provides contention-free frame transfer using a poll and response protocol. It is a centrally controlled access mechanism. Use of the PCF is an optional part of the IEEE 802.11 standard but every station should support it. A point coordinator (PC) located in an access point controls the PCF. Stations request the PC to register them on a polling list. The PC regularly polls this list for traffic. The PCF coexists with the DCF by employing the shorter PIFS interval. The PC begins operation during the contention-free period (CFP). When the PC gains access to the medium using the normal DCF procedures the CFP begins. It transmits a beacon frame periodically to allow mobile stations to locate and identify the BSS. During this time, access to the medium is completely controlled by the PC and the DCF does not gain access to the medium due to the PC using the shorter PIFS interval. The CFP repetition interval determines the frequency of occurrence of the PCF.

2.2.3 IEEE 802.11b

The Direct Sequence Spread Spectrum Radio (DSSS) PHY provides both 1 and 2 Mbps operation. The IEEE 802.11b providing high data rate of 5.5 Mbps and 11 Mbps is the PHY layer extension of the IEEE 802.11 DSSS scheme. It is referred to as the high rate direct sequence spread spectrum (HR/DSSS). Fall back to 1 and 2 Mbps is supported. Using the channel ability option defined in the IEEE 802.11b the IEEE 802.11 FHSS 1 and 2 Mbps networks can operate with the HR/DSSS 11 Mbps networks. Data rates are achieved by using different modulation schemes. The basic data rate is 1 Mbps Differential Binary Phase Shift Keying (DBPSK) modulated. The extended rate is 2 Mbps Differential Quadrature Phase Shift Keying (DQPSK) modulated. The basic and the extended rate use the 11-bit Barker word as the spreading format. The enhanced rate operates at 5.5 Mbps and 11 Mbps using CCK modulation and Packet Binary Convolutional Coding (PBCC). PBCC is an option for enhanced performance networks.

802.11b operates in the 2.4-GHz ISM (Industrial, Scientific, and Medical) band. 802.11b provides 11 channels (for North America), each channel being 22 MHz in width, and each channel centered at 5 MHz intervals (beginning at 2.412 GHz and ending at 2.462 GHz) as shown in Figure 2-3. Both overlapping and non-overlapping channels are supported. When using non-overlapping channels wireless LANs operating in the same area will not interfere with each other. In North America channels 1, 6 and 11 are used for non-overlapping networks.

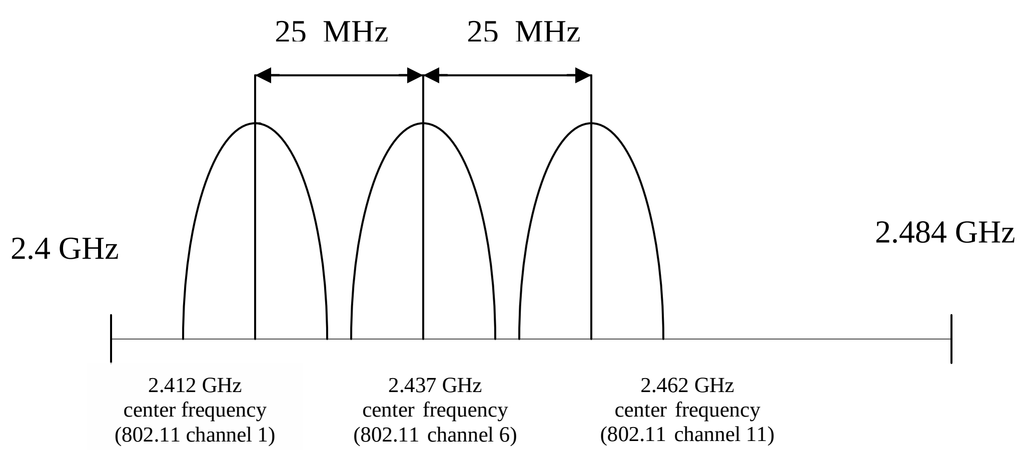


Figure 2-3 Nonoverlapping 802.11b Channels

2.2.4 WLAN Network parameters

The throughput corresponds to the amount of data in bits that is transmitted over the channel per unit time. It is measured in bits per second (bps). The throughput of a network connection depends on the type of protocol used and the type of traffic sent over the network. The bandwidth of the network connection also affects the throughput.

We consider unidirectional delay corresponding to the time taken for voice packets to reach the receiver, measured in milliseconds. Round trip time, the time required for the data packet to travel from one host to another and return back to the original host is not considered.

Jitter corresponds to the average variation in the arrival time of a packet. The delay variation or jitter is an important parameter affecting performance in Voice over IP applications [5]. Jitter is measured in terms of milliseconds.

Data bandwidth is the maximum available rate at which data can be transmitted over the network. Unlike throughput the data bandwidth includes the protocol overhead, checksum and assumes the protocol to be fully efficient. So the data bandwidth is always greater than the throughput. It is measured in bits per second (bps).

Packet loss is expressed as a ratio of the number of packets lost to the total number of packets transmitted. In our experiment packet loss is an important parameter affecting the performance of the network. Beyond a particular packet loss percentage transmission of voice packets over the wireless network results in unacceptable voice quality [22].

2.3 The Internet Protocol and the User Datagram protocol

We look at the Internet protocol (IP) and the User Datagram Protocol (UDP) to understand the structure of the voice packet. Packet switched computer communication networks use this network layer protocol. IP provides connectionless, unreliable, best-effort packet delivery service. The protocol implements two basic functions namely addressing and fragmentation. IP receives segments from the Host-to-Host layer and fragments them into datagrams (packets). These blocks of data are transmitted from sources to destinations. IP then reassembles datagrams back into segments on the receiving side. Hosts are identified using fixed length addresses. All hosts on a network have a logical ID called an IP address. Each IP packet is treated independently, each packet is stamped with the address of the receiver and the sender. Routing decisions are made on a packet-by-packet basis. Intermediate nodes inspect the address of each packet and choose the best path using a routing table.

Bit 0 Bit 15 Bit 16 Bit 31

<table>
<tr><td>Version (4)</td><td>Header Length (4)</td><td>Priority and Type of Service (8)</td><td colspan="2">Total Length (16)</td></tr>
<tr><td colspan="3">Identification (16)</td><td>Flags (3)</td><td>Fragment Offset (13)</td></tr>
<tr><td colspan="2">Time to Live (8)</td><td>Protocol (8)</td><td colspan="2">Header Checksum (16)</td></tr>
<tr><td colspan="5">Source IP address (32)</td></tr>
<tr><td colspan="5">Destination IP address (32)</td></tr>
<tr><td colspan="5">Options (0 or 32 if any)</td></tr>
<tr><td colspan="5">Data (varies if any)</td></tr>
</table>

20 bytes

Figure 2-4 The IP Header

Figure 2-4 shows the format of the IP header. The priority or ToS field tells how the datagram should be handled. This field allows four request options: precedence, low delay, high throughput, and high reliability. The length field gives the total length of the packet including the header and the data. The identification field is a sequence number that together with the source, destination address and user protocol helps uniquely identify a datagram. The flags indicate if fragmentation may occur. If the packet is too large to fit into a MAC layer frame the Fragment Offset field provides fragmentation and reassembly. The time to live value is set when the packet is originally generated and is decremented as the packet pass through intermediate nodes, thus preventing it from continuously circling the network. The protocol field indicates the protocol. The value for TCP port is 6 and that for the UDP is 17. The IP option field is not required in every packet but can be used for special treatment of the packet in terms of network testing, debugging and security. The data field contains the data provided by the transport layer protocol.

The User Datagram Protocol (UDP) is a transport layer protocol for use with the IP network [19]. It is considered a connectionless protocol as it neither creates a virtual circuit nor contacts the destination before delivering information to it. It does not guarantee delivery or duplicate protection. UDP assumes that the application will use its own reliability method. It takes messages from the application process, attaches source and destination port number fields for multiplexing or demultiplexing services (to communicate with the destination process). The network layer encapsulates the UDP datagram into an IP packet and then passes the packet to the Link layer protocol. UDP neither sequences the segments nor orders the segments at the destination. It neither checks if the packets are received nor acknowledges receipt and is therefore referred to as an unreliable protocol. If a segment arrives at the receiving host, UDP uses the port numbers and the IP source and destination addresses to deliver the data in the segment to the correct application process. The simplicity of UDP requires minimal overhead. Figure 2-5 shows the User Datagram Protocol header format. UDP has low overhead as compared to TCP. The length field corresponds to the length of the UDP header and the UDP data fields. The CRC is a checksum of both the UDP header and the UDP data fields. The data field corresponds to the data from upper-layers.

Bit 0 … Bit 15	Bit 16 … Bit 31
Source Port (16)	Destination Port (16)
Length (16)	Checksum (16)
Data (if any)	

8 bytes

Figure 2-5 UDP Segment

0	4	10	16 … 31
Source Port			Destination Port
Sequence Number			
Acknowledgement Number			
Data Set	Reserved	Code Bits	Window
Checksum			Urgent Pointer
Options			
Data			

Figure 2-6 TCP Header Format

2.4 The Real-time transport protocol

The real-time transport protocol (RTP) provides end-to-end transport functions for real-time audio and video applications [6]. RTP provides payload type identification, sequence numbering, timestamping and delivery monitoring services. It does not guarantee quality of service. Quality of service on the internet and in other networks corresponds to the idea that transmission rates, error rates and other characteristics can be measured, improved, and, to some extent, guaranteed in advance. Typically applications run RTP on top of UDP making use of its multiplexing and checksum services. However, RTP may be used with other network or transport protocols. RTP neither guarantees delivery nor prevents out-of-order delivery. At the same time it does not assume that the underlying network is reliable and delivers packets in sequential order. The receiver reconstructs the transmitted packets in sequence using the sequence number and timing information in the RTP packet. Usually RTP is integrated into the application program Real-Time Control Protocol (RTCP) works in conjunction with RTP. RTCP control packets are periodically transmitted to all other participants by each participant in an RTP session. The RTP control protocol monitors quality of service and conveys information about participants in an on-going session. Its primary function is to provide feedback on the quality of data distribution. The RTP packet is a data packet consisting of a fixed RTP header, a list of sources (may be empty) and payload data. Encapsulation of the RTP packet may be needed by the underlying protocols. Normally a packet of the underlying protocol has a single RTP packet. The format of the RTP header is as shown in Figure 2-8.

IP Header	UDP Header	RTP Header	RTP payload

Figure 2-7 The Structure of Voice Packet

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td colspan="2">V</td><td>P</td><td>X</td><td colspan="4">CSRC count</td></tr>
<tr><td colspan="2">M</td><td colspan="6">Payload type</td></tr>
<tr><td colspan="8">Sequence number (2 bytes)</td></tr>
<tr><td colspan="8">Timestamp (4 bytes)</td></tr>
<tr><td colspan="8">SSRC (4 bytes)</td></tr>
<tr><td colspan="8">CSRC (0-60 bytes)</td></tr>
</table>

Figure 2-8 The RTP header format

V identifies the version of the RTP. If the padding bit (P) is set one or more additional padding octets are placed at the end of the packet. The last octet of the padding has a count of how many padding octets should be ignored. Padding may be needed for carrying several RTP packets in a lower - layer protocol data unit or for some encryption algorithms with fixed block sizes. X indicates an extension to the fixed header. The extension mechanism allows experimentation with new payload-format-independent functions that require additional information to be carried in the RTP data packet header. The contributing source (CSRC) list identifies the contributing sources for the payload contained in this packet. The CSRC count (CC) field indicates the number of CSRC identifiers that follow the fixed header. A profile defines the interpretation of the marker (M). It allows significant events such as frame boundaries to be marked in the packet stream. A profile specification document defines a set of payload type codes and their mapping to payload formats (e.g., media encodings). A profile may also define extensions or modifications to RTP that are specific to a particular class of applications. Typically an application will operate under only one profile. The profile may define additional marker bits or specify that there is no marker bit by changing the number of bits in the payload type field.

The Payload Type (PT) field identifies the format of the RTP payload and its interpretation by the application. A default static mapping of payload type codes to payload formats is specified by the profile. Seven bits are allocated to this field. Sixteen bits are reserved for storing the sequence number. The sequence number which increments by one for each RTP data packet is used by the receiver to detect packet loss and for restoration of packet sequence. A random value is chosen for the start of the sequence number to make known-plaintext attacks on encryption more difficult.

The information pertaining to the sampling instant of the first octet in the RTP data packet is contained in the timestamp field. 32 bits are used to represent this field. The sampling instant is derived from a clock that increments monotonically and linearly in time allowing synchronization and jitter calculations. The resolution of the clock must be sufficient for the desired synchronization accuracy and for measuring packet arrival jitter. Time is represented in seconds using the timestamp format of the Network Time Protocol (NTP) [21]. As in the case of sequence numbers the initial value of the timestamp is random.

The synchronization source (SSRC) identifier field represented by 32 bits identifies the synchronization source. This identifier is chosen randomly so that no two synchronization sources within the same RTP session will have the same SSRC identifier. In the case where participants in one area are connected through a low-speed link to the majority of participants who enjoy high-speed network access, instead of forcing everyone to use a lower-bandwidth, reduced-quality audio encoding, an RTP-level relay called a mixer may be placed near the low-bandwidth area. This mixer resynchronizes incoming audio packets, mixes these reconstructed audio streams into a single stream, translates the audio encoding to a lower-bandwidth one and forwards the lower- bandwidth packet stream across the low-speed link. Only in the case when a mixer is involved the list of CSRC identifiers are present.

2.5 Types of Queuing

The type of queuing on the device's interface manages the way packets sent through the interface are queued. When there is more data than can be sent immediately the queuing determines the priority to send packets. Thus traffic congestion management is done at the interface especially important at intermediate nodes (routers). The queue refers to the collection of packets waiting for processing in physical blocks of memory (buffers). A node may also receive more packets than it can process. The buffer holds the packets until the node can catch up. If it cannot catch up, the buffers fill up and the new incoming packets are dropped. This is termed as a tail drop. Four common queuing algorithms are: first-come-first-serve, weighted fair queuing, priority queuing and custom queuing. In Cisco routers weighted fair queuing is used by default for serial interfaces at 2.048 Mbps and below [12]. All other interfaces use FCFS by default when no other queuing strategy is configured.

2.5.1 Weighted Fair Queuing

Weighted Fair Queuing (WFQ) provides traffic priority management that automatically sorts among individual traffic streams without requiring defined access lists [12]. This queuing technique can manage duplex data streams such as those between pairs of applications and simplex data streams such as voice or video. From the perspective of WFQ there are two categories of data streams namely high-bandwidth sessions and low-bandwidth sessions. Low-bandwidth traffic has effective priority over high bandwidth traffic. High-bandwidth traffic shares transmission service proportionally based on the weights assigned. Thus high-bandwidth traffic is prevented from having an unfair share of resources. New messages for high-bandwidth conversations are discarded after the default or set congestive-messages threshold is met. However low-bandwidth sessions including control-message conversations continue to enqueue data. As a result the fair queue may sometimes have more messages than specified by the threshold value. The WFQ moves real-time interactive traffic, typically low-bandwidth to the front of queues and shares the remaining bandwidth fairly among other flows. IP precedence is also considered in determining the allocated bandwidth. The Versatile Interface Processors (VIP) refers to the interface card used by Cisco 7500 series and Cisco 7000 series with Route Switch Processor 7000 (RSP7000) routers. When the interface has Versatile Interface Processors (VIP) cards the queues are treated with the same weight. This is called distributed weighted fair queuing. DWFQ uses the VIP to provide a faster implementation of WFQ. The whole processing of the WFQ algorithm is done on the VIP.

2.5.2 First-in First-out (FIFO) queuing

FCFS or FIFO queues are a basic store and forward technique. Packets are queued on a first come first served basis. The packet waiting the longest is transmitted first. When the queues become full, congestion is said to occur and the incoming packets are dropped. FIFO queuing technique relies on end systems to control congestion using congestion control mechanisms. FIFO queuing works well on uncongested high-capacity links having minimal delay or when differentiation of services for packets traveling through the device is not needed. One big disadvantage with this queuing technique is that a station starting a file transfer can consume all the link bandwidth. This is referred to as a packet train and can be detrimental to interactive sessions.

2.6 VOIP Performance Requirements

In this section we look at the network parameters having an impact on voice quality in IP telephony. We take a look at the parameters that have a significant effect on the quality of voice. Network Packet Delay corresponds to the time taken for a packet to traverse the network. The typical end-to-end VoIP delay ranges from 150 - 400 milliseconds (ms). Delay values upto 200 ms is acceptable even for business purposes [4]. The G.114 ITU-T standard describes that a 150 milliseconds one-way delay is acceptable for high voice quality. Every element of the network adds to the packet delay. This is inclusive of the delay introduced by the routers, switches and delay buffers. The delay added on by the routers depends on the hardware and configuration settings such as queuing method, access list and transmission modes. There are inherent delays in a public network that one cannot control. Delay can be better controlled in a private network environment.

