ATHABASCA UNIVERSITY

SURVEY OF VOIP-BASED TECHNOLOGY AND SUGGESTIONS FOR IMPROVEMENTS

BY

SCHUBEN DIAS

A thesis submitted in partial fulfillment

Of the requirements for the degree of

MASTER OF SCIENCE in INFORMATION SYSTEMS

Athabasca, Alberta

October 2008

ABSTRACT

VoIP over WLAN is poised to become an important Internet application. However, there are two major technical problems that stand in the way. These are low VoIP capacity in WLAN and unacceptable VoIP performance in the presence of coexisting traffic from other applications. In this paper we review literature [11] (Ganguly, 2006), containing ways of improving VoIP performance. One method of improving VoIP performance is giving higher priority to VoIP packets at the Access Point when there is a mixture of TCP and VoIP packets. Another method is multiplexing packets from several VoIP streams. This helps increase the amount of data sent by reducing the number of packets as there is an overhead associated with each packet. Reducing the number of packets would allow more data to be transmitted. The jitter buffer repeats old packets to prevent reduced quality due to delayed packets. Algorithms can be applied to reject packets which have been delayed too long to prevent the jitter buffer from growing too long at the access point. The sampling rate of the voice analog message can be adjusted to prevent too many packets from being transmitted and still keep the VoIP signal quality high. The number of hops (Comeras, 2007) is studied [6] to find the optimal number of hops without saturation of bandwidth. End-to-end model is simulated [7] with different parameters, like packet size, to find an optimal set of parameters. Interference to VoIP devices from blue-tooth devices is investigated [8]. When multiple devices send packets through a media, collisions occur. When a collision occurs, the devices have a back-off period.

Random back-off period is investigated [11] to increase capacity. When VoIP devices leave one Access Point area and enter another access point area, there is a handoff from one Access Point to another. A novel Access point handoff is discussed [10].

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of Purpose

The purpose of this research paper is to determine ways of improving VoIP capacity and performance.

Research Question or Purpose

Voice over Internet Protocol (VoIP) is becoming more important because of the widespread availability of WLAN networks and the low cost of making voice calls over the network. However, the VoIP capacity and performance in the presence of existing traffic are major technical problems.

A typical VoIP packet at the IP layer consists of a large header and a payload about the size of the header. One way of increasing capacity is to reduce the header or have the same header with a larger payload. This is done using aggregation of packets.

A method of reducing Access Point contention would be to give priority to high priority VoIP packets at the Access Point and to queue the remaining VoIP packets in an Access Point buffer. A scheme priority [1] is given at Access Point, which reduces or increases the contention window allowing some packets priority over others. Various data applications, which are not real-time, also exist on the network. The scheme priority can be used to give higher priority to the VoIP packets over the data packets. Multiplex-multicast (M-M) scheme is performed [2] by multiplexing packets from several VoIP streams into one multicast packet. This method increases the amount of data sent by reducing the number of packets and overhead sent.

Adaptive Jitter Buffering (AJB) and latency management can be implemented at the endpoints [4] to manage the jitter buffer. Breaks of a few milliseconds can be introduced at the end points to improve the quality of playback. Over time, the delay accumulates to cause the playback to be out of phase with the recording. This also, causes the jitter buffer to become very long and reduces quality. The AJB algorithm prevents the jitter buffer from growing too long.

VoIP transmission involves sampling the voice signal at a rate which influences the speech quality transmitted. A high rate of sampling of the voice signal transmits a high quality signal [5]. The amplitude of signal recorded also affects the quality of signal. Also, the jitter buffer at the receiver helps to synchronize packets that have been delayed. This study [5] finds an optimum sampling rate for a reasonable voice quality.

VoIP calls have to be routed through multiple hops. Disconnections trigger rerouting which may increase the number of hops. An investigation (Comeras,

2007) [6] determines the number of hops that can be made without saturating the bandwidth of the network.

Various devices operate in the same frequency range as wireless VoIP devices such as microwave ovens and Bluetooth devices. Bluetooth technology operates at the same frequency range as wireless devices. Wireless VoIP Speech quality in the presence of Bluetooth interference (McKay, 2003) [8] is investigated.

Capacity improvement of Wireless LAN VoIP [9] makes use of a random back-off period to prioritize the scheduling of requests in a Distributed system.

A Novel AP [10] is introduced which involves the use of the AP to handoff a mobile host to another AP with higher signal strength to the mobile host. When a mobile host moves into the range of another AP while connected to a different AP, the handoff time would involve searching for an AP and connecting to the AP. If the previously connected AP would negotiate the connection with its neighbour AP, it saves connection time.

The footprint of each Access Point (AP) in a WLAN is limited to about 250 meters outdoors and about 100 meters indoors. If we need to provide wireless connectivity to a large area without having a wired backplane, we need a wireless mesh network.

Mesh networks add routing functionality to APs of WLAN. Recently, (Ergen, 2006) [11], with the coming of dual interface cell phones with WiFi interfaces and ubiquitous availability of wireless LAN, VoIP over WLAN is gaining popularity.

Packet losses and delay due to interference in a multiple hop mesh network with limited capacity can significantly degrade the end-to-end VoIP call quality [11]. This study [11] presents and evaluates the practical optimizing techniques that can enhance the network capacity, maintain the VoIP quality and handle user mobility efficiently.

Maintaining calls without disruptions for mobile VoIP clients that handoff to different APs during an ongoing voice call is crucial. A design is provided which provides a fast handoff scheme that uses link layer feedback mechanism to detect client movement and update the routes within the mesh network accordingly. The voice packets at the old AP are buffered and rerouted via the new AP to the client.

CHAPTER II

REVIEW OF RELATED LITERATURE

Solutions to Performance Problems in VoIP over 802.11 Wireless LAN.

Voice over Internet Protocol (VoIP) over a wireless local area network (WLAN) is ready to become an important Internet application. At the same time, driven by huge demands for portable access, wireless local area network (WLAN) market is taking off quickly. Due to its convenience, mobility and high-speed access WLAN is becoming more popularly used for the “last mile” Internet access. VoIP exploits the advanced voice-compression techniques and bandwidth sharing in packet switched networks. VoIP can dramatically increase bandwidth efficiency. There are two major technical problems. These are: 1) low VoIP capacity in WLAN and 2) unacceptable VoIP performance in the presence of existing traffic. A VoIP stream with 10kb/sec, and 802.11b WLAN operating at 11Mb/s could in principle support more than 500 VoIP sessions. In real life only a few sessions can be supported due to protocol overheads. We can improve the VoIP capacity by close to 100% without changing the standard 802.11 CSMA/CA protocol [2]. This can be done by optimizing the DIFS and SIFS time.