Network Jitter corresponds to the variation in inter-packet arrival time due to the variable transmission delay over the network. A jitter buffer is incorporated at the receiver to compensate for the network jitter. To have a smooth packet flow the jitter buffer holds incoming packets for a certain period of time before decompressing them. The inclusion of the jitter buffer causes additional packet delay. The size of the jitter buffer is generally twice the largest statistical variance between packets. Dynamic jitter buffers give good voice quality. Apart from the choice of the size of the jitter buffer, the queuing method on the router has a significant impact on jitter. The topology of the network also affects jitter. For example the jitter is less on a hierarchical switched network than on a flat hub-based network due to lesser number of collisions.

Packet Loss results when packets sent are not received at the final destination. For good voice quality in a VoIP network, packet loss of compressed speech such as G.723 and G.729 should be not be more than 1% [5]. A higher packet loss is acceptable in the case of G.711. When the voice payload is larger the packet loss becomes more noticeable as a larger chunk of the signal is lost. Also continuous packet loss is worse than random packet loss over time. When a node receives an RTP audio stream for Voice over IP it must compensate for the jitter that is encountered. The mechanism that handles this function is the playout delay buffer at the end nodes. The playout delay buffer must buffer these packets and then play them out in a steady stream to the digital signal processors (DSPs) to be converted back to an analog audio stream. The playout delay buffer is also sometimes referred to as the de-jitter buffer. Packets are discarded if they arrive later than the jitter buffer can hold them since they cannot be played. In voice over IP networks the individual packets are sent over different routes. This may result in misordered packets. Packet misorder may also be because of re-routing due to congestion or due to load balancing.

2.7 Compression Standards and Codec Parameters

The CODEC (CODER/DECODER) is a standard through which voice information can be encoded into data or decoded back to voice information. Speech compression produces a compact representation of speech sounds which when reconstructed is close enough to the original to be understandable. Speech codecs are divided into three classes - waveform codecs, source codecs and hybrid codecs. Waveform codecs produce high bit rates and result in very good speech quality at the receiver. The simplest form of waveform coding is Pulse Code Modulation involving merely sampling and quantizing the input waveform [5]. Source codecs operate at very low bit rates but produce speech sounding synthetic [5]. Hybrid codecs give good quality speech at intermediate bit rates. They use techniques from both source and waveform coding. A common benchmark used to determine the quality of sound produced by specific codecs is the mean opinion score (MOS) [5]. A wide range of listeners judge the quality of a voice sample for a particular codec on a scale of 1 to 5. One corresponds to bad sound quality and five corresponds to excellent sound quality. The scores obtained are then averaged providing the MOS for the sample. Though low-bit rate codecs save on infrastructure costs there is signal distortion due to multiple encodings and codec-induced delay. Table 2-1 and Table 2-2 show the MOS scores for the codecs and the codec-induced delay.

Table 2- 1 Compression Methods and MOS Scores

Compression Method	Bit Rate (kbps)	Framing Size	MOS Score
G.711 PCM	64	0.125	4.1
G.726 ADPCM	32	0.125	3.85
G.728 LD-CELP	16	0.625	3.61
G.729 CS-ACELP	8	10	3.92
G.729 x 2 Encodings	8	10	3.27
G.729 x 3 Encodings	8	10	2.68
G.729a CS-ACELP	8	10	3.7
G.723.1 MP-MLQ	6.3	30	3.9
G.723.1 ACELP	5.3	30	3.65

Table 2- 2 CODEC-Induced Delays

CODEC	Compression Delay (ms)
G.711 PCM	0.75
G.726 ADPCM	1
G.728 LD-CELP	3 to 5
G.729 CS-ACELP	10
G.729a CS-ACELP	10
G.723.1 MP-MLQ	30
G.723.1 ACELP	30

In speech coding the value of the next sample is predicted from previous samples due to the correlations present in speech samples. Differential Pulse Code modulation schemes quantize the difference between the original and predicted signals. Adaptive differential PCM (ADPCM) exploits this redundancy using a simple linear filter to predict the next speech sample. ADPCM is simple to implement and has very low delay. The complexity of a coding algorithm is the processing effort needed to implement the algorithm and is measured in terms of arithmetic capability and memory requirement. A large complexity can result in high power consumption in the hardware. Speech coders often process speech in blocks. This processing introduces communication delay. Coding efficiency is expressed in bits per second (bps).

2.8 Summary of Prior Research

In the study by Boonchai Ngamwongwattana the model created represents best-effort Internet service. The model assumes a one-way average inherent delay of 70 milliseconds and a standard deviation of 10 milliseconds. The voice call traffic is assumed to be a constant bit rate in which techniques, such as voice activation detection, are not applied. The queue in the model represents the weakest network link, essentially in the access network that the voice call must cross. The link capacity is assumed to be 256 Kbps as the bottleneck. The background traffic shares the same queue as the voice flow is modeled as a Poisson arrival. Wireless hop is added to the existing wired network with two different queuing techniques namely FIFO and WFQ. Varying packet delay can change end-to-end delay. Smaller packet delay does not always decrease the end-to-end delay [13]. The packet delay resulting in minimum end-to-end delay varies depending on the codec [13]. For G.711 18 to 20 milliseconds (144 to 160 bytes of payload size) is needed [13]. For G.726 it is 14 to 16 milliseconds (56 to 64 bytes of payload size) [13]. For G.729 the minimum end-to-end delay is at 20 milliseconds (20 bytes of payload size) [13]. Maximum number of VoIP users over 802.11(2 Mbps) is 12 [23]. But this is just bandwidth not delay. The distribution of packet delay for VoIP over wired networks has been used to better define the playback point at the receiving end of a voice application [17].

Chapter 3

EXPERIMENTAL RESULTS

3.1 Measurement System Setup

In this chapter, the actual experimental setup used to establish the effect of packetization delay on the end-to-end delay for different load conditions is discussed. The setup used for the experiment is as shown in Figure 3-1. The key software components of this setup consists of a network emulation package NISTNet [9], a traffic generator MGEN [10] and a protocol analyzer Ethereal [11]. The hardware components employed in this setup include a pair of Cisco 3640 series routers [12] with voice over IP capabilities, a pair of Cisco 2500 series routers representing a WAN network link, a pair of analog phones for call setup, a Cisco Aironet 1200 series wireless access point and a WaveLAN Ethernet Converter comprising the wireless part of the network. Apart from the above hardware components Cisco 2900 series switches, Cisco 1900 series switches and hubs are used to connect the various components. Each component of the setup is described in detail below.

The Cisco 3640 series router pair, Rc1 and Rc2 is directly connected to two analog phones through the FXS interface. These routers provide analog and digital voice capability using VIC-2FXS voice interface cards. Using this packet voice technology, voice is sampled, digitized, coded and packetized for various packetization delay values using the G.729 codec. Constant bit rate traffic is generated.

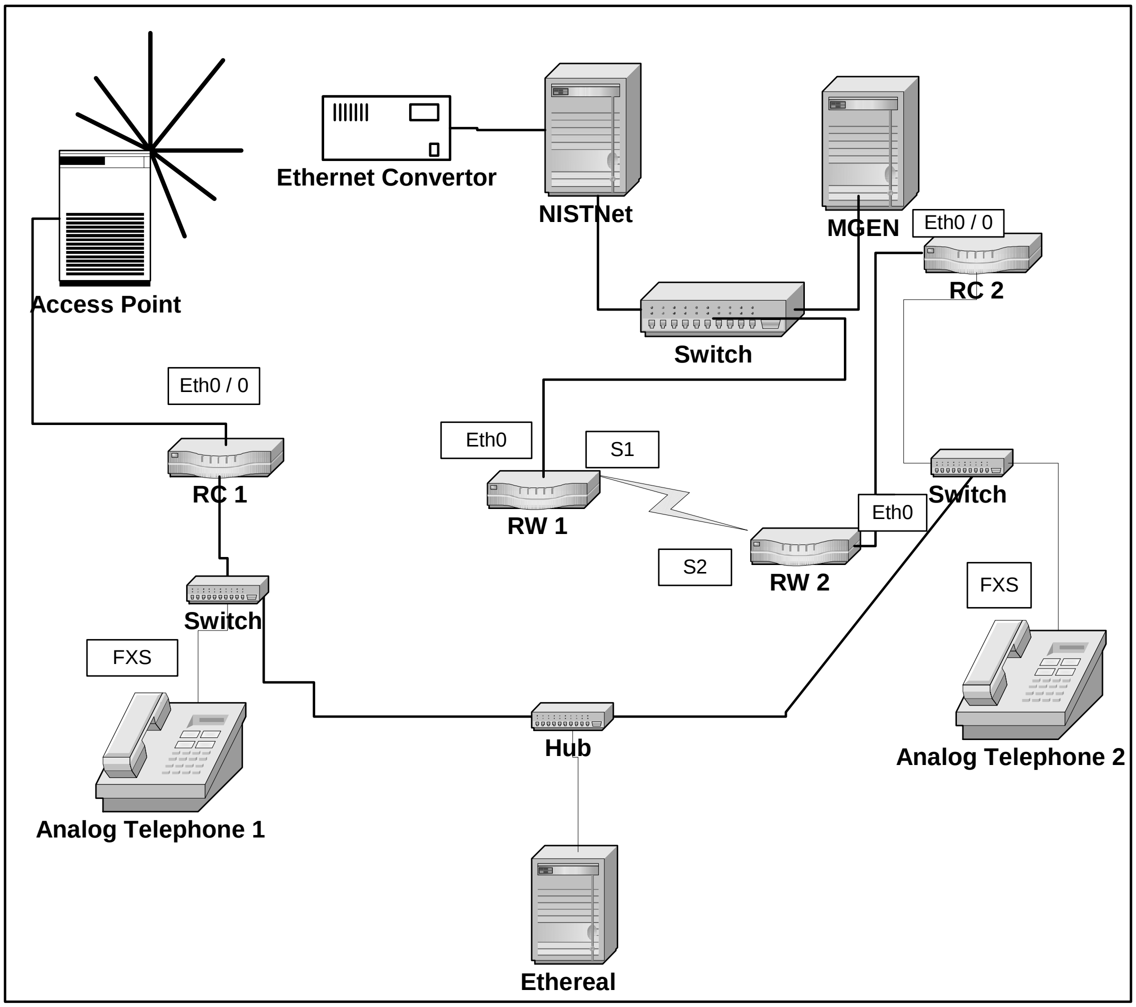


Figure 3-1 The Experimental Setup

The Cisco wireless access point connected to the Cisco 3640 router transmits the generated voice packets over 802.11b wireless LAN environment. This access point connects the

Ethernet network to the wireless network. The WaveLAN Ethernet Converter receives the voice packets generated by the access point and hands it over to the network emulator in the Ethernet network.

The NIST Network Emulation tool (v.2.0.10) running on Linux operating system allows a wide range of parameters. They can be set depending on the network condition to be emulated. The PC running NIST Net emulates a network of routers with two Ethernet cards each supporting a LAN subnetwork. Among the various parameters, the inherent Internet delay is what is set using this software. The Pareto-normal delay distribution is chosen with a mean of 70 milliseconds and a standard deviation of 10 milliseconds corresponding to reference studies from [13].

The MGEN packet generator running on Sun Solaris operating system generates background traffic emulating different levels of load as in the Internet [10]. The rate of packet generation for the required load condition is computed as the product of the link capacity and required load percentage divided by the packet size. The packet size is set at a value of 250 bytes, the average size of a packet in the Internet [12]. It is assumed as traditional data traffic and hence the pattern of traffic generation chosen is Poisson [13]. This traffic flows through the WAN link and is directed to the receiving router on the WAN link rather than emulating multiple flows. To observe the variation in end-to-end delay with load, the offered load levels are varied from 0% to 50% and increased until the delay value is above 200 milliseconds, which is unacceptable for voice traffic.

The Cisco 2500 series routers, Rw1 and Rw2 emulate a wide area network. The capacity across this link was restricted to 250 Kbps so that this link acts as a bottleneck for the voice packets that flow through them [13].

Effective Voice Bandwidth = f / t_p Kbps

Overhead Bandwidth = $h / N\, t_p$ Kbps

Where

f	Size of compressed data (bit)
t_p	Voice frame delay (ms)
h	Header size of voice packet (bit)
N	The number of compressed data frames in payload

The network bandwidth is the sum of the effective voice bandwidth and the overhead bandwidth. Substituting values for the G.729 codec we have

Effective Voice Bandwidth = 8 Kbps

(a) Packetisation Delay = 10 ms

Overhead Bandwidth = 32 Kbps

Network Bandwidth = 40 Kbps

(b) Packetisation Delay = 60 ms

Overhead Bandwidth = 5.33 Kbps

Network Bandwidth = 13.33 Kbps

For a load of 50% we see that around 64 packets of size 250 bytes are issued by MGEN per second. This comes to 128 Kbps. Guided by the above calculations and practically checking the bandwidth consumed with the help of the NIST Net emulator we settle on a value of 250 Kbps on the WAN link. The background traffic generated by the packet generator sets the load at different levels, thus emulating the model of the Internet. The queuing technique enabled on these routers is also considered. The default queuing mechanism on routers is the Weighted Fair queuing (WFQ). The other popular queuing mechanism enabled on routers is the First-In First-out (FIFO). The Internet as we see it today is not optimized for transmission of voice. By default it does not have priority queuing enabled specifically for voice packets [12]. Hence the WFQ and FIFO queuing mechanisms are considered here. The two factors that affect the packet drop across the network are the router's queue and the nature of the wireless environment

The Ethereal protocol analyzer (version 0.9.1) captures the data packets sent over the live network connection, from the caller to the called party [11]. Data can also be read from a capture file, programmatically edited and browsed via a GUI. Currently 366 protocols can be dissected. Live data can be read from Ethernet, FDDI, PPP, Token-Ring, IEEE 802.11, classical IP over ATM, and loopback interfaces. Display filters may be used to refine data display. Switches are used at the sending and receiving end. For precision and accuracy synchronizing clocks at sender and receiver located at two distant points is needed. Since network emulation is used in the setup, the transmitting and receiving devices are placed nearby and simple measurement can be effective, overcoming problems with respect to fault tolerance, cost and administration. By enabling port mirroring on them a copy of every voice packet at the sending and receiving end is captured. The switches are in turn connected to a hub. The voice packets captured at both ends is sent to the protocol analyzer also connected to the hub. Each voice packet generated has an RTP sequence number stamped on it. The end-to-end delay experienced by every packet, identified by the RTP sequence number is calculated with the help of the time-stamp on it. Subtracting the time-stamps on the packets at the sending and receiving end gives the network delay. Adding the value of packetization delay to this gives the total end-to-end delay. The packetization delay is varied from 10 to 60 milliseconds. The voice packets are generated from recorded speech played for a time interval of 120 seconds. The number of packets transmitted depends on the packetization delay. At a value of 10 milliseconds around 10,000 packets are generated. When the set packetization delay is 60 milliseconds the number of packets reduces to around 2000 packets. The packet loss depends on the background load.

3.2 Performance Results

3.2.1 WFQ over a wired network

In the experiment when weighted fair queuing is enabled on the routers in the WAN link the end-to-end delay changes as shown in figure [3-4]. Figure [3-2] shows the percentile 10 and 90 percentile value. This network is completely wired. The delay increases in a linear fashion. As the load on the network (simulated by the MGEN traffic generator) increases the delay increases.