VoIP can be used with novel applications that integrate voice with data. It facilitates the creation of new services that combine voice communication with other media and data applications such as video, white boarding, and file sharing. In order to do this, VoIP and TCP traffic has to coexist harmoniously over the WLAN. This can be largely solved by simple solutions that require only changes to the medium access control (MAC) protocol at the access point.

Voice over Internet Protocol (VoIP) over IEEE 802.11 is becoming very popular due to its low cost and high availability. IEEE 802.11 standard is being widely used not only for data communications but also for voice communications. The result is a widespread availability of WLAN networks and low cost voice communications. Transmitting of voice traffic over WLAN does not utilize all of the bandwidth efficiently. This paper [2] proposes a scheme for increasing capacity. Wireless LAN uses two modes for channel access. These are Distributed Co-ordination Function (DCF) and Point Co-ordination Function (PCF). DCF [1] is suited for traffic, which does not require to be real time. The traffic is sporadic in nature. PCF mode is based on a polling mechanism and is better suited for real time traffic. DCF is more widely used because a large number of devices currently used do not support the PCF mode.

DCF mode is based on CSMA/CA (carrier sense multiple access with collision avoidance). Figure 1 shows the timing diagram [1].

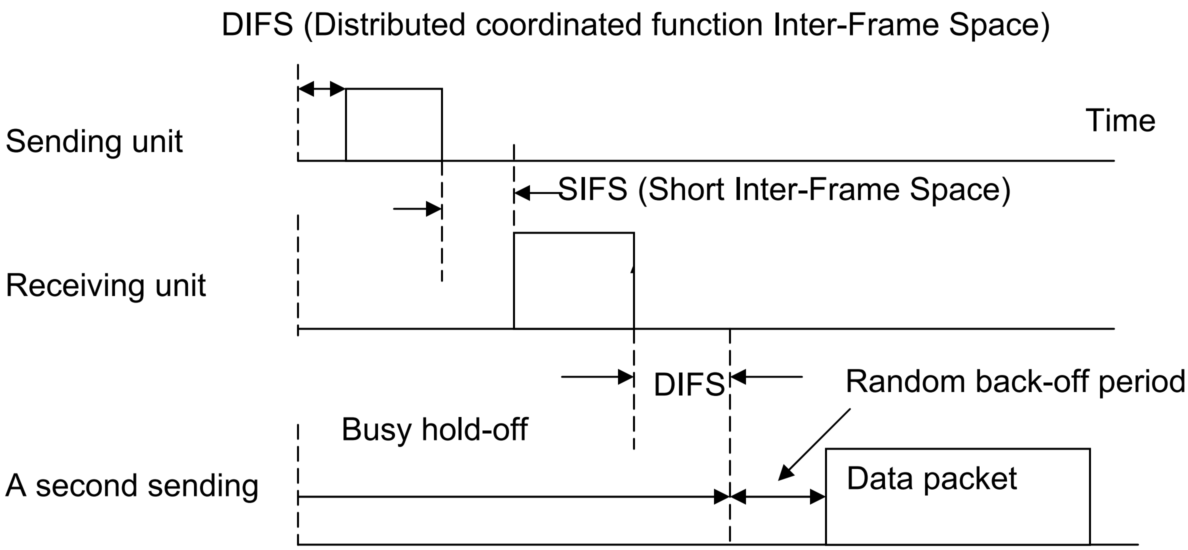


Figure 1: Timing Diagram of DCF

Each station with data to send selects a random number from the set (0, Contention Window). It then waits for a Distributed Inter-frame (DIFS) period of time. If the channel is found idle, it decrements the counter by one every slot till the channel remains idle. When the counter reaches 0, the station transmits. If the channel is sensed busy during the period the counter is decremented, it stops counting. When it senses that the channel is idle, it starts counting down again from the value it reached when the channel became busy. When the receiver receives an error free packet, it sends back an acknowledgment after a Short Inter-Frame Space period (SIFS). There may be a chance that two stations count down to zero at the same time. If both stations transmit at the same time, a collision will occur. To reduce the possibility of another collision, the stations generate another contention window equal to 2*(CW+1)-1 where CW is the contention window measured in microseconds. After a successful transmission, the contention window is reduced to the initial value CW. DCF supports the optional RTS/CTS control where the stations sends an RTS packet and only starts transmitting when a CTS packet is received from the receiving station. The RTS/CTS control is not valid for small real time packets because it would add overhead in the form of RTS and CTS packet without providing a lot of throughput gain.

Transmitting Voice traffic over a packet based network presents Quality of Service challenges. A packet-based network has more overheads in terms of header information. This results in poorer utilization of available bandwidth. This challenge is not related to the media. The wireless channel leads to a more stringent requirement on the multiple access technique used. Wireless channel degrades traffic to a greater extent than a wired link. This is due to interference from other wireless devices such as Bluetooth devices, which operate in the same frequency. Both of these lead to an increase in overhead and lower bandwidth utilization.

IEEE 802.11b can support rates up to 11Mbps. Typical bandwidth requirements for Voice over IP (VoIP) are less than 10Kbps for a single communication stream. This suggests that the number of simultaneous VoIP streams a WLAN can support is around 11M/10K = 1100. This corresponds to 550 VoIP sessions. However, a WLAN working under DCF with a rate of 11 Mbps mode can only support 10 VoIP sessions. This difference is due to the packet overhead and the inefficiency inherent in the WLAN MAC protocol. For large IP packets of 1000 bytes, the transmission time of the payload of 1000 bytes at 11Mbps is 1000 * 8 / 11 = 727 μseconds. The total transmission time for the whole MAC frame is 1562 μseconds. The utilization factor at the MAC layer is 46.5%. A typical VoIP packet at the IP layer consists of a header of 40 bytes and a payload of 10-30 bytes for a total of 50-80 bytes. The transmission time for payload in this VoIP packet is only 56-58 microseconds and the overhead is more than 800 microseconds.

A solution to this problem is to use a multiplex-multicast (M-M) scheme. A typical VoIP packet at the IP layer consists of 40-Bytes IP/user datagram protocol (UDP)/real-time transport protocol (RTP) headers and payload ranging from 10 to 30B. The efficiency at the IP layer is less than 50% [2]. At the 802.11 MAC/physical (PHY) layers, the efficiency is much worse [2]. The transmission time for a 30B payload at 11Mb/second is 30*8/11 = 22μs. The transmission time for a 40B IP/UDP/RTP header is 40*8/11 = 29μs. The 802.111 MAC/PHY layers have additional overhead of more than 800μs attributed to the physical preamble, MAC header, MAC back-off time, MAC acknowledgement (ACK), and intermission times of packets and acknowledgement. The overall efficiency drops to less than 3%.

In an enterprise WAN, VoIP becomes more complicated because WLAN needs to simultaneously support other applications besides VoIP. Providing room for the applications may reduce the capacity for VoIP sessions. Even when VoIP is limited to half the capacity, interference from TCP will cause unacceptably large increases in delay and packet-loss rate of VoIP traffic.

The multiplex-multicast (M-M) scheme eliminates inefficiency in downlink VoIP traffic by multiplexing packets from several VoIP streams into one multicast packet. Many constituent VoIP packets share the overhead of the multicast packet. Various tests have been made [2] with popular voice codecs like constant bit rate (CBR) and variable bit rate (VBR) voice encoding with results showing 80%-90% higher than ordinary [2] VoIP over WLAN with the M-M scheme. The delay incurred in the M-M scheme is well below 125ms leaving plenty of delay margin for the backbone network as VoIP packets travel from one WLAN network to another. However, there is a high loss rate when there is interference by the downlink VoIP traffic.

The loss rate of multicast packets is excessive when there is [2] upstream TCP packets due to packet collisions. The reason for this is that, in multicasting there is no automatic repeat request (ARQ). ARQ is present only in unicasting.

Collided multicast packets are not retransmitted. Excessive multicast packet loss due to collision is a large problem in WLAN. This problem does not exist in the Ethernet.

For VoIP, the analog or pulse-coded modulation (PCM) voice signals are encoded and compressed into a low-rate packet stream by codecs. Codecs generate constant bit-rate audio frames consisting of 40-B IP/UDP/RTP headers followed by a small payload. For GSM 6.10, the payload is 33B. The time between adjacent frames is 20ms. This corresponds to a rate of 50 packets per second per VoIP stream.

<u>Access Point priority based Capacity Enhancement scheme for VoIP over WLAN.</u>

When an Access Point has to transmit N simultaneous voice calls, it must obtain access to the channel. That channel has an equal probability of being granted to another channel, which has just one voice call to transmit. This leads to longer packet queues at the Access Point causing a larger delay.

A proposed solution [1] to this is to create a scheme priority, which is given to an Access Point. This priority is implemented by decreasing the contention window for the Access Point or by decreasing the inter-frame space for the Access Point and making the contention window for the Access point equal to one for voice calls that have a high priority. There is still a chance of collisions but the algorithm, which determines the contention window, should minimize the number of collisions.

VoIP stream typically requires 10kb/sec, an 802.11b WLAN operating at 11Mb/s. If VoIP packets are generated every 20 milliseconds and have a size of 160 bytes, the total transmission time of the packet is:

DIFS + Average Random Back-off time + Data + SIFS + MAC Header + Physical Header

Data size = payload + Header (IP + UDP + RTP) = 160 + 40 = 200 bytes

Total time = 50 + 31*20/2 + 200*8/11 + 10 + 248 = 955.45 microseconds.

For full duplex, one packet needs to be transmitted in the uplink and downlink direction. This implies that one channel needs about 2000 microseconds of channel time. The voice packet arrives every 20 milliseconds, which allows for only 10 simultaneous sessions.

Using the solution where Access points are provided priorities by decreasing their contention windows or decreasing their inter-frame space, we can reduce the time and overhead for each packet to be transmitted to its receiver. The Random Back-off time can be removed to allow for more simultaneous sessions. Packets from an Access Point do not collide with other packets leading to fewer retransmissions and larger call capacity.

After a successful transmission, the contention window size is decreased. This ensures that the AP continues to have higher priority over the channel as compared to other stations.

Reducing the inter-frame space period ensures that if the Access Point has traffic to transmit, it has access to the medium as soon as the channel becomes free. Throughput in a normal DCF degrades after 10 channels.

Optimized Adaptive Jitter Buffer Design for Wireless Internet Telephony.

In order to compare the quality of VoIP with regular telephony [4], one must correlate the performance impact of various parameters on VoIP over WiFi networks. The Adaptive Jitter Buffering (AJB) and latency management can be implemented at the endpoints. The discussed AJB implementation is codec aware, E-model audio-quality aware and coexists with wireless optimizations like frame packing and repeating. It is based on actual network delay characteristics obtained via signalling protocol. The domain of AJB has been researched for a decade since the start of media streaming over the Internet. Newer trends are focusing on codec-aware aspects of jitter buffer technology, error concealment aware aspects and audio quality. It is recognized that no other voice processing functionality has a bigger impact on latency than jitter buffer.

The Adaptive Jitter Buffer (AJB) Design has the main purpose to smooth out variations in network delay. If an audio frame is delayed by a few milliseconds, the buffer waits for it to arrive, playing white noise or some audio reconstructed from previously received frames. Introducing breaks of a few milliseconds in playback does not adversely affect quality of playback. Delays beyond a threshold are assumed to be due to packet loss. The jitter buffer skips to the next available frame to play back. Over some time, the waiting times for delayed frames accumulate and playback is out of phase with the recording by a large time lag. The jitter buffer becomes very long as newer frames arrived on time while the buffer is waiting for older (delayed) frames. When the buffer becomes very long, real-time voice quality deteriorates because playback delays become untenable. It is important that the jitter buffer reduces playback delays by voluntarily omitting playback of a few frames, deleting them from the buffer and continuing play back from a later frame. However, this technique reduces the quality. In the attempt to improve voice quality by decreasing delay, the buffer must drop frames, which in turn causes voice quality to worsen. There is a trade off required to satisfy both decrease in delay and quality.

The AJB algorithm must ensure that the combination of delay and loss in playback stream keeps resultant audio quality at a reasonable level for the human ear. The key to intelligent AJB design is to know how long to wait for delayed frames, when to voluntarily skip frames (when to shrink the buffer), and how many frames to skip without adversely impacting audio quality. This will depend on the type of codec used and error concealment techniques. The overall quality depends on the play-out buffer algorithm used. The play-out buffer algorithm has to work with both capacity optimization techniques and redundancy mechanisms. Often, multiple audio frames are packed into a single IP packet to improve WLAN transmission efficiency by accumulating protocol overhead over several frames. If we assume that 60 ms worth of audio frames are packed into a single IP packet in order to improve network transmission efficiency in the WLAN. If the maximum length of the buffer is less than 60 ms, then the audio frames will be lost from each and every IP packet. If we assume that an audio frame contains 10 ms of encoded audio information, we can use the *Np* (packing factor) to denote the number of 10 ms audio frames packed into a single IP packet.