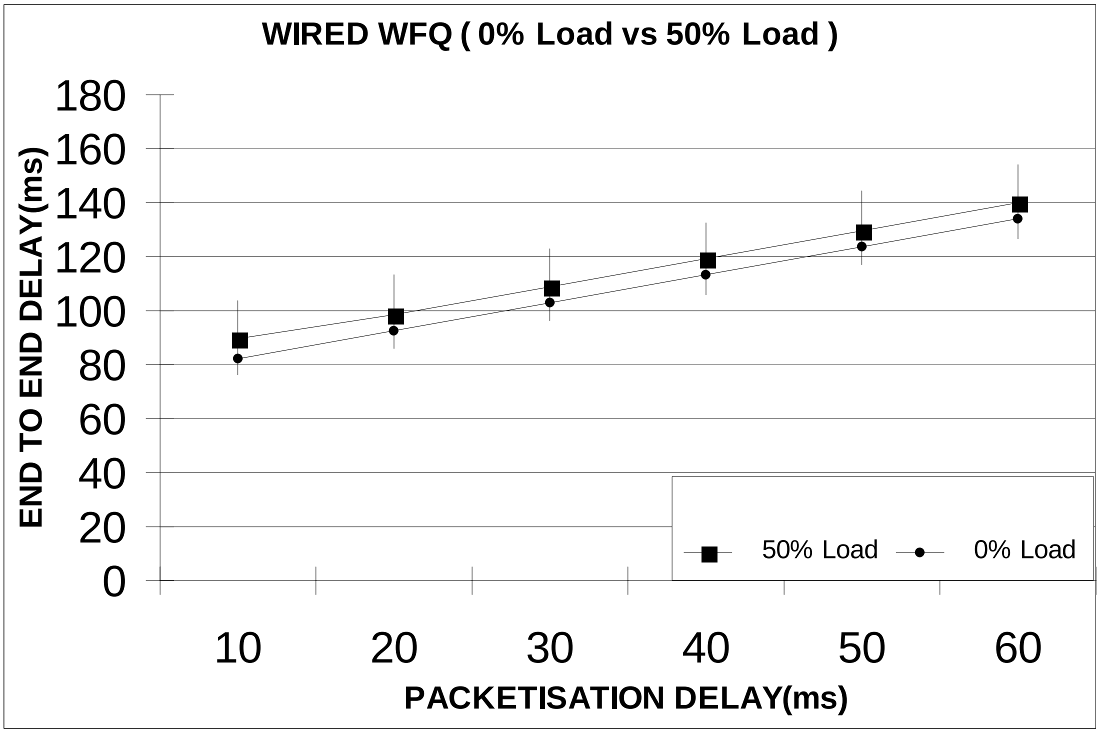


Figure 3-2 Percentile Values for End-to-End delay for Wired WFQ

Figure 3-2 shows the percentile values for end-to-end delay for Wired WFQ. Packetization delay value of 10 milliseconds when using WFQs results in an 82 ms delay (± 10 ms). This packetization delay when summed with the network delay it yields the lowest end-to-end delay. The network delay is got as the difference between the time stamps on the packet at the destination and at the source as shown in Figure 3-3. Each packet is uniquely identified using the RTP sequence number on it.

	RECEIVED PACKET	**TRANSMITTED PACKET**
No	94	2
Len	64	64
Time	**0.178735**	**0.003588**
Src hw	Cisco_55:f0:22	Cisco_75:4b:00
Source	192.168.117.52	192.168.117.52
Destination	32.0.0.51	32.0.0.51
Protocol Info	RTP Payload type=ITU-T G.729, SSRC=136869172, **Seq=31308**, Time=4280942295	RTP Payload type=ITU-T G.729, SSRC=136869172, **Seq=31308,** Time=4280942295

Network Delay	0.178735 - 0.003588 0.175147 seconds
Packetization Delay	0.010 seconds

End-to-End Delay	0.185147 seconds

Figure 3-3 Calculation of End-to-End delay with a live example

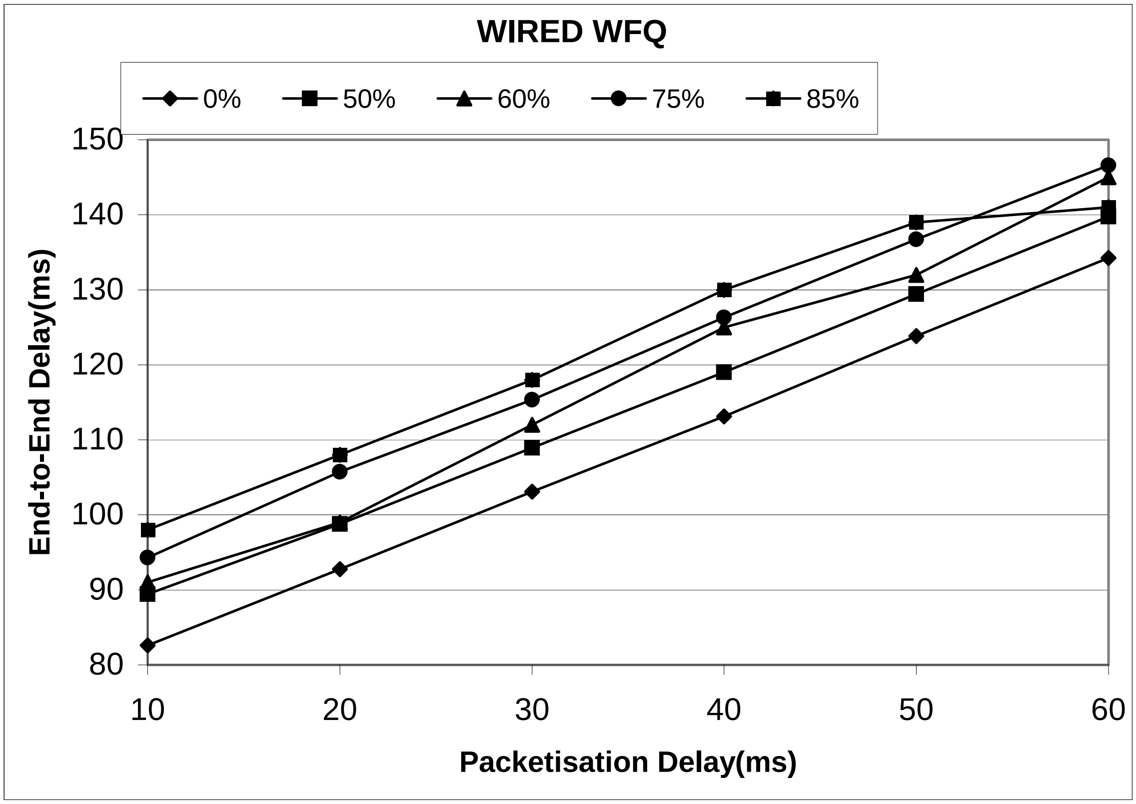


Figure 3-4 Delay Characteristic for WFQ over a wired network

Figure 3-4 shows the delay characteristic for WFQ over a wired network. The packetization delay is varied from 10 ms to 60 ms. When the size of the packet is large the end-to-end delay increases due to packetization delay. When the load on the network increases the end-to-end delay curve shifts up. Hence we see that the end-to-end delay increases from 82 ms at a packetization delay of 10 ms with 0% load to 149 ms at 75% load with a packetization delay of 60 ms. Small packetization delay relates to a small payload size but the overhead bandwidth consumed will be more. Even when the load on the network is very high the maximum end-to-end delay is around 150 milliseconds. These sets of curves are statistically different. The t-test is used to compare the distribution of two population of points. The packet loss remained zero even when the offered load is as high as 80%.

3.2.2 WFQ over wired network with a wireless subnet

When a wireless subnet is added as shown in Figure 3-6 the end-to-end delay in the network still follows a linear trend for packetization delay values from 20 ms to 60 ms. Figure 3-5 shows the percentile values. The delay value increases by around 7 milliseconds when compared to a fully wired network. The end-to-end delay touches a value of 160 milliseconds. The above stated behavior is not observed when the packetization delay is 10 milliseconds. At this packet size the end-to-end delay is the highest 140 ms.

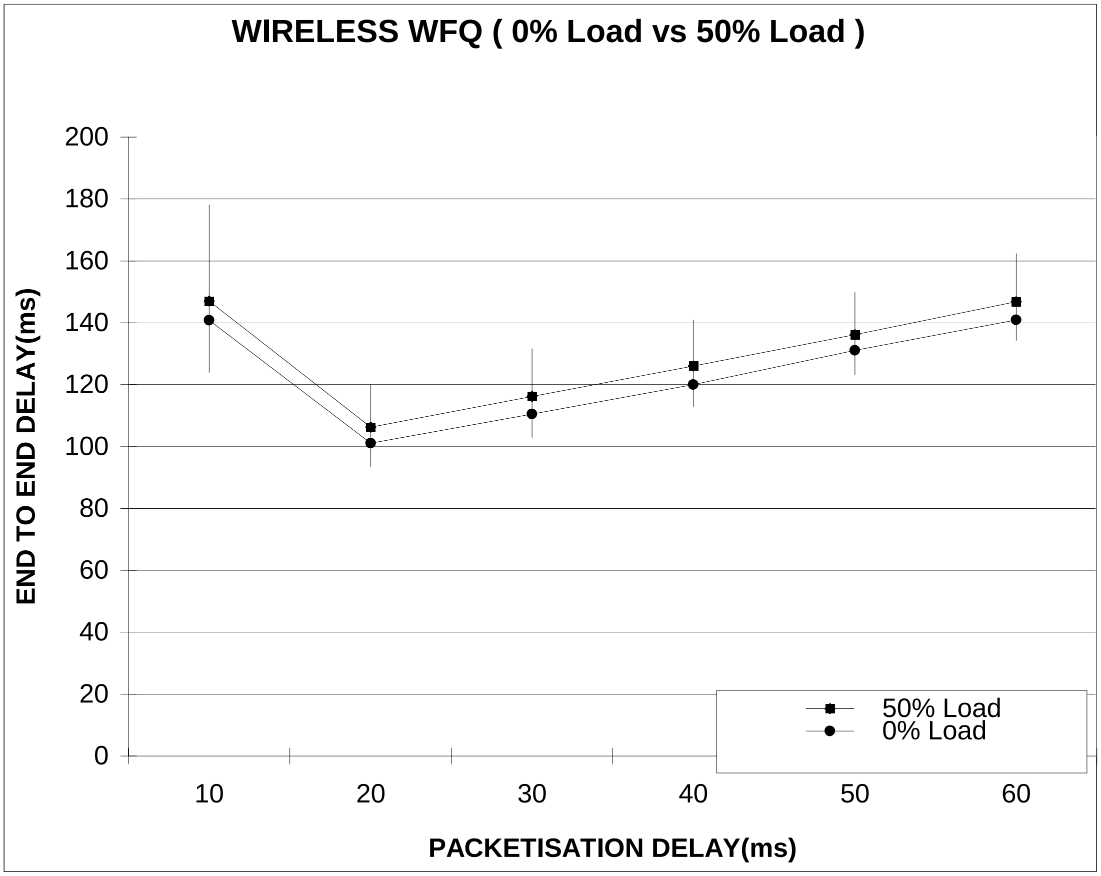


Figure 3-5 Percentile Values for End-to-End delay for Wireless WFQ

Figure 3-5 shows the percentile values for end-to-end delay with wireless WFQ. Comparing the curves in Figure 3-2 representing wired WFQ and Figure 3-5 representing wireless WFQ the curves shift up with increase in load. The additional delay with the increase of load to 50% in the case of both wired WFQ and wireless WFQ is 8 ms for a packetization delay of 20 ms, 7 ms for 30-50 ms packetization delays and 6 ms for a packetization delay of 60 ms. Difference in the shape of the curves is seen at 10 ms when the delay is 58 ms for wireless WFQ.

The dip in the delay curve when the packetization delay changes from 10 ms to 20 ms is statistically significant.

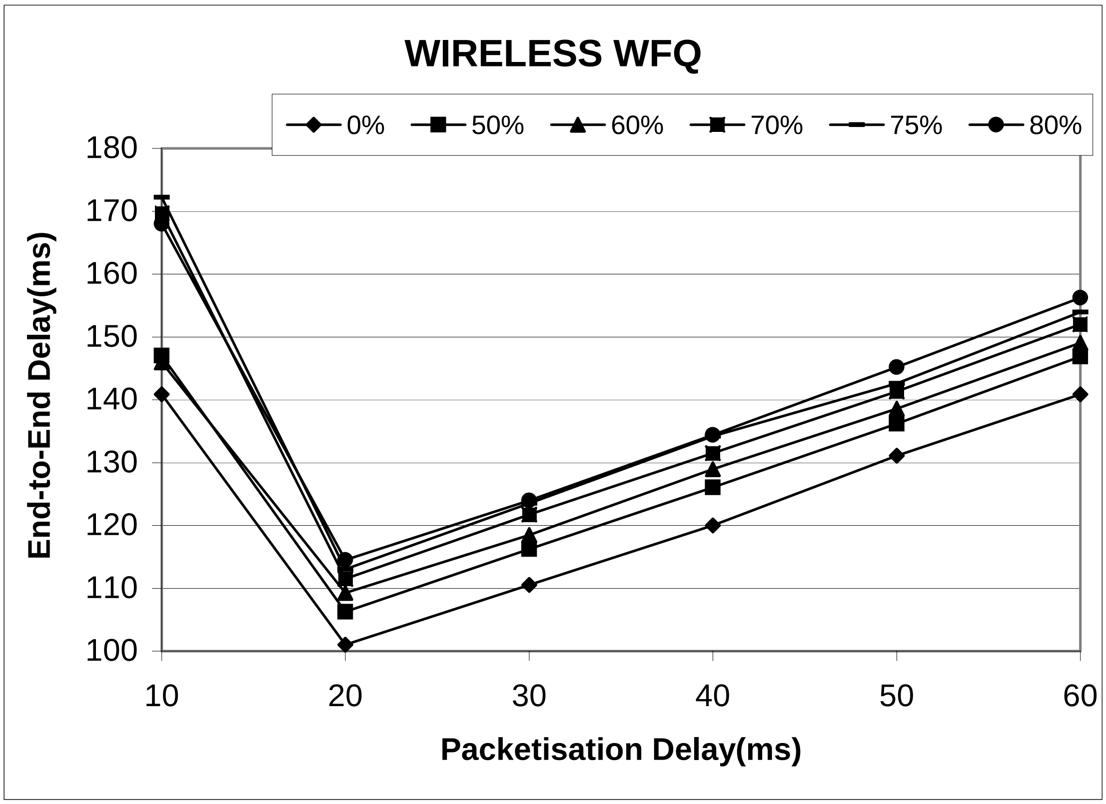


Figure 3-6 Delay Characteristic for WFQ over a wireless subnetwork

Figure 3-6 shows the delay characteristic for WFQ over a wireless subnetwork. When the size of the packet is 10 ms the end-to-end delay is 140 ms. Packetization delay curve shifts to the right compared to wired WFQ. It is seen that the packet delay producing minimum end-to-end delay packetization delay is 10 ms for wired WFQ while it is 20 ms for wireless WFQ. As the load on the network increases the end-to-end delay curve shifts up. Hence we see that for 0% load, the end-to-end delay dips from 141 ms at a packetization delay of 10 ms to 101 ms at 20 ms packetization delay and increases back to 141 ms as compared to a dip to 113 ms from 172 ms at 10 ms packetization delay accompanied by a rise to 154 ms for a packetization delay of 60 ms when the load is 75%. These sets of curves are statistically different. However the overall shape stays the same.

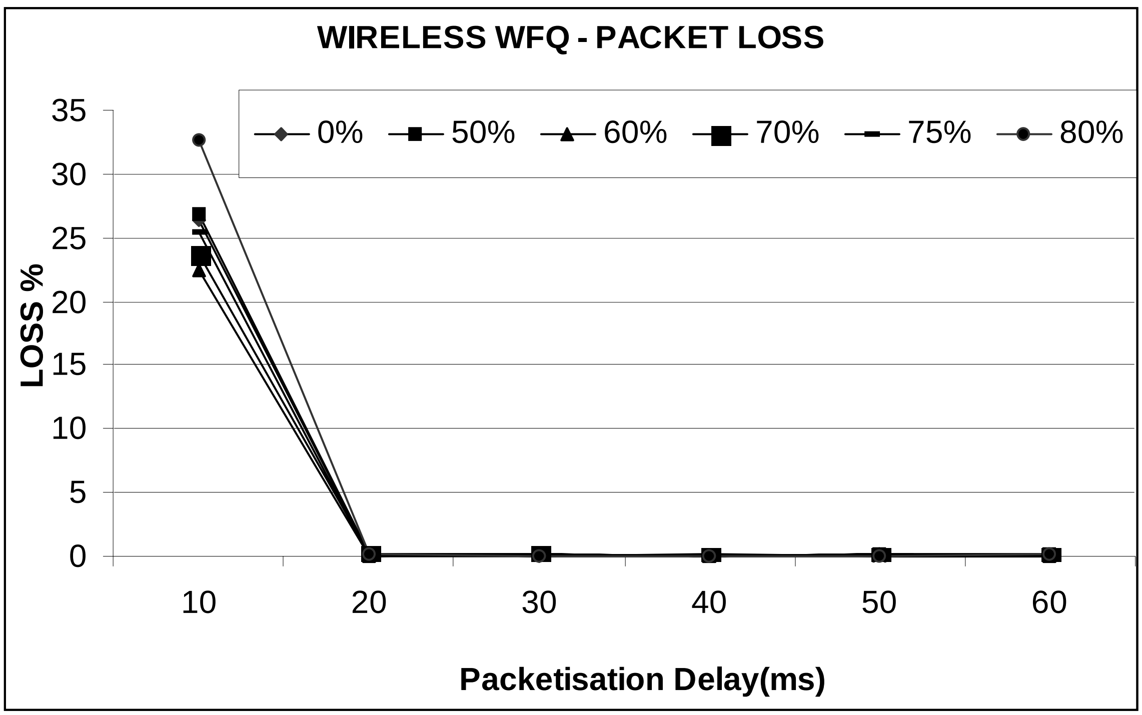


Figure 3-7 Packet Loss percentage for WFQ over wireless network

Figure 3-7 shows the packet loss for WFQ over a wireless subnetwork. The packet loss is the highest 26.44% at a packetization delay of 10 milliseconds.