One of simplest and most commonly used redundancy mechanisms is frame repeating. The transmitter piggybacks older audio frames with newer ones in an IP packet so that if previously sent audio information has been lost in the network, the piggybacked frames can be used at the receiver to reconstruct the audio. Sequence numbers in the RTP header are used to order the frames appropriately so that the receiver can arrange the frames for play back. We use *Nr* (repeating factor) to denote the number of older audio frames packed into the single IP packet. To ensure equal redundancy for every audio frame, *Nr* is an integral multiple of *Np*. For example, if Np = 6 and Nr = 2 * Np, then an IP packet contains 60ms of new audio data along with 120ms of older audio data that immediately preceded the new 60ms worth of data.

Using a repeating factor Nr of 2 * Np, with Np = 6 means that jitter buffer must be willing to expand to 180 ms. It must be willing to insert a frame delayed by up to 120 ms into the buffer rather than discarding it for being too late.

The optimal levels of frame packing and frame redundancy (Np and Nr), depend primarily on the network characteristics such as capacity and packet loss rate.

The analysis below is an analysis of the factors that affect audio quality. The output of the analysis is a set of guidelines for an optimal adaptive jitter buffer (AJB). The analysis is based on the ITU-T E-model.

The E-model conversational audio quality rating R is defined as:

$R = R0 – Is – Id – Ie + A$

ITU-T G.107 recommendations mention that a rating of over 90 means that users are very satisfied while one of over 80 means users are satisfied. A rating of over 70 implies that users are dissatisfied.

R0 is the inherent SNR (Signal to Noise Ratio)-dependent quality of the audio.

Is depends on the amplitude of the signal.

Id is the impairment to the audio quality due to delay.

Assuming perfect echo cancellation, Figure 2 plots the delay impairment Id as a function of one-way mouth to ear delay. This includes the packetization delay, the end-to-end network delay (transmission and queuing) experienced by the IP packet, and any delay experienced by the decoded audio frame waiting in the AJB.

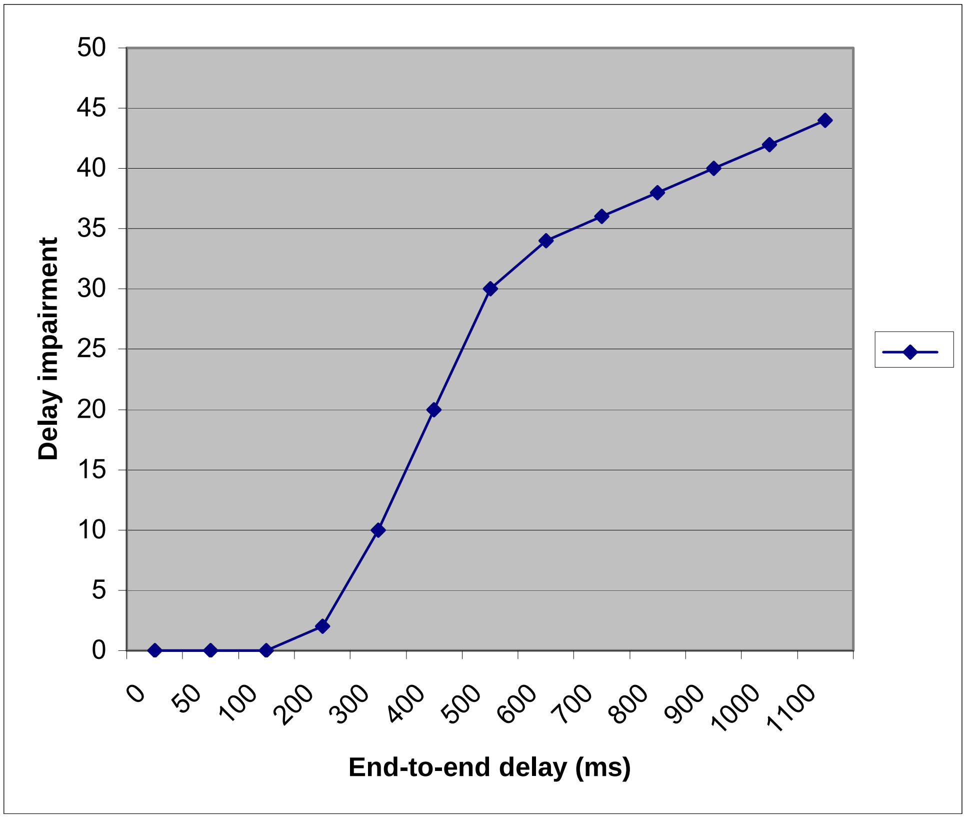


Figure 2: E-model delay impairment Id as a function of one-way mouth to ear delay

Ie is the impairment to the audio quality due to errors.

The value of Ie depends on inherent loss due to the use of lossy codec, residual network loss, and any loss introduced by the AJB algorithm. The AJB drops audio frames that are deemed to have arrived too late. It also drops audio frames during its shrinking process. A is the access advantage which is the factor that determines how much poorer audio quality a user is willing to endure for the convenience of using advanced communication technology over POTS.

The value for A is 10 for mobile telephony and 20 for satellite telephony. There is no number available for VoIP or VoIP over wireless because these are relatively newer technologies.

AJB Algorithm

The proposed AJB algorithm is a) codec-aware, b) E-model audio quality aware, c) co-existing with wireless optimizations, and d) based on actual network delay characteristics as opposed to packet inter-arrival times.

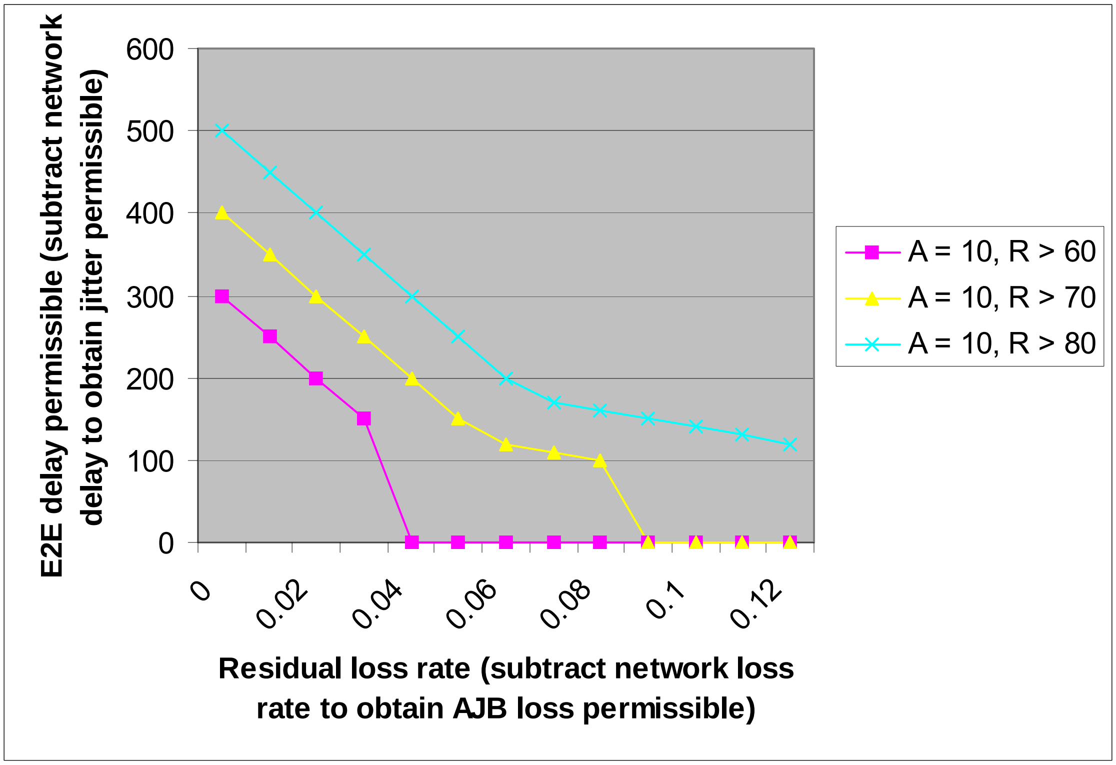


Figure 3: Adaptive jitter buffer (AJB) design constraints - delay is in ms.

The AJB algorithm depends on the results of an analysis to determine the acceptable delay-loss area for a specific codec chosen. The result of the analysis is a plot of what playout delay is permissible at various loss rates. The acceptable region is largest for the lowest quality (R > 60). The acceptable region is smallest for the highest quality (R > 80).

The plot shows that for a voice quality rating R > 80, the end to end delay one way should be less than 250ms for a 1-2% loss rate. The three sources that comprise frame playout delay are: 1) one-way end-to-end network delay, 2) packetization delay arising from the frame packing optimization, 3) delay because the original frame was lost and only a redundant frame was received later. The optimizations are chosen such that the total delay and loss rate combination lie in the shaded region. Once this operating delay-jitter point is plotted, the area to the right and above this point, up to the R > 80 line is available for the AJB to adapt within.

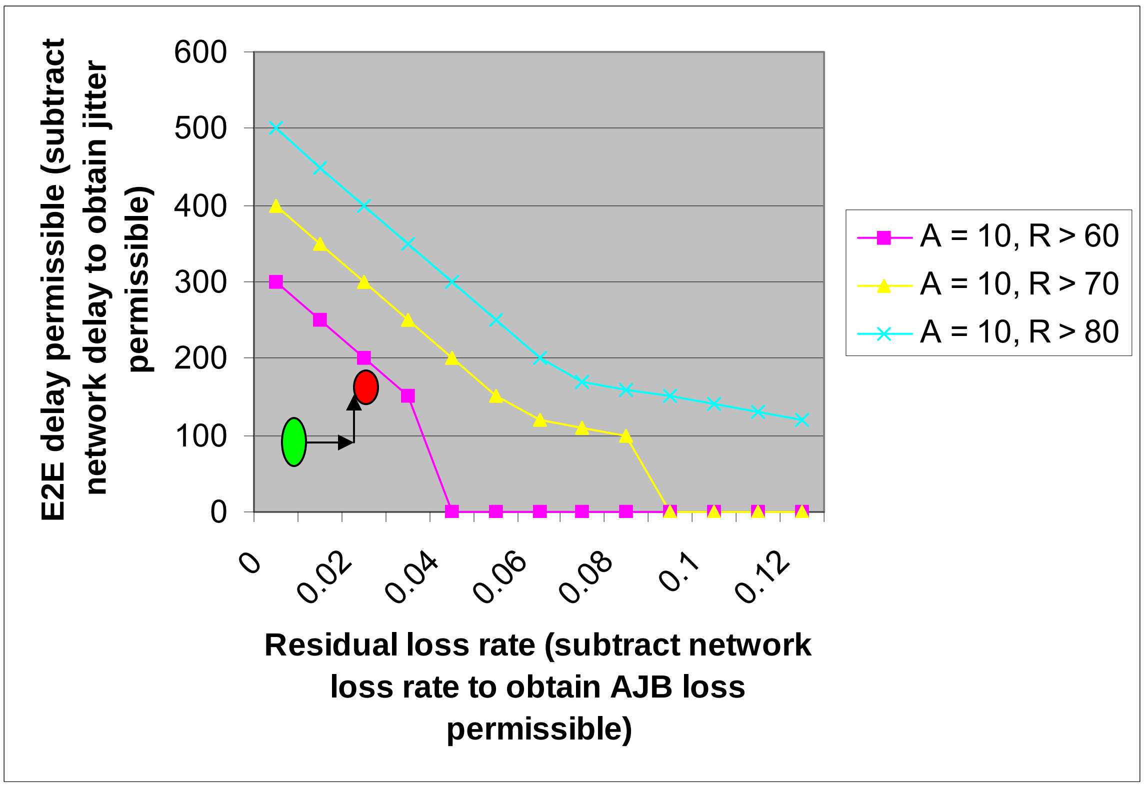
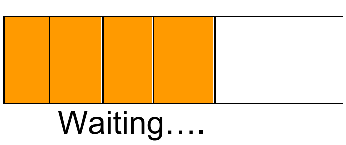


Figure 4: Adaptive jitter buffer (AJB) design constraints - delay is in ms (waiting for packet).