3.2.3 FIFO over wired network

On enabling FIFO queuing strategy the end-to-end delay follows a linear trend for low load conditions. Here again the lowest end-to-end delay is observed at 10 milliseconds. As the load on the network increases to around three-fourth the link capacity the end-to-end delay. This happens only when the packet size is at its smallest i.e. 10 milliseconds. Figure [3-10] shows this change in the trend. Figure [3-8] shows the percentile values. When the network load exceeds three-fourths the link capacity the delay is greater than 200 ms for all packet sizes. The packet loss follows the curves as in figure [3-9]. Previous studies on end-end delay as a function of packetization delay with the G.729 codec, using the same experimental network yield similar results. The delay characteristics follow the same trend. We observe a difference in the load value at which the end-to-end delay becomes unacceptably high. The end-to-end delay was below 200 ms at network loads upto 80% [13] while Figure [3-9] shows end-to-end delay upto 70%. This variation could be due to difference in value of the link capacity on the WAN link.

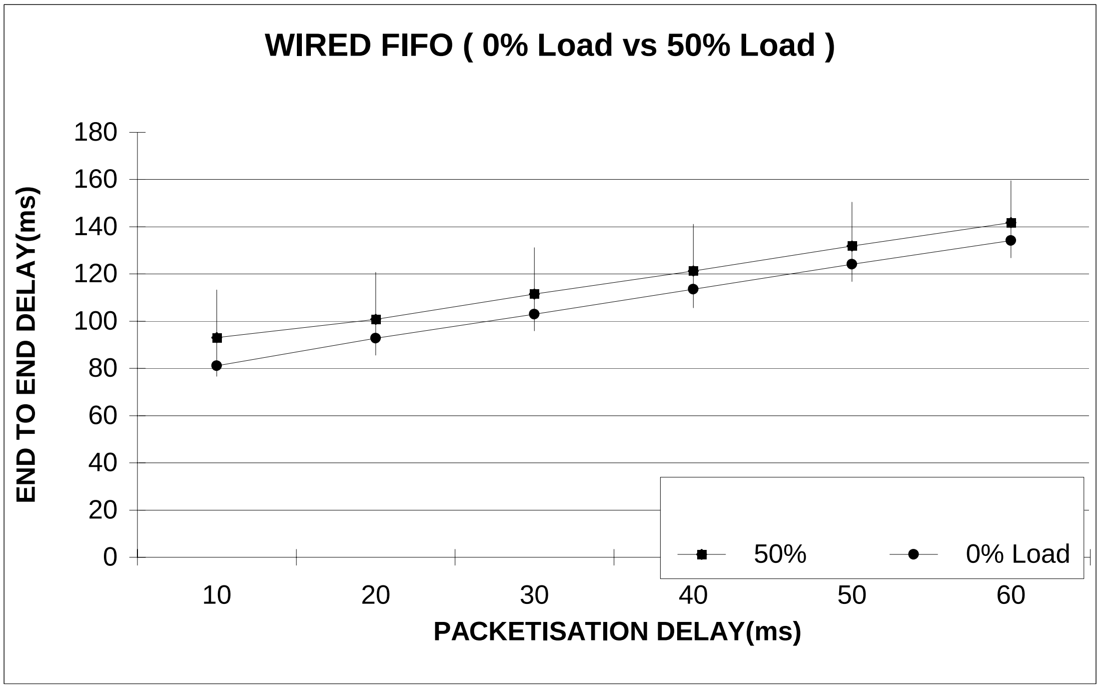


Figure 3-8 Percentile Values for End-to-End delay for Wired FIFO

Figure 3-8 shows the percentile values for end-to-end delay for wired FIFO at 0% load and 50% load. Comparing this to WFQ shown in Figure 3-2 the minimum end-to-end delay is still at 10 ms. When packetization delay of 10 ms is used at high loads (more than 70%) there is network congestion.

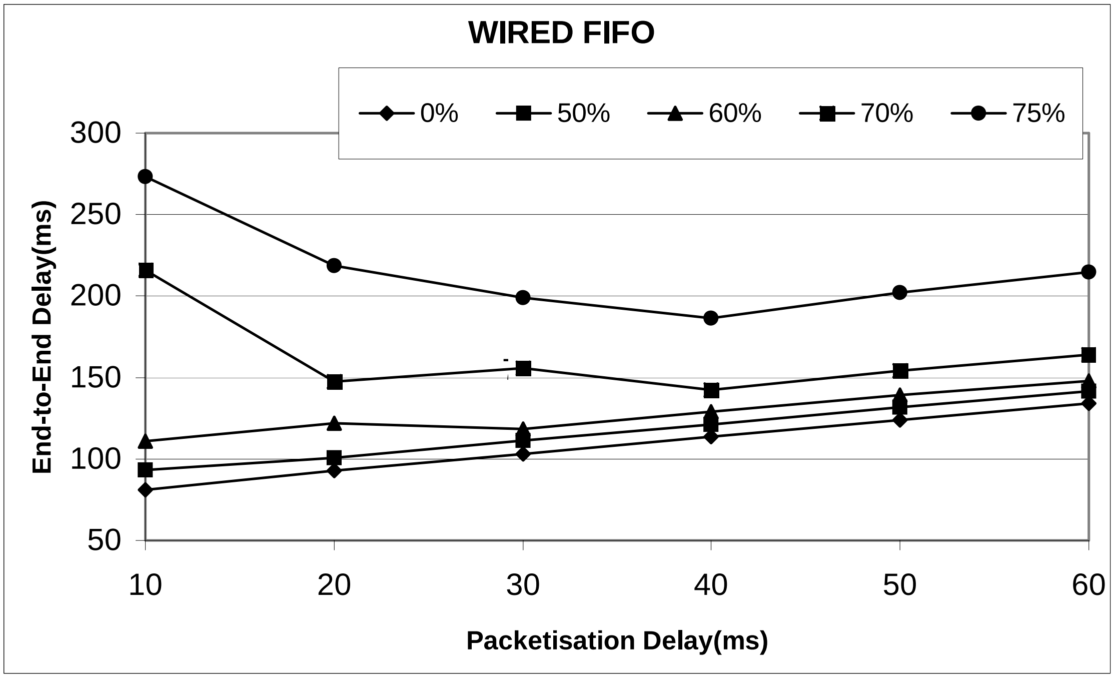


Figure 3-9 Delay Characteristic for FIFO over a wired network

Figure 3-9 shows the delay characteristic for FIFO over a wired network. With FIFO queuing strategy at low loads the end-to-end delay increases with increase in size of packets due to packetization delay. As the load on the network increases the end-to-end delay curve shifts up. Hence we see that the end-to-end delay increases from 81 ms at a packetization delay of 10 ms with 0% load to 215 ms at 75% load with a packetization delay of 60 ms. At high loads the packetization delay of 10 ms yields high end-to-end delay values. For 75% load, the end-to-end delay dips from 273 ms at a packetization delay of 10 ms to 218 ms at 20 ms packetization delay. These sets of curves are statistically different. The bottom of the 75% curve is at 40 ms packetization delay. Packetization delays from 30 – 50 ms meet the ITU specification. Comparing Figure 3-9 for wired FIFO, Figure 3-4 for wired WFQ and Figure 3-6 for wireless WFQ the inflection point happens for wired FIFO and wireless WFQ but not wired WFQ.

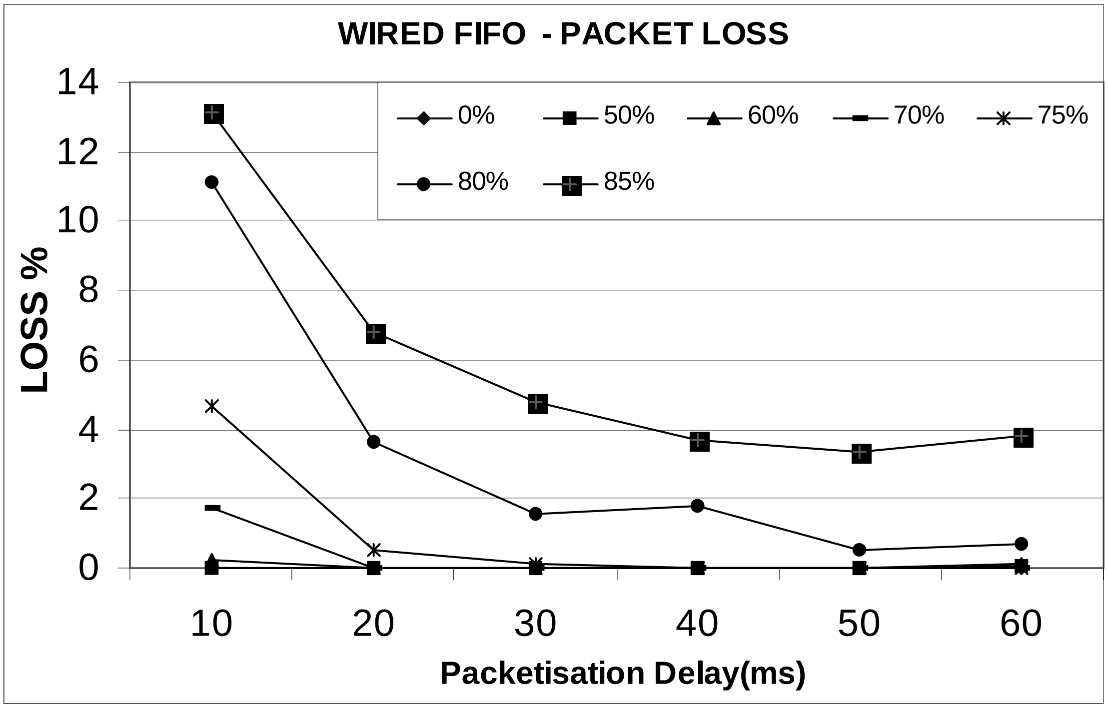


Figure 3-10 Packet Loss percentage for FIFO over a wired network

Figure 3-10 shows packet loss percentage for FIFO over a wired network**.** When WFQ is used the packet loss in the wired network is zero.

3.2.4 FIFO over wired network with a wireless subnet

The packetization delay of 10 milliseconds yields the worst results for all acceptable load values as shown in figure [3-12] and figure [3-13]. Figure [3-11] shows the percentile values. Keeping apart values at this packet size, the trend is similar to that of wired networks.

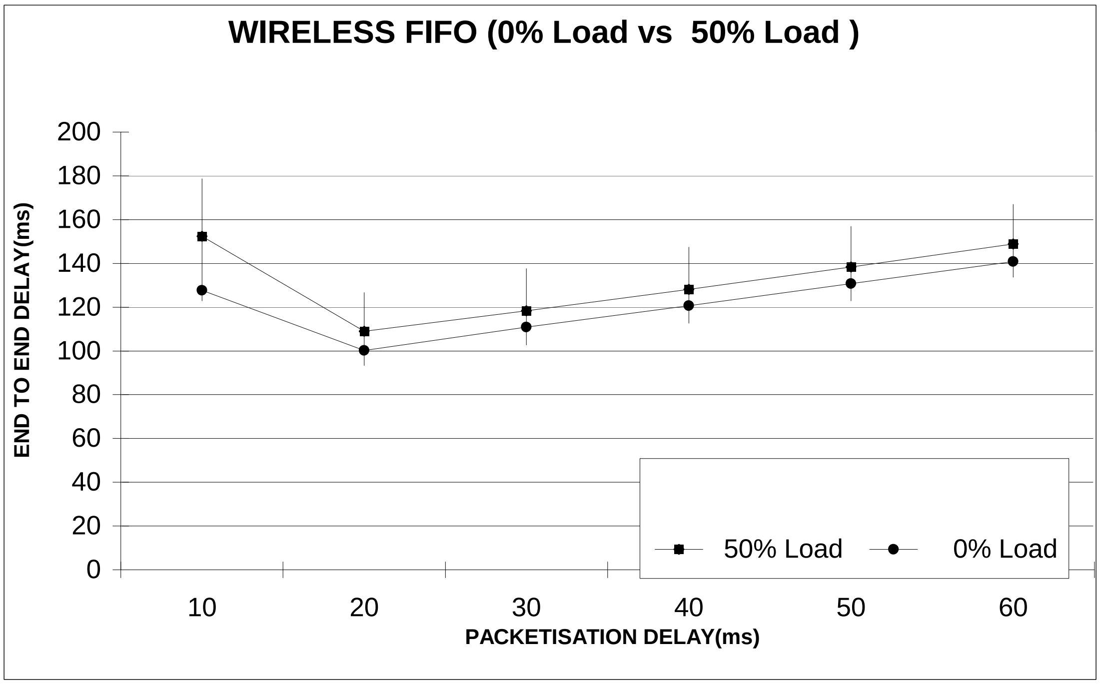


Figure 3-11 Percentile Values for End-to-End delay for Wireless FIFO

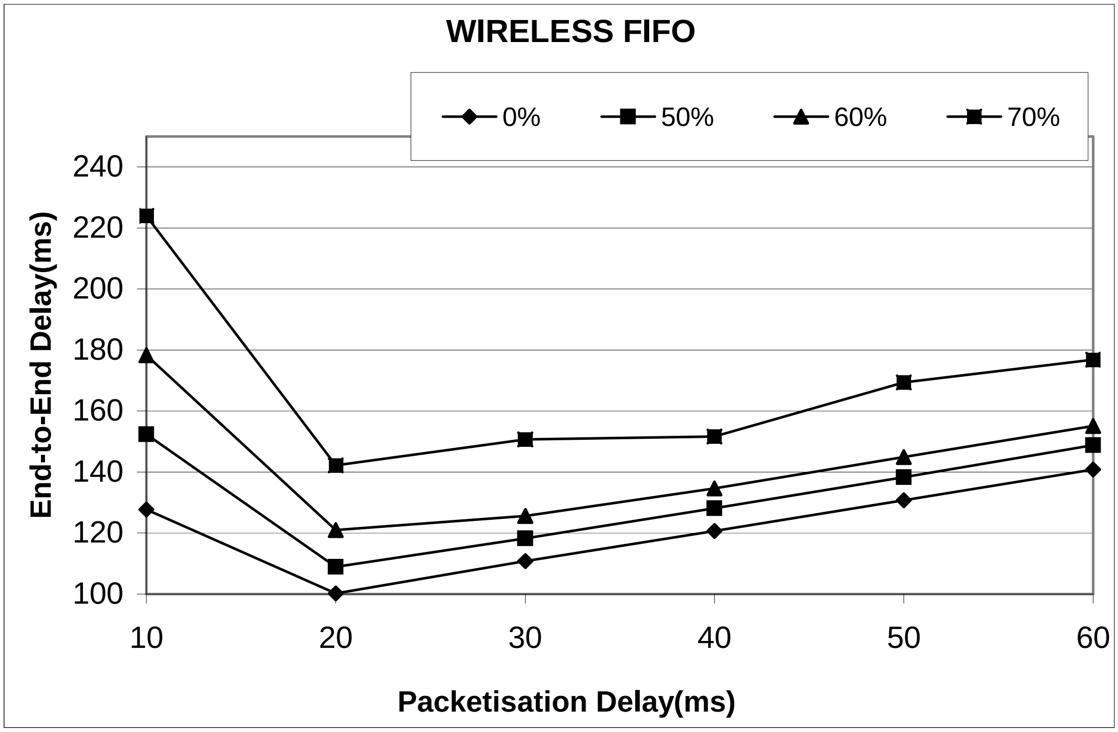


Figure 3-12 Delay Characteristic for FIFO over a wireless subnetwork

Figure 3-12 shows the delay characteristic for FIFO over a wireless subnetwork. Unlike WFQ the variation in load is significant with increase in load with FIFO queuing on the wireless network. These sets of curves are statistically different. At a load of more than 70%, a 10 ms packetization delay causes network congestion. When WFQ is used a step-by-step increase in delay is noted as the load on the network increases. Network condition remains stable. Wireless FIFO and wired FIFO exhibit similar delay characteristics. In the case of wireless FIFO the delay is more pronounced. An average positive shift 7 ms is noted. On comparing the packet loss for FIFO queuing to WFQ the difference is not statistically significant. The t-test is used to compare their distributions.

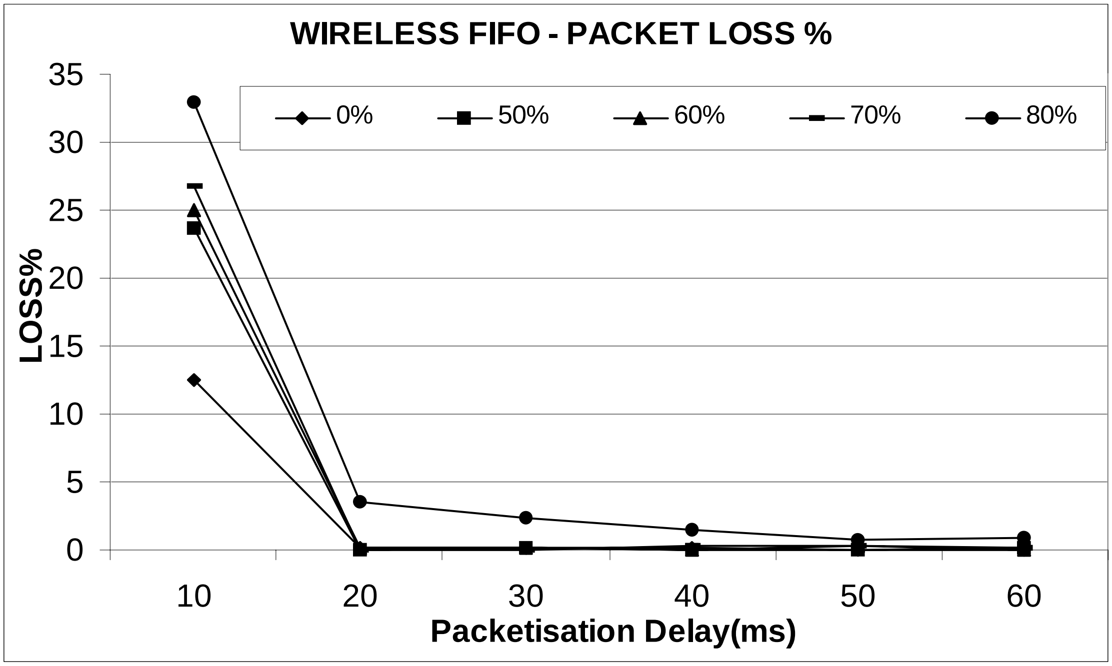


Figure 3-13 Packet Loss percentage for FIFO over a wireless subnetwork

Figure 3-13 shows the packet loss for FIFO over a wireless subnetwork. Even at 0% load the packet loss at 10 ms packetization delay does not meet the ITU specifications. Also at 80% load for packetization delay from 10 to 40 ms the packet loss percentage does not confer to the ITU specifications.

3.3 Changes in Delay Distribution

Since speech is sent through a packet-switched network, different packets of the stream experience variable source-to-destination delays. The packet delay variations have a substantial effect on the quality of a voice over IP call. To preserve continuity of the stream ideally this delay difference should be smoothed. Jitter buffers are devices implemented in terminal equipment or software at the receiving end of a voice connection on a packet switched network. Their function is to ensure that voice packets are delivered at regular intervals to the voice decoder, even if packets experience variation in time travelling over the network. Jitter buffer size [byte], is the physical size of the dejittering buffer at the receiver. Setting the size of a jitter buffer (measured in time, i.e. milliseconds) is crucial as the allocated buffer size should be utilized efficiently. If the jitter buffer is too large, unnecessary delay is introduced. If it too small, extra loss is introduced which affects voice quality. Many VoIP systems employ a fixed or adaptive jitter buffer to mitigate the effects. The size of a static jitter buffer is configured once, or fixed in hardware by the manufacturer. The size of a dynamic jitter buffer is adapted constantly by software in the receiving end to accomodate for changes in the network delay. If a frame incurs a particularly long end-to-end delay, the frame may not be ready to be played at the time when the preceding frame has been played. This will result in a gap in the playout. For continuous playouts every frame must be played with a fixed latency that is greater than the worst case end-to-end delay. Low latency with some gaps in playout will be preferable to high latency and no gaps. To reduce the gap in playout a playout buffer is used which buffers all the incoming packets at the receiver for a duration of time before playing out. This introduces extra buffer delay called the display latency. Two known jitter management policies are the I-Policy and the E-Policy. Under the I-Policy [15], frames are played with a single fixed latency and each frame that arrives with an end-to-end delay greater than this latency is discarded. Under the E-policy [15], late frames are played at the next opportunity. Due to this all frames after the late frame are delayed.

Destination buffering schemes help smoothing of the packet delay variability. The source-to-destination delay is a parameter used for the analysis of these schemes [15]. The distribution function is required for the analysis. For example in [15] the exponential function is chosen. With an exponential distribution most packets would have little deviation from the minimum delay but occasionally one would experience high delay. An adaptive jitter buffer would discard the packet, recognize that delay had increased and increase the jitter buffer size. From the experimental results we see that the distribution changes for different values of packetization delay and offered load as shown in figures 3-14 to 3-25. It takes the form of a poisson or exponential distribution.

3.3.1 Delay Distribution for wired WFQ

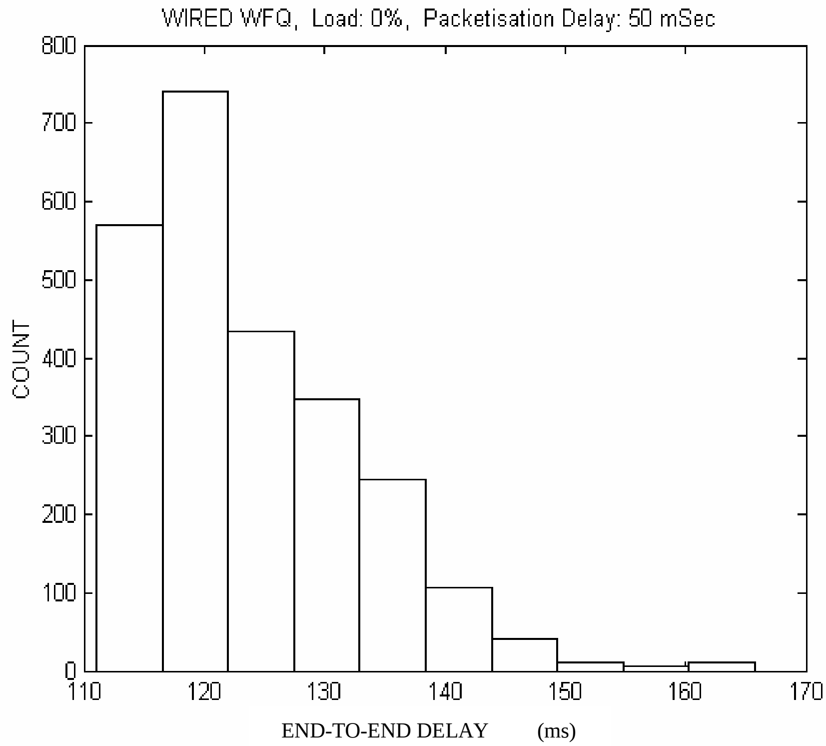


Figure 3-14 Delay Distribution for wired WFQ at 0% Load, 50ms

Figure 3-14 shows the delay distribution for a wired network using WFQ when voice traffic was generated using 50 ms packet delays and a background traffic load of 0% was employed. Visually analyzing the resulting curve we establish a poisson distribution, but again this is expected because the background traffic was generated with a poisson arrival rate. The distribution has an average value of 122 ms.

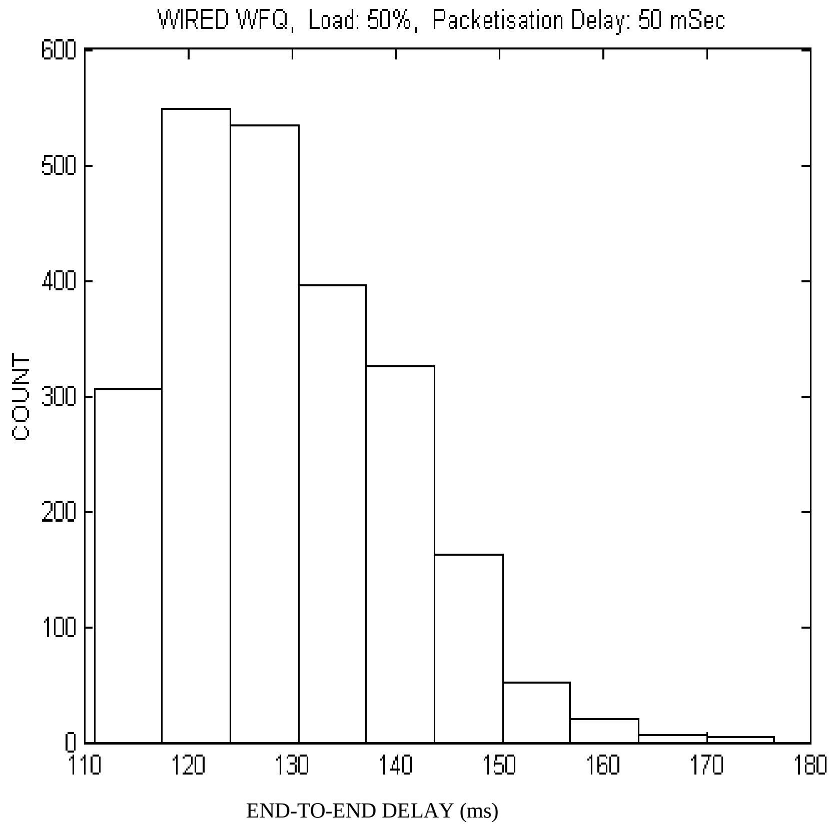


Figure 3-15 Delay Distribution for wired WFQ at 50% Load, 50ms

Figure 3-15 shows the delay distribution for a wired network using WFQ when voice traffic was generated using 50 ms packet delays and a background traffic load of 50% was employed. The resulting curve resembles a poisson distribution, but again this is expected because the background traffic was generated with a poisson arrival rate. Comparing this curve with the case for 0% load (Figure 3-14) the entire distribution shifts to the right. The delay distribution has an average value of 129 ms.

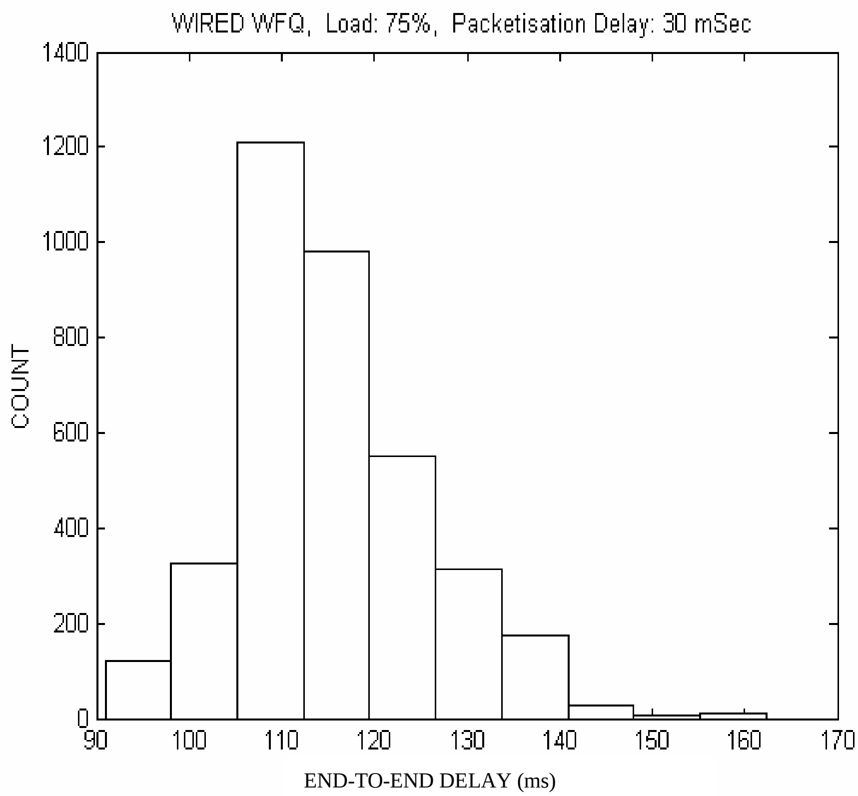


Figure 3-16 Delay Distribution for wired WFQ at 75% Load, 30ms

Figure 3-16 shows the delay distribution for a wired network using WFQ when voice traffic was generated using 30 ms packet delays and a background traffic load of 75% was employed. The resulting curve looks like a poisson distribution, but again this is expected because the background traffic was generated with a poisson arrival rate. The distribution has an average value of 115 ms. It is seen that the average value of delay is lesser than the value at 0% load (Figure 3-14). This is because a packetization delay of 30 ms is used compared to 50 ms as in Figure 3-14. In Figure 3-14 and Figure 3-15 the curves are shifted by their corresponding packetization delay. Also some of the packets are delayed less (90 ms – 100 ms) due to the time taken for the queue to build up initially.

3.3.2 Delay Distribution for wireless WFQ

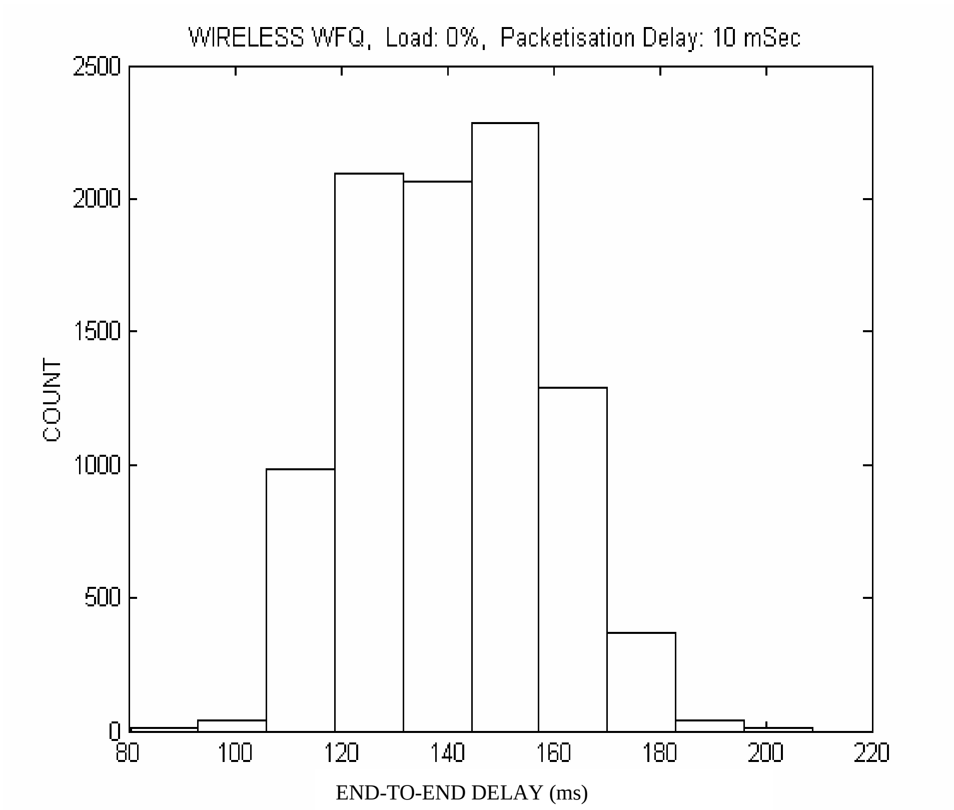


Figure 3-17 Delay Distribution for wireless WFQ at 0% Load, 10ms

Figure 3-17 shows the delay distribution for a wireless network using WFQ when voice traffic was generated using 10 ms packet delays and a background traffic load of 0% was employed. The resulting curve is no longer poisson shaped. It appears gaussian shaped. The delay distribution has an average value of 141 ms. Additive effect of the wireless environment results in significant delay. Since the packetization delay value of 10 ms is extremely small the end-to-end delay is considerable for the majority of packets.