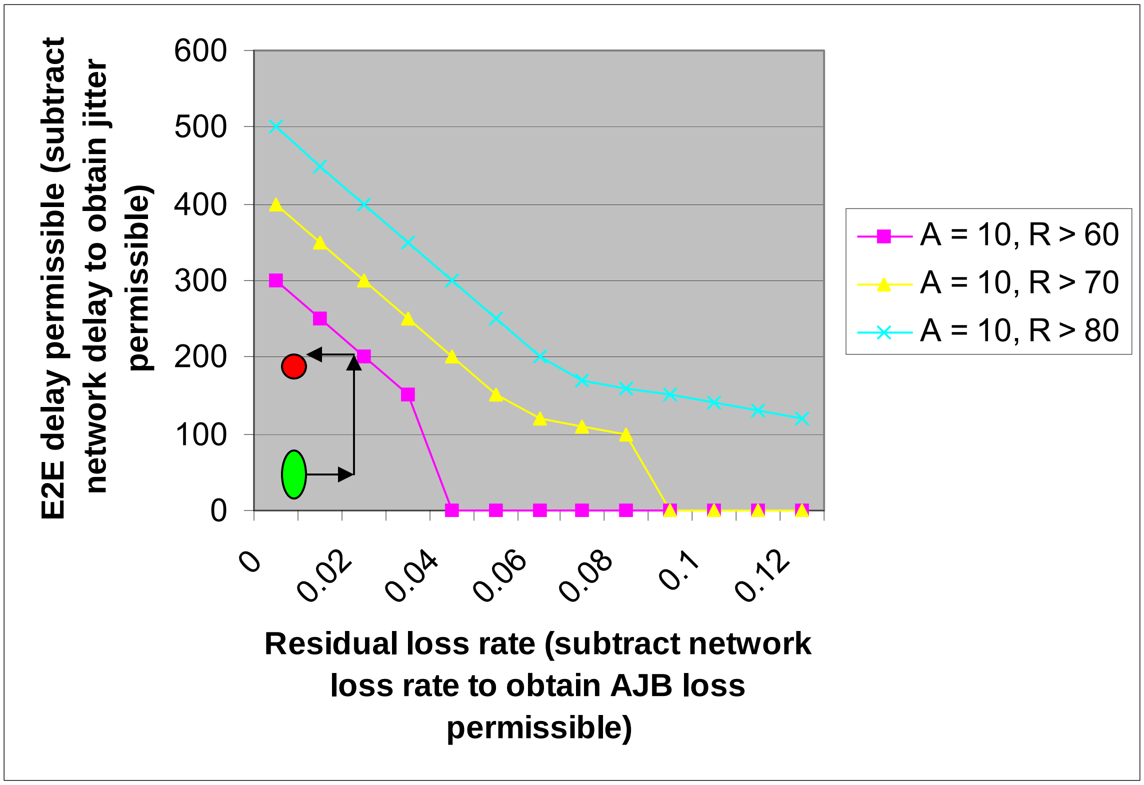
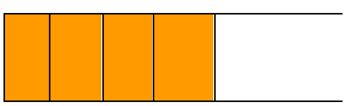


Figure 5: Adaptive jitter buffer (AJB) design constraints - delay is in ms (packet arrived).



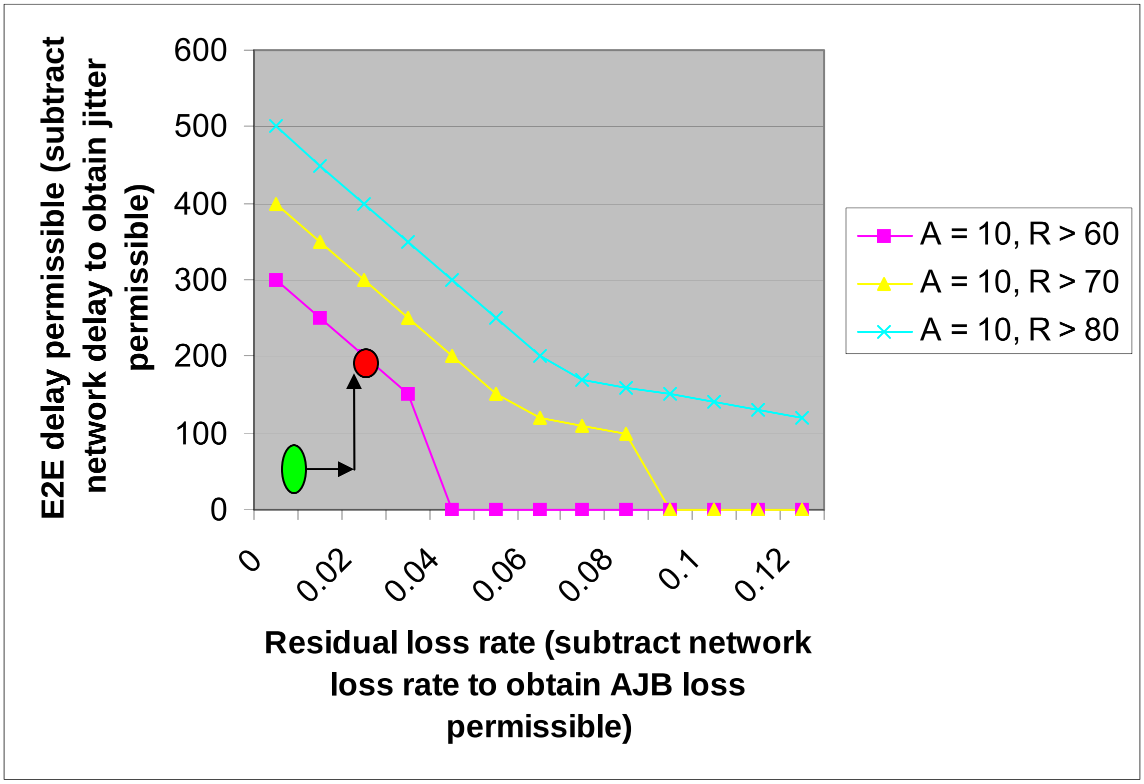
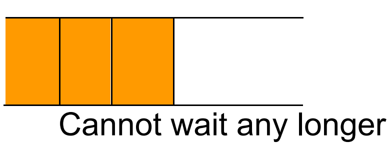


Figure 6: Adaptive jitter buffer (AJB) design constraints - delay is in ms (cannot wait any longer for packet).