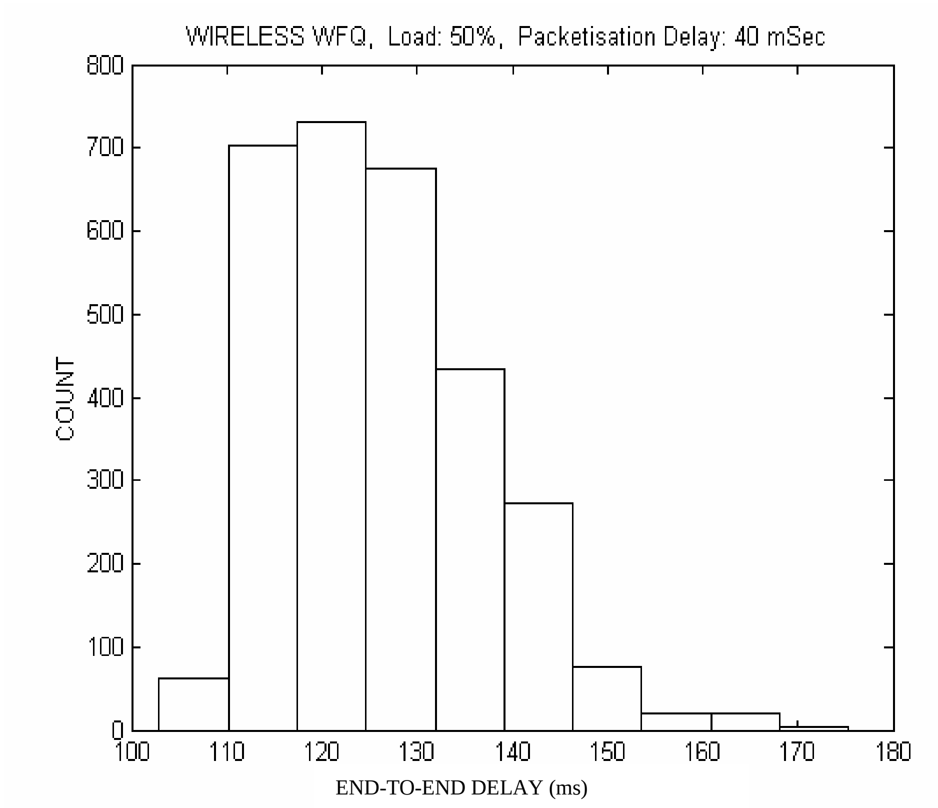


Figure 3-18 Delay Distribution for wireless WFQ at 50% Load, 40ms

Figure 3-18 shows the delay distribution for a wireless network using WFQ when voice traffic was generated using 40 ms packet delays and a background traffic load of 50% was employed. The distribution has an average value of 126 ms. On increasing the load to 50% the delay distribution shifts to the right. On comparing Figure 3-18 to the delay distribution for wired WFQ (Figure 3-15) we see the added delay of 10 ms and the delay due to the wireless medium.

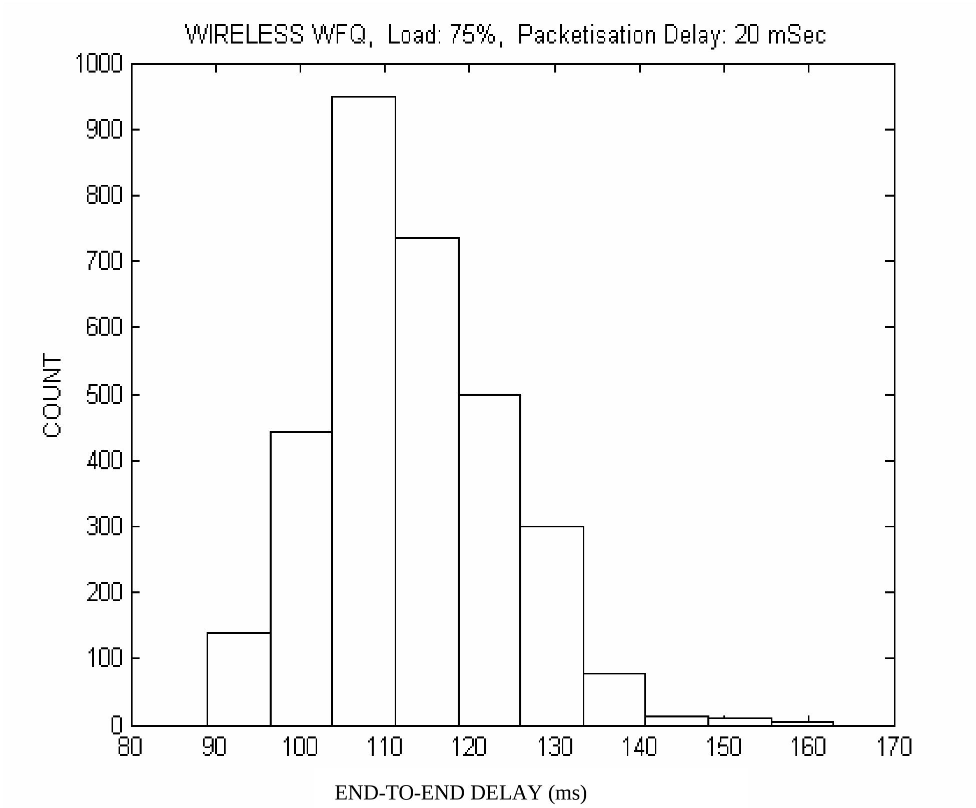


Figure 3-19 Delay Distribution for wireless WFQ at 75% Load, 20ms

Figure 3-19 shows the delay distribution for a wireless network using WFQ when voice traffic was generated using 20 ms packet delays and a background traffic load of 75% was employed. The resulting curve is no longer poisson shaped. It is appears gaussian shaped. The distribution has an average value of 113 ms. On comparing Figure 3-19 to the delay distribution for wired WFQ we see the added delay of 10 ms and the delay due to the wireless medium. The time taken for the queue to build up for high load results in small value of delay (90 ms – 100 ms) initially.

3.3.3 Delay Distribution for wired FIFO

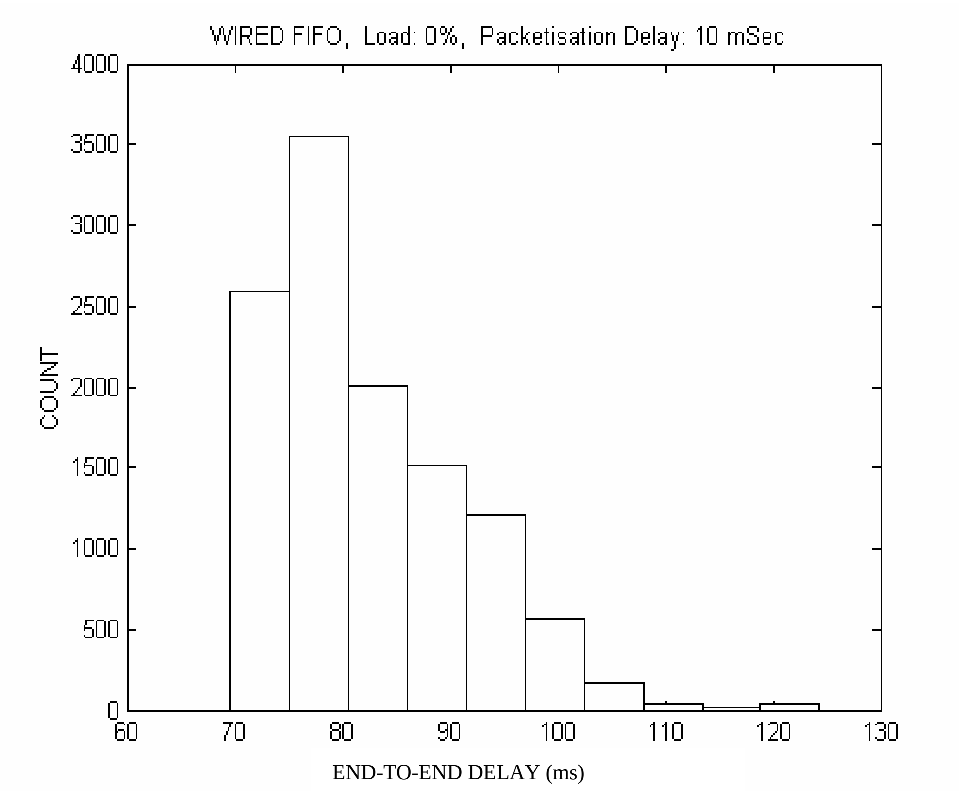


Figure 3-20 Delay Distribution for wired FIFO at 0% Load, 10ms

Figure 3-20 shows the delay distribution for a wired network using FIFO when voice traffic was generated using 10 ms packet delays and a background traffic load of 0% was employed. The resulting curve resembles a poisson distribution, but again this is expected because the background traffic was generated with a poisson arrival rate. The distribution has an average value of 81 ms. Compared to wired WFQ as shown in Figure 3-14 at 0% load for a packetization delay of 10 ms the delay distribution is left shifted.

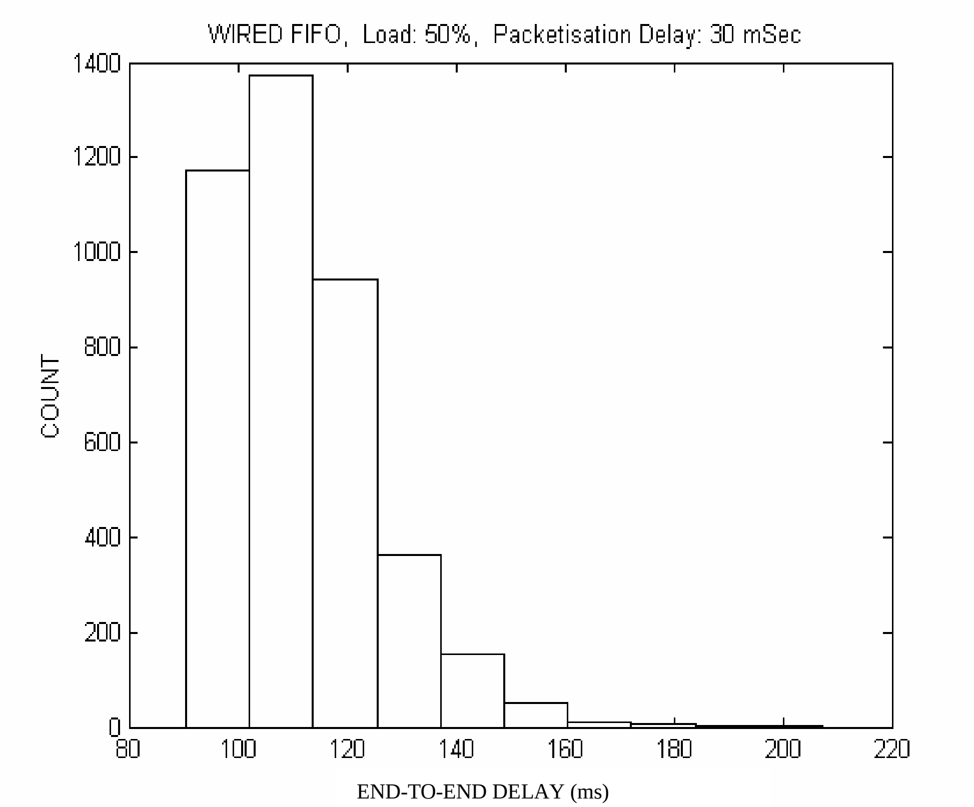


Figure 3-21 Delay Distribution for wired FIFO at 50% Load, 30ms

Figure 3-21 shows the delay distribution for a wired network using FIFO when voice traffic was generated using 30 ms packet delays and a background traffic load of 50% was employed. The resulting curve looks like a poisson distribution, but again this is expected because the background traffic was generated with a poisson arrival rate. The distribution has an average value of 111 ms. Compared to the distribution at 0% load (Figure 3-20) we see that the curve has right shifted.

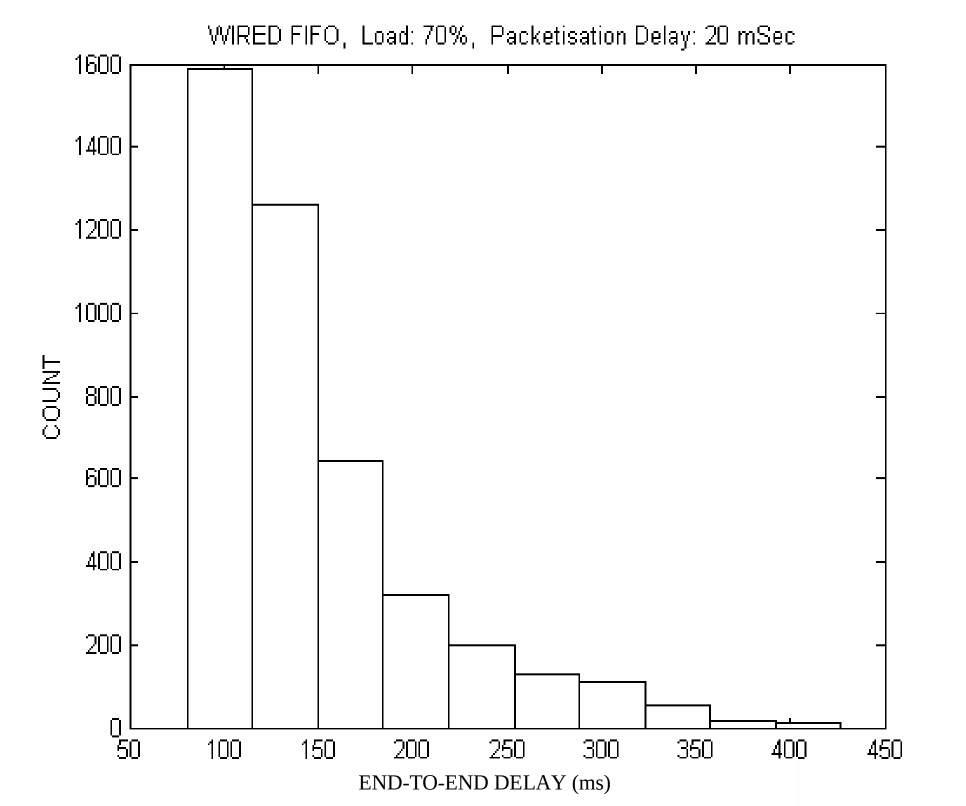


Figure 3-22 Delay Distribution for wired FIFO at 70% Load, 20ms

Figure 3-22 shows the delay distribution for a wired network using FIFO when voice traffic was generated using 20 ms packet delays and a background traffic load of 70% was employed. The distribution has an average value of 147 ms and has an exponential shape when compared to lower loads where the distributions were poisson. When compared to Figure 3-21 we see that as the load increases the number of packets suffering delay exceeding 200ms increases. While the maximum delay experienced by the packets is 180 ms for 50% load, at 70 % load the delay goes upto 350 ms.

3.3.4 Delay Distribution for wireless FIFO

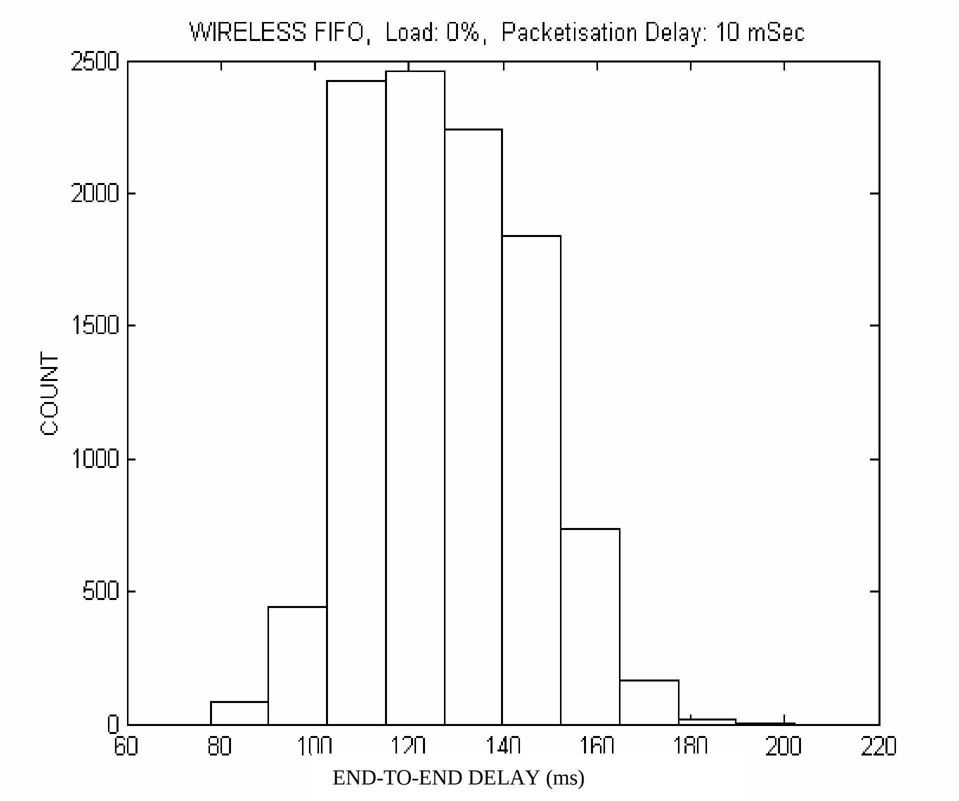


Figure 3-23 Delay Distribution for wireless FIFO at 0% Load, 10ms

Figure 3-23 shows the delay distribution for a wireless network using FIFO when voice traffic was generated using 10 ms packet delays and a background traffic load of 0% was employed. The distribution has an average value of 128 ms and shows an approximate gaussian shape. The packet delay is more in the case of wireless WFQ as in Figure 3-17.

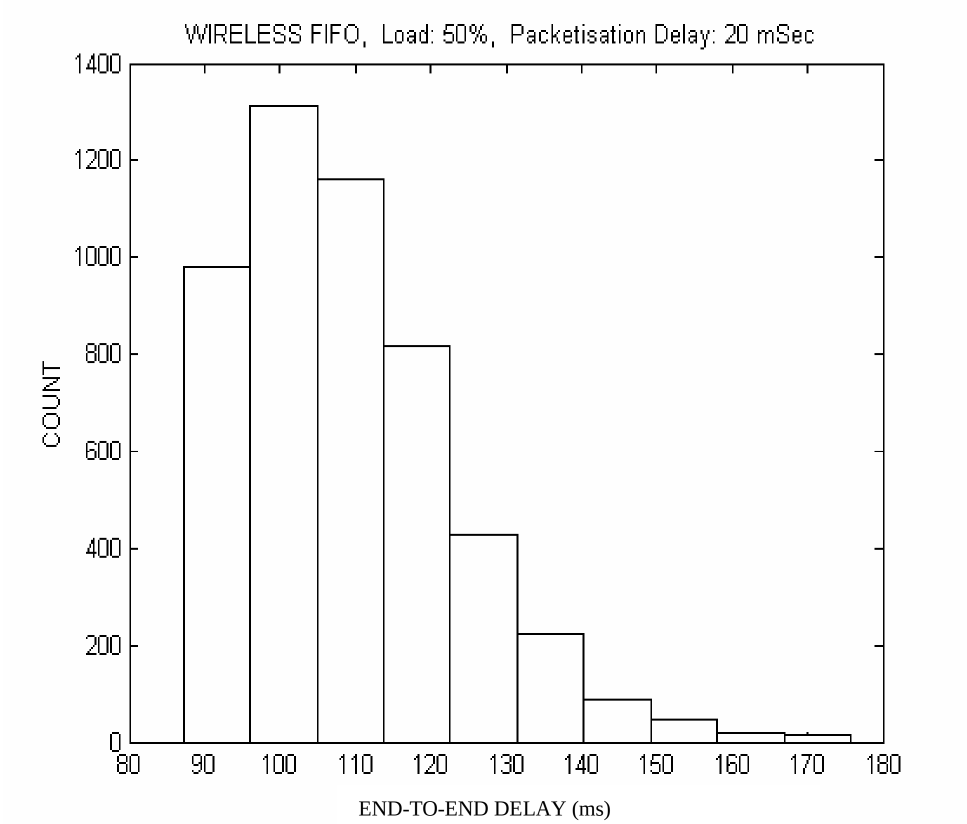


Figure 3-24 Delay Distribution for wireless FIFO at 50% Load, 20ms

Figure 3-24 shows the delay distribution for a wireless network using FIFO when voice traffic was generated using 20 ms packet delays and a background traffic load of 50% was employed. The resulting curve resembles a poisson distribution, but again this is expected because the background traffic was generated with a poisson arrival rate. The distribution has an average value of 109 ms. With respect to Figure 3-21 showing the distribution of wired FIFO there is a right shift.

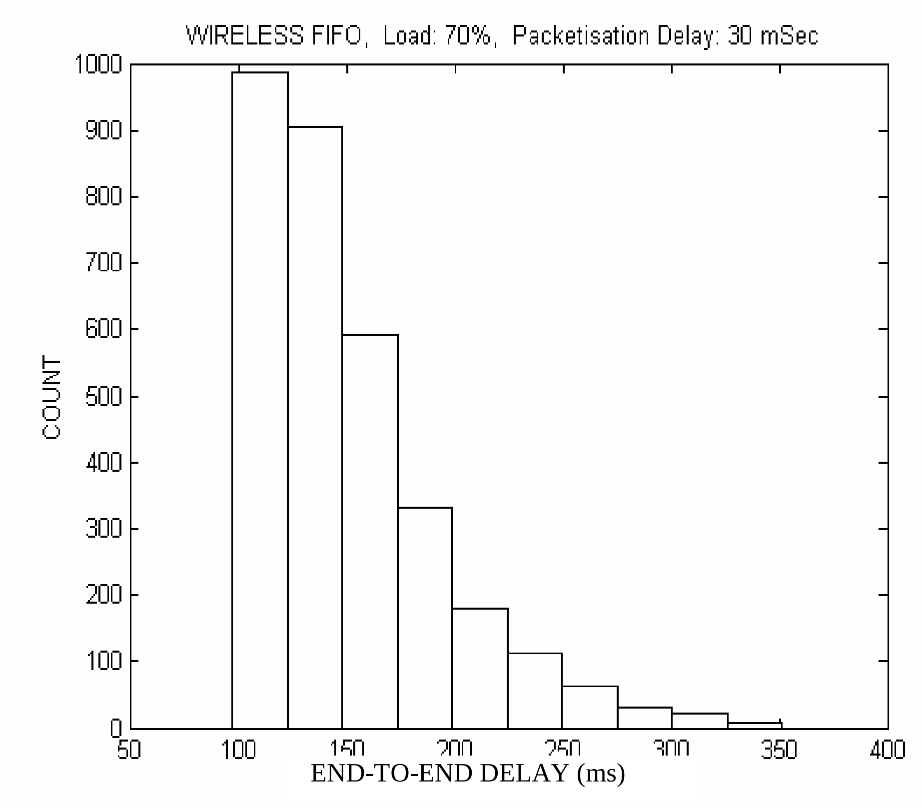


Figure 3-25 Delay Distribution for wireless FIFO at 70% Load, 30ms

Figure 3-25 shows the delay distribution for a wireless network using FIFO when voice traffic was generated using 30 ms packet delays and a background traffic load of 70% was employed. The distribution has an average value of 151 ms and has an approximate exponential shape. Comparing the delay distribution of Figure 3-22 the distribution is slightly right shifted.

Chapter 4

ANALYTICAL RESULTS

4.1 System Model

Voice packets generated are sent over a network prototype best representing a real -time Internet service. The overall end-to-end delay is broken into three significant delay components namely packetization delay, inherent Internet delay and network delay. The variation in packetization delay affects end-to-end delay as demonstrated in this study. At certain delay values considerable savings in the delay budget is achieved. The packetization delay introduces a fixed delay and is controllable. The model assumes an inherent Internet delay of 70 milliseconds with a standard deviation of 10 milliseconds [13]. The network delay is highly variable. Constant Bit Rate traffic is assumed for simplicity. As the network congestion experienced in each direction is different unidirectional traffic flow is taken into account. Analysis is done with the G.729 codec. The background data traffic generated simulates varying load conditions as in a typical Internet network. It follows a Poisson distribution. A link capacity of 250 Kbps is set across the WAN link. This acts as the weakest link.

4.2 Effect of Queuing strategy on End-to-End delay

With the FIFO queuing strategy, a M/M/1 queuing system model is used [14]. M/M/1 is Kendall's notation of this queuing model. The first part represents the input process, the second the service distribution, and the third the number of servers. The M represents an exponentially distributed interarrival or service time, specifically M is an abbreviation for Markovian. The M/M/1 Waiting line system has a single channel, single phase, Poisson arrival rate, exponential service time, unlimited population, and First-in First-out queue discipline. When the traffic intensity increases the amount of queuing increases. The delay of the packets being processed increases as the queue size increase. This is due to the fact that as more and more packets wait in the queue, the delay for the packets increases in time. When using small packet sizes the packetization delay is very small but the queuing delay leads to high end-to-end delay. The queue becomes full and packets are dropped. To analyze this type of queue, we need to know only the mean arrival rate λ and the mean service rate μ. The utilization of the processor is given by the probability of having one or more packets in the system. The average utilization $\rho = \lambda/\mu$. The mean response time is given as $1/(\mu-\lambda)$. The p^{th} percentile of the waiting time is given as $-W \ln((1-p)/\rho)$.

4.3 Analytical Results

The average increase in the end-to-end delay on addition of a wireless hop is found for voice traffic at zero load condition. Figure 4-1 and Figure 4-2 shows this difference. At a packetization delay value of 10 milliseconds there is a huge increase in the end-to-end delay. The variance of the shift is shown in Figure 4-3 and Figure 4-4. At zero background load conditions both the queuing strategies yield comparable results. Figure 4-5 and figure 4-6 show the shift in delay at different load conditions for WFQ and FIFO queuing. These values are obtained as the difference between the average values for WFQ and FIFO given in Section 3.

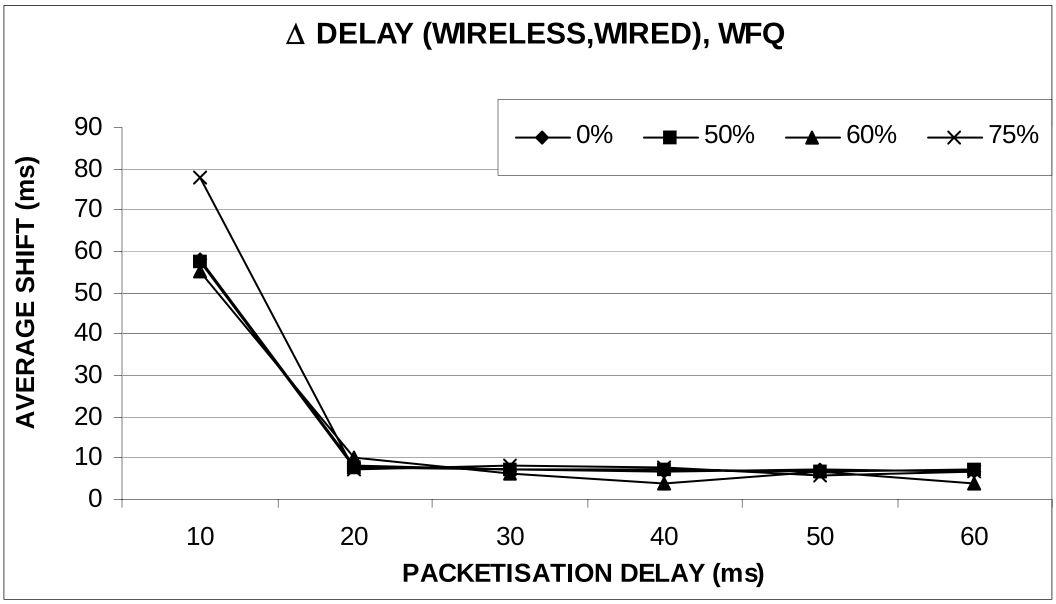


Figure 4-1 Average increase in delay on addition of a wireless hop with WFQ at 0%

Figure 4-1 shows the positive shift in the delay on addition of a wireless hop to the wired network with WFQ enabled. For packetization delay values ranging from 20 ms to 60 ms the shift lies between 6.5 ms and 8 ms. The 10 percentile and 90 percentile values show the delay dispersion. Figure 4-1 shows the value of packetization delay does not seem to significantly impact the positive shift got on insertion of a wireless link.

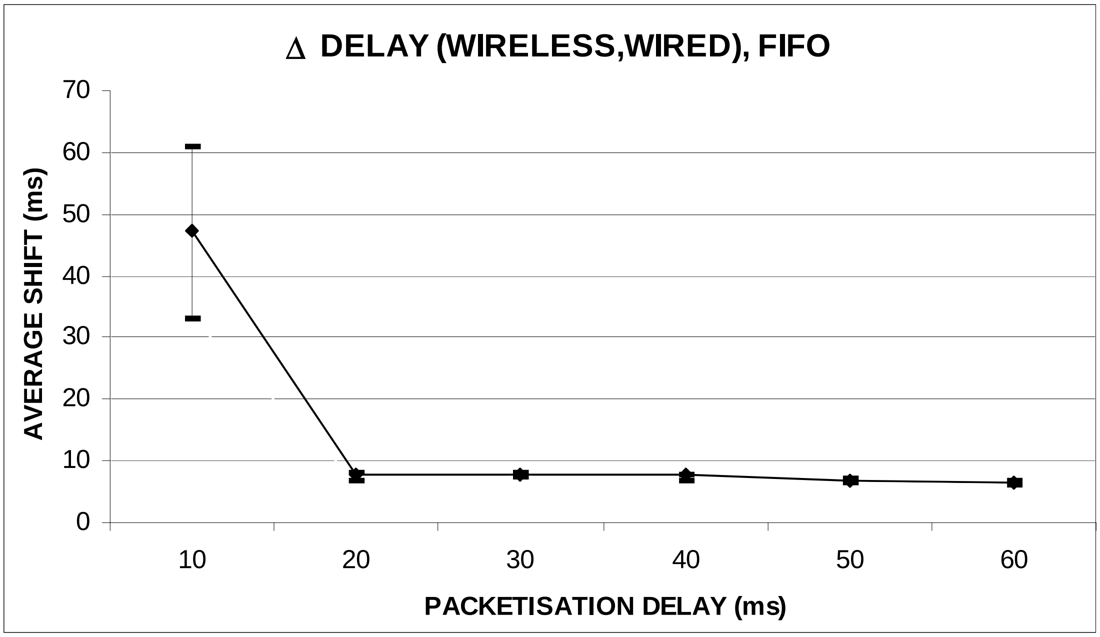


Figure 4-2 Average increase in delay on addition of a wireless hop with FIFO at 0% Load

Figure 4-2 shows the positive shift in the delay on addition of a wireless hop to the wired network with FIFO enabled. For packetization delay values from 20 ms to 60 ms the shift is between 6.5 ms and 8 ms. Shift in delay is observed at 0% load. The difference got is in the same range as WFQ as shown in Figure 4-1. As there is no background traffic, voice packets traverse the network without a bottleneck. Hence the type of queuing does not produce a significant difference.

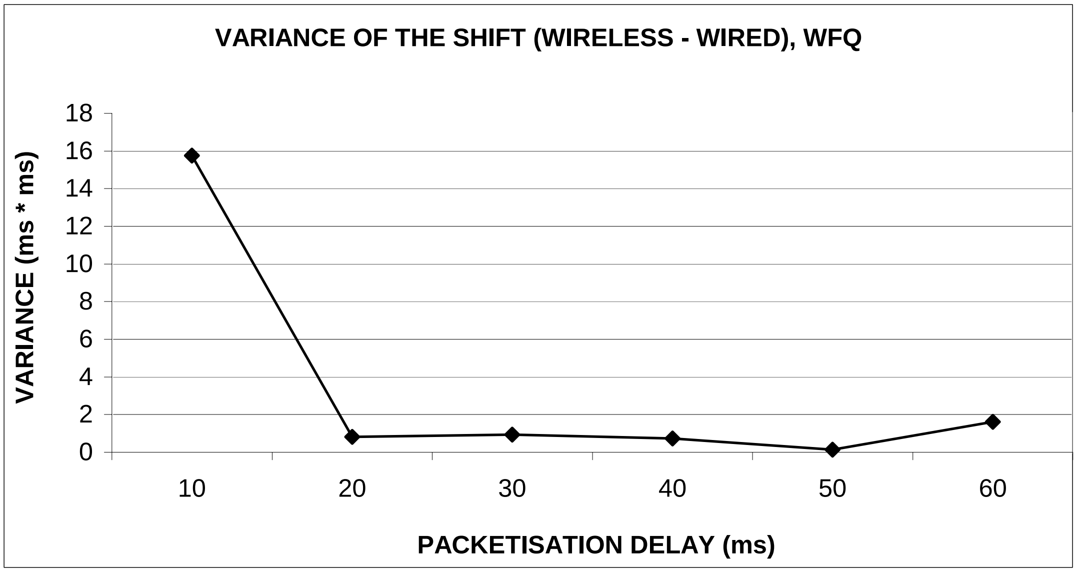


Figure 4-3 Plot of the variance of the shift between a wired and wireless network

Figure 4-3 shows the variance of the shift between a wired and wireless network using WFQ. At a packetization delay of 10 ms the variance is extremely high. This is because the packets are extremely small. The packet loss is very high and the packets undergo highly varying delays.