The above Figure shows that for a voice quality rating R > 80, the end-to-end delay one way should be less than 250 ms for a 1-2% loss rate. Knowing the network delay, an optimal value of Np >= 3 can be obtained that improves call capacity while leaving sufficient room for the AJB to adapt to instantaneous delay jitter. If the end-to end network delay estimate is 120 ms one-way, then choosing Np = 6 means that the permitted jitter is only 70ms (250ms – (120 ms network delay + 60 ms packetization delay)), for R > 80. Since each packet contains 60 ms of audio, there can only be one out of order packet permitted. Therefore, there is no point in having Nr > 1 * Np because repeated frames are out of order frames. This also means that the 1-2% loss rate on the basis of which 250 ms tolerable delay was calculated. The value of Np = 3 is preferable since jitter buffer of up to 100 ms (250 ms – (120 ms + 30 ms)) of delay jitter or 3 out of order packets. This means that the redundancy factor Nr = 3 * Np is permissible and yields a loss rate of 1-2%.

If the requirement for R > 80 cannot be met, the algorithm attempts to attain an audio quality rating of R > 70 and so on. The above figure rules out a frame packing ratio Np of as much as 9 because the delay jitter characteristics of Np = 9 and Nr = 2 * Np means that the packetization delay is 90ms and jitter to be allowed is 180 ms. This adds up to 270 ms which exceeds the 250 ms and we have not yet accounted for network delay.

In the above diagram, the bigger dot represents the current operating delay/loss point of the AJB algorithm. When a frame is not available at playback time, the horizontal arrow indicates the increase in loss rate due to this event (burstiness of the loss must be accounted for here). The horizontal time indicates how long the AJB algorithm is permitted to wait for the frame. The time the AJB is permitted to wait for the frame must be greater than the frame-redundancy delay.

The following is the outline for the AJB algorithm.

Initialization Phase

1. Choose highest E-model quality
2. Find one way end to end network delay estimate
3. Find Np/Nr combination that keeps:
    a) End to End delay + packetization delay + Out-of-order delay <= Delay-for-loss rate-1%, and loss rate <= 1% OR
    b) End to End delay + packetization delay + Out-of-order delay <= Delay-for-loss rate-2%, and loss rate <= 2% OR
    c) and so on until loss rate exceeds the maximum allowable loss rate for this E-model quality.
4. If no such Np/Nr combination is found, that also improves network capacity, repeat for next highest E-model quality.

Operating Phase

1. Maintain running average of loss rate so far.

2. When frame is unavailable for play-out, allow expansion of AJB up to area demarcated for chosen E-model quality, and computed average loss rate.
3. Allow periodic shrinking provided the increased loss rate and decreased play-out delay still fall within area demarcated for chosen E-model quality.
4. If a drastic change in loss rate running average or end-to-end delay estimate occurs, re-start Initialization phase.

The above figure shows the working of the AJB algorithm, specifically Step 2 of the operating phase. When a frame is missing at playback time, the horizontal and vertical arrows are used to compute resultant loss rate, if it is lost and waiting time permitted. If the frame arrives during this waiting time, the loss rate is reset. The smaller dot shows the new operating point of the algorithm. If the frame does not arrive in time, the operating point is updated as shown and the algorithm continues.

<u>Throughput Performance of a Wireless VoIP Model with Packet Aggregation in IEEE 802.11.</u>

The components of VoIP include grabbing/reconstruction, compression/decompression, transmission/reception over IP in sender/receiver parts respectively [5]. The analog speech signal is first encoded into a digital representation to be able to be transferred over IP network. At regular small intervals, blocks of digitized speech information are sent over the network from the transmitter to the receiver. On the receiver side, the digitized block is transformed back to an audio signal, which is output to the speakers. Digitization of voice data includes sampling and quantization. The sampling rate and number of bits used in quantization determines the rate of data transmission before data compression. The speech signals of humans can be frequencies of beyond 12 kHz. High quality communication is attained on a telephone system by transmitting frequencies of below 4 kHz. Nyquist theorem suggests a sampling rate of 8 kHz is enough for digitization of speech.

The range of amplitudes of the voice signal is covered by at least 12 bits when uniform quantization is used. A uniform quantization with 8-bit quantization also gives telephone quality. The required bandwidth for telephone quality conversation is 64kbps, which is 8 kHz times 8 bits per sample. To avoid delay jitter, which is the difference in packet arrivals, a buffer is used at the receiver. Instead of playing the voice data immediately after the reception of the packet, the packets are buffered. Although the buffer slightly increases the delay, it increases the probability of playing the packets consecutively without interruption.

The number of bits in each packet is very important in terms of delay and packet loss. To reduce the amount of loss, a packet should contain only a small amount of voice signal. If the packet is lost, only a small part of the conversation is lost. If the length of data in each packet decreases, the overhead in the network increases due to the low ratio of data length to length of the header fields.

Performance issues for VoIP call routing in a hybrid ad hoc office environment.

Designing a routing protocol able to support VoIP calls over a multi-hop [6] environment is not straightforward. VoIP calls are sensitive to packet loss and should be bounded on transmission delay to assure correct communication. A single VoIP flow traversing 6 wireless hops each transmitting at 2 Mbps can completely occupy the bandwidth in a wireless network. Modern implementations of wireless adhoc routing protocols limit the maximum number of hops that a flow can traverse. They rely on a fixed networking infrastructure in a hybrid ad hoc network fashion to complete communication between all stations in the network. Determining the maximum number of hops to be traversed highly depends on the scenario, application and number of concurrent flows to be transported.

On the other side, the end to end delay of a VoIP flow traversing a wireless multi-hop network is rarely an issue when flows traverse a reduced number of hops or when low-end devices are used. When low-end devices are used, forwarding nodes generally have short buffering capacity that rapidly overflows in case of saturation of the network, resulting in packet loss, not delay. Jitter fluctuations are generally overcome using play-out buffers at receivers.

Disconnections have an impact on quality of VoIP calls. Wireless multi-hop network paths are prone to suffer disconnections. Disconnections trigger the route re-discovery process in the routing protocol to find a new appropriate path to support the communication. This may lead to packet loss and/or delay that affect the quality of the ongoing VoIP calls. When designing a routing protocol for wireless multi-hop networks, the number of route rediscovery processes should be kept low and fast in order to support VoIP calls.

The topics covered are:

1) The impact of number of hops on the quality of VoIP calls
2) Providing a framework to analyze the impact of route disconnections on the quality of a VoIP as perceived by the end user.