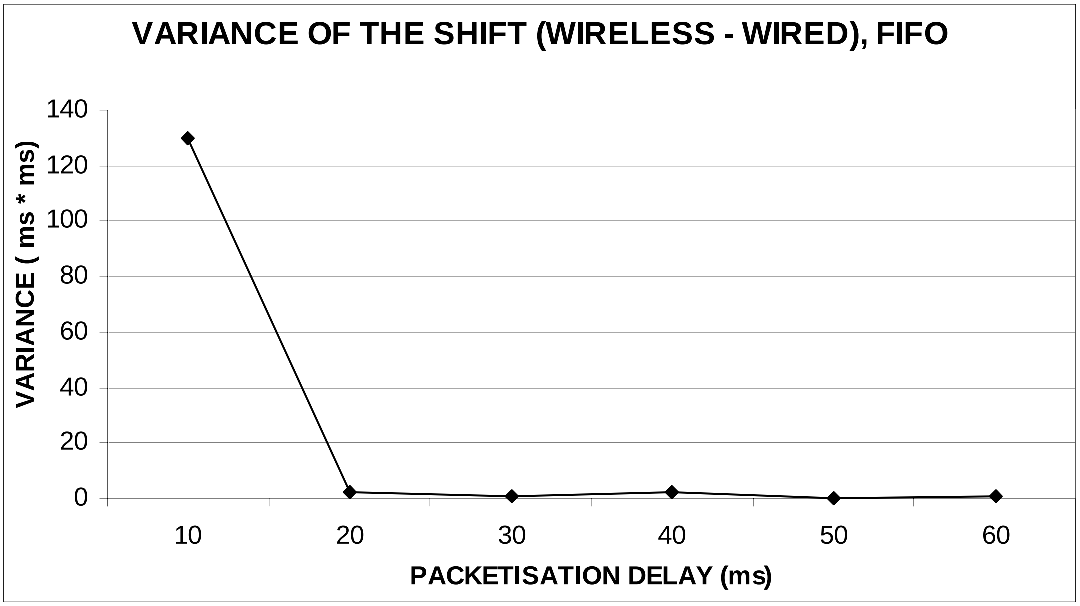


Figure 4-4 P1ot of the variance of the increase in delay between a wired and wireless network

Figure 4-4 shows the variance of the shift between a wired and wireless network using FIFO. Again at a packetization delay of 10 ms the variance is extremely high. The variance for FIFO is more than the variance for WFQ as shown in Figure 4-3

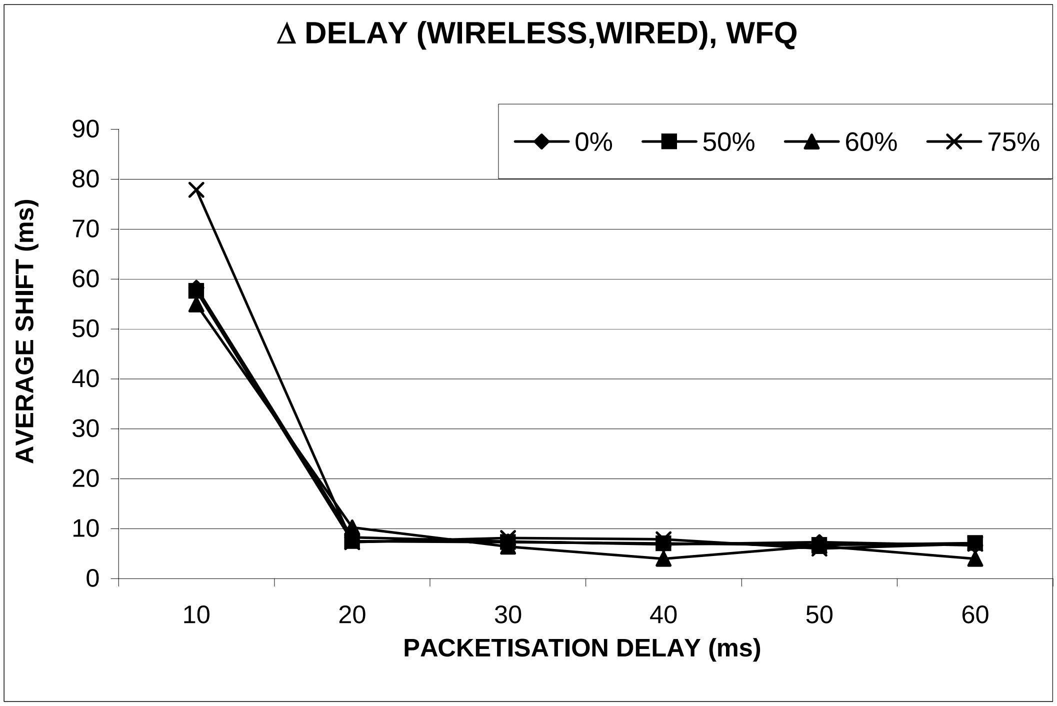


Figure 4-5 Average increase in de1ay on addition of a wireless hop with WFQ

Figure 4-5 shows the added delay when a wireless hop is added to the existing network is pronounced when the packetization delay is 10 ms. Small value of packetization delay contributes to a low payload-to-overhead ratio. When transmitted in a wireless environment it induces a high delay. For packetization delay values from 20 ms to 60 ms the positive shift is less than 10 ms.

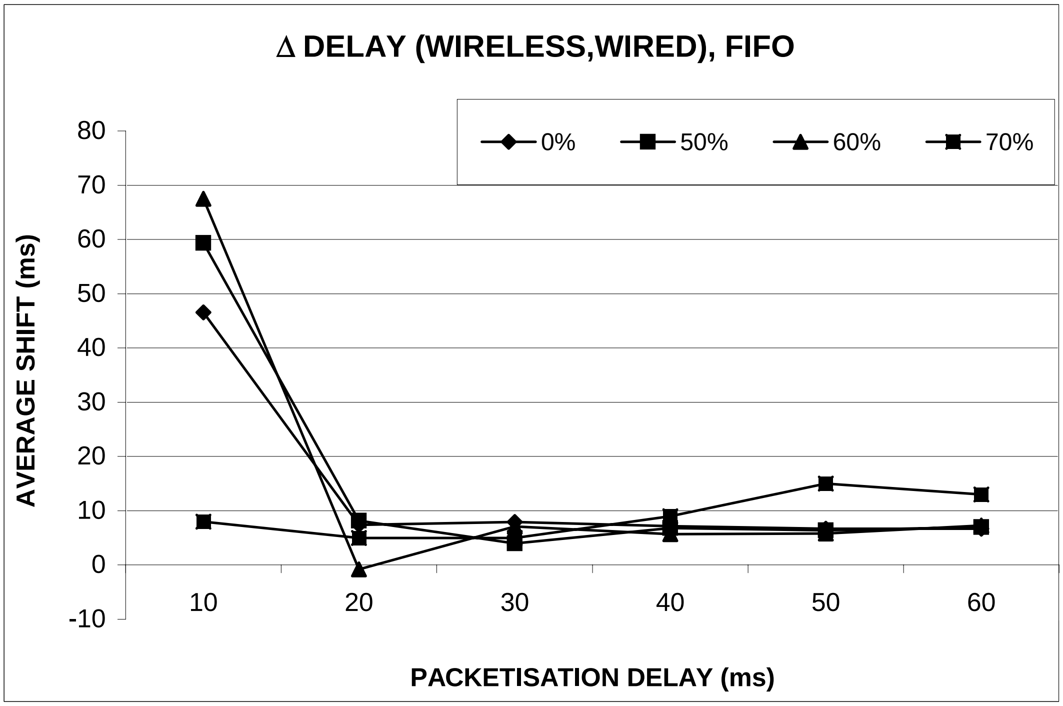


Figure 4-6 Average increase in de1ay on addition of a wireless hop with FIFO

Figure 4-6 shows the increase in delay when a wireless hop is added to the existing network is pronounced at a packetization delay is 10 ms. Dip in the curve seen at 20 ms is significant. This might be because at 10 ms the number of packets generated is very high. Hence the buffer in the access point could be filled up and the packets experience delay. At 20 ms the scenario changes since lesser number of packets are generated. At a load of 70% the difference in delay between wired and wireless FIFO at 10 ms is not very significant. For packetization delay values between 20 ms and 60 ms the difference n delay is less than 15 ms.

Chapter 5

CONCLUSIONS AND COMMENTS

WLANs offer advantages in terms of mobility, connectivity and ease of installation. They provide long-term cost savings in dynamic environments. At the same time wireless LANs don't provide a complete solution to networking. They adequately supplement wired LANs. Devices operate only over a limited distance and security is difficult to guarantee. As the number of computers using the network increases, the data transfer rate to each computer accordingly decreases. The transmission speeds for wireless LANs depends on parameters such as the distance from the access point, the local environment, the file size and the number of users. However the glaring benefits provided by WLANs overshadow their shortcomings. Today we can see many networks with a wireless extension to the wired LAN backbone. Separate networks may be linked via a wireless connection.

Voice over IP is still an inexpensive alternative to traditional telephone service. Deployment of VoIP in wireless local area networks has more stringent requirements. Delay and packet loss values have to be within the maximum acceptable vales. Since we are considering a public network, minimization of end-to-end delay may be done only at the sending and receiving nodes. Minimization in the remaining part of the network is not viable. For better results, techniques such as loss concealment and voice activation detection can be used. Emphasis may be placed on enhancement of speech coding and sophistication of jitter buffers used. In this thesis the packet size at the end nodes is considered. Transmitting voice packets at a particular size yields minimum end-to-end delay in a network with a wireless hop. Considerable savings in the delay budget can be achieved. These savings tend to be critical for the time sensitive voice packets.

In the analytical part the effect of introducing a wireless extension to the existing wired network is shown. The effect is additive leading to a positive shift in the end-to-end delay. The queuing technique enabled on the router affects the delay characteristic. This relationship is discussed in light of the end-to-end delay introduced in the network. The experimental analysis concludes that there is a correspondence in the delay characteristics of a fully wired network and one having a wireless extension. The G.729 codec, the most practical choices for voice over IP is used.

The wireless extension adds to the end-to-end delay. The transmit power is set at 100 mW (20 dBm). As the SNR decreases the packet loss increases. The SNR has little effect on the end-to-end delay compared to its effect on packet loss. In a wireless environment there are more packet losses. The packets are being dropped in the queue of the buffer in the access point. There are more packet losses. There is a striking difference between the two networks when the packetization delay is 10 milliseconds. Adding a wireless link adds a large delay when 10 ms voice packets are used. Otherwise it adds a fairly constant 7 ms delay across all of the packetization delays and traffic loads if WFQ is used. If FIFO is used than again a large delay is added at 10 ms and the delay for 20-40 ms is again 7ms or more and at high loads (70%) the delay increases to 15 ms. Overhead of the header leads to very low efficiency when 10 ms voice packets are used. This packet size is highly infeasible for VoIP in a wireless environment. The problem is more intense at high network loads for FIFO enabled queues. When the size of the compressed data frame is 20 bytes, corresponding to 20 milliseconds of speech samples, the end-to-end delay across the network is minimum. With respect to this value there is a savings in the delay budget of 30% compared to a data frame size of 10 bytes and 25% compared to a data frame size of 60 bytes.

It is observed that the packet loss with Weighted Fair Queuing is negligible. When tested over FIFO queuing the packet loss exceeds the acceptable cutoff for voice transmission at a data frame size of 10 bytes. The packetization delay value which produces the minimum delay when using IP telephony over a WFQ enabled network having a wireless subnet where the delay is 114 ms ± 13 ms for a load of 80% comes out to be 20 milliseconds. FIFO queuing touches 220 milliseconds for 70% load while it does not go beyond 172 milliseconds for a load a high as 80% with WFQ. When FIFO is enabled over the network with a wireless subnet packet telephony is barely feasible beyond three-fourths the link capacity. At high loads the end-to-end delay is impractical. A small increase in the variance occurs, in WFQ it ranges from 0.8 – 1.6 $msec^2$ and for FIFO it ranges from 0.1–1.9 $msec^2$ for packetization delay values between 20 and 60 ms. It is interesting to note the variation in delay distribution of voice packets for different packet sizes and load conditions. This data could be used in designing adaptive jitter buffers. The shape of the distribution changes, even if the background traffic is generated from a poisson process that does not necessarily mean that the delay will be a poisson distribution.

Security is an important issue in wireless networks. As future work the delay characteristics for a wireless network with encryption enabled may be obtained. The voice quality with different encryption schemes could be considered. Several standards are developed under the 802.11 umbrella. When there is mobility of clients between access points or when additional traffic is introduced on the wireless link the scenario may change significantly. It would be interesting to do a relative study for optimization for voice transfer over IP. VoIPv6 has a bigger packet size in comparison with VoIPv4. The tolerance of VoIPv6 can be tested in a wireless environment.

APPENDIX

IP Address of Cisco Routers

3600 Series Cisco Routers
RC1: 192.168.117.52

2500 Series Cisco Routers
RW1:

Ethernet Interface: 30.0.0.11

Serial Interface: 31.0.0.50

2500 Series Cisco Routers
RW2:

Serial Interface: 31.0.0.51

Ethernet Interface: 32.0.0.50

3600 Series Cisco Routers
RC2: 32.0.0.51

RC1

interface Ethernet0/0

ip address 192.168.117.52 255.255.255.0

dial-peer voice 2 voip

destination-pattern 301

codec g729ar8 bytes 10

session target ipv4:32.0.0.51

exit

exit

wr

router rip

network 192.168.117.0

ip route 30.0.0.0 255.0.0.0 192.168.117.50

ip route 31.0.0.0 255.0.0.0 192.168.117.50

ip route 32.0.0.0 255.0.0.0 192.168.117.50

RC2

```
interface Ethernet0/0

ip address 32.0.0.51 255.0.0.0


dial-peer voice 2 voip

destination-pattern 51

codec g729ar8 bytes 10

session target ipv4:192.168.117.52

exit

exit

wr


router rip

network 32.0.0.0


ip route 30.0.0.0 255.0.0.0 32.0.0.50

ip route 31.0.0.0 255.0.0.0 32.0.0.50

ip route 192.168.117.0 255.255.255.0 32.0.0.50
```

RW1

interface Ethernet0

ip address 30.0.0.11 255.0.0.0

interface Serial1

ip address 31.0.0.50 255.0.0.0

clockrate 250000

router rip

network 31.0.0.0

network 30.0.0.0

ip route 32.0.0.0 255.0.0.0 32.0.0.50

ip route 192.168.117.0 255.255.255.0 30.0.0.50

RW2

```
interface Serial0

ip address 31.0.0.51 255.0.0.0


interface Ethernet0

ip address 32.0.0.50 255.0.0.0


router rip

network 32.0.0.0

network 31.0.0.0


ip route 30.0.0.0 255.0.0.0 31.0.0.50

ip route 192.168.117.0 255.255.255.0 31.0.0.50
```

BIBLIOGRAPHY

[1] Todd Lammle, “CCNA Cisco Certified Network Associate Study Guide”, SYBEX Inc. 2002.

[2] William Stallings, “Data And Computer Communications”, Prentice-Hall, Inc.1997.

[3] William A. Shay, "Understanding Data Communications & Networks", Brooks/Cole, 1999.

[4] B. Vlaovic, Z. Brezocnik, "Packet Based Telephony", EUROCON'2001, Trends in Communications, Volume: 1, 2001.

[5] David J. Wright, "Voice over Packet Networks", John Wiley & Sons, 2001.

[6] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RFC 1889: RTP: A Transport Protocol for Real-Time Applications", January 1996.

[7] "Voice over IP for Cisco 3600 Series Software Configuration Guide", Cisco Systems, Inc. 1992--2002.

[8] Bob O'Hara & Al Petrick, "The IEEE 802.11 Handbook: A Designer's Companion", IEEE, 2000.

[9] M. Carson, National Institute of Standards and Technology, http://www.itl.nist.gov/div892/itg/carson/nistnet/.

[10] B. Adamson, The Naval Research Laboratory, http://manimac.itd.nrl.navy.mil/MGEN/.

[11] http://www.ethereal.com

[12] http://www.cisco.com

[13] Boonchai Ngamwongwattana "Optimizing Packetization for Minimal End-to-End delay in VoIP network" Masters Thesis, University of Pittsburgh 2001

[14] Raj Jain, “The Art of Computer Systems Performance Analysis”, John Wiley & Sons,1991.

[15] William E.Naylor, Leonard Kleinrock, "Stream Traffic Communication in Packet Switched Networks: Destination Buffering Considerations", IEEE Trans. Commun., vol. COM-30, Dec. 1982.

[16] Donald L.Stone, Kevin Jeffay, "An Empirical Study of Delay Jitter Managment Policies", Multimedia Systems, vol. 2, July 1994.

[17] Klepec, B., Kos, A., "Performance of VoIP Applications in a Simple Differentiated Services Network Architecture", EUROCON'2001, Trends in Communications, vol.1, 2001.

[18] Lakaniemi, A., Rosti, J., Raisanen, V.I., "Subjective VoIP speech quality evaluation based on network measurements", Communications, 2001. ICC 2001. IEEE International Conference, vol.3, 2001.

[19] Postel, J., "Internet Protocol", RFC 760, USC/Information Sciences Institute, January 1980.

[20] Postel, J., "Transmission Control Protocol," RFC 761, USC/Information Sciences Institute, January 1980.

[21] David L. Mills, “Network Time Protocol (Version 3) Specification, Implementation and Analysis”, RFC 1305, University of Delaware, March 1992.

[22] Andrew S. Tanenbaum, “Computer Networks”, Prentice-Hall, Inc.2001.

[23] Jeffrey Feigin, Kaveh Pahlavan, Mika Ylianttila, “Hardware-Fitted Modeling and Simulation of VoIP over a Wireless LAN”, IEEE, 2000.