Results presented aim at providing a meaningful guide to the design of efficient strategies and protocols to support VoIP communications over wireless multi-hop networks. The results show that controlling the mutual interference between multi-hop nodes collaborate in the forwarding of a single VoIP flow, unbounds the number of hops that this flow can traverse, as long as end to end delay is kept acceptable. Secondly, it shows how codec adaptation can help increase the number of clients supported in a multi-hop collaborative environment. Thirdly, it proposes an aggregation strategy at the application layer and shows how efficiently it increases the number of VoIP flows supported in a wireless multi-hop network. It proposes a framework to analyze the impact of route rediscovery disconnections using extensions to the E-model.

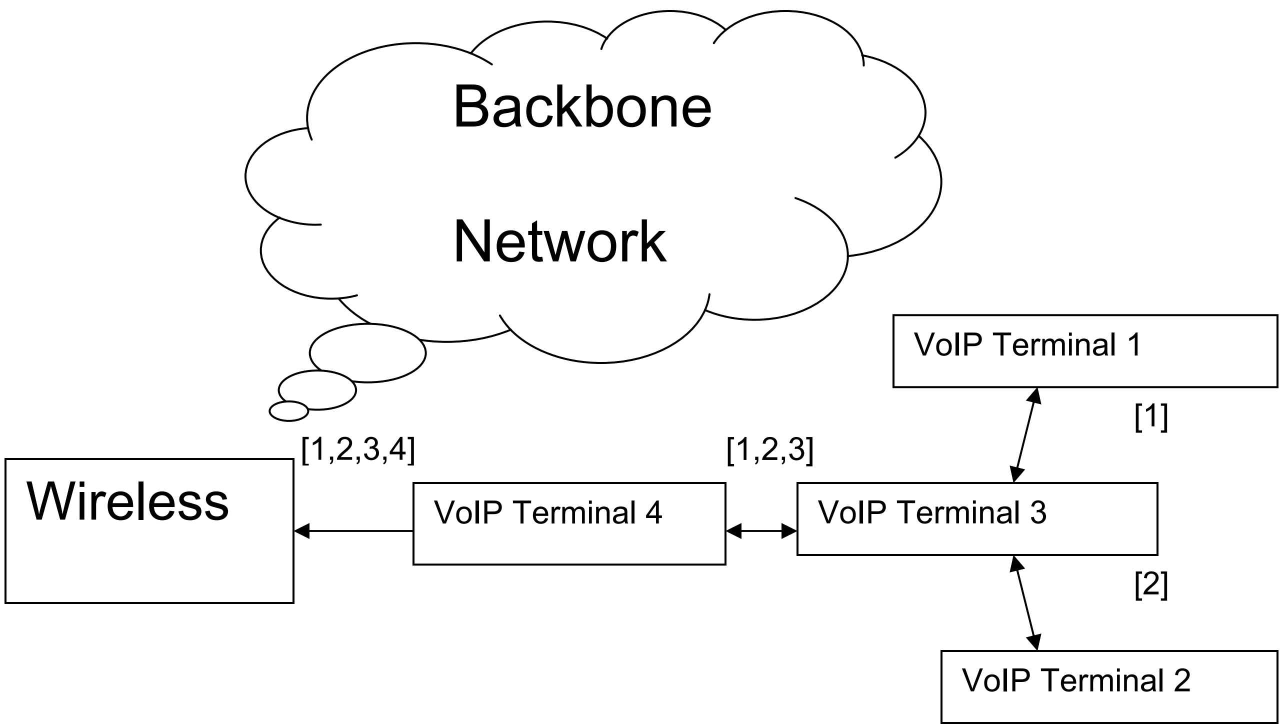


Figure 7: Multi-hop network scenario

Experiments were carried out within EXTREME framework. This is a multi-purpose networking experimental platform currently under development within the Centre Tecnologic de Telecomunicacions the Catalunya (CTTC) in Barcelona.

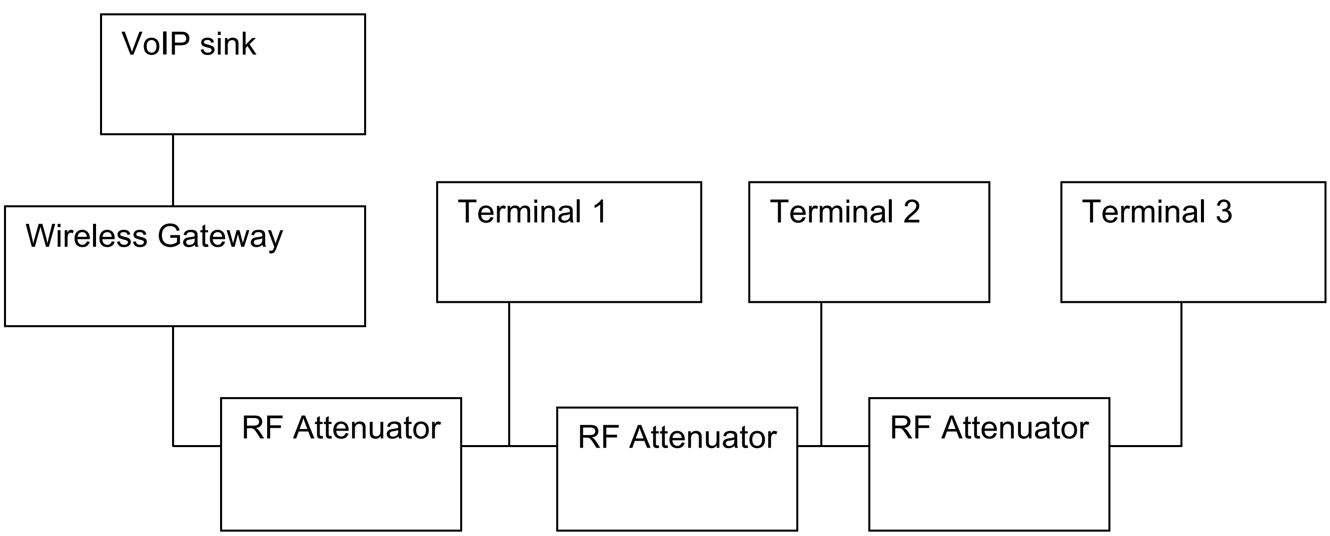


Figure 8: CTTC Experimental set-up

The impact of the number of hops traversed on the quality of a VoIP calls is analyzed. Figure 9 shows the R-factor obtained when computing the E-model at the VoIP sink when the terminal it communicates with is located from 1 to 7 hops away. While it is a bi-directional communication, we do not print the curve obtained at the wireless terminal as it is practically the same as the one obtained at the VoIP sink node. All nodes are configured to transmit at 2Mbps physical rate.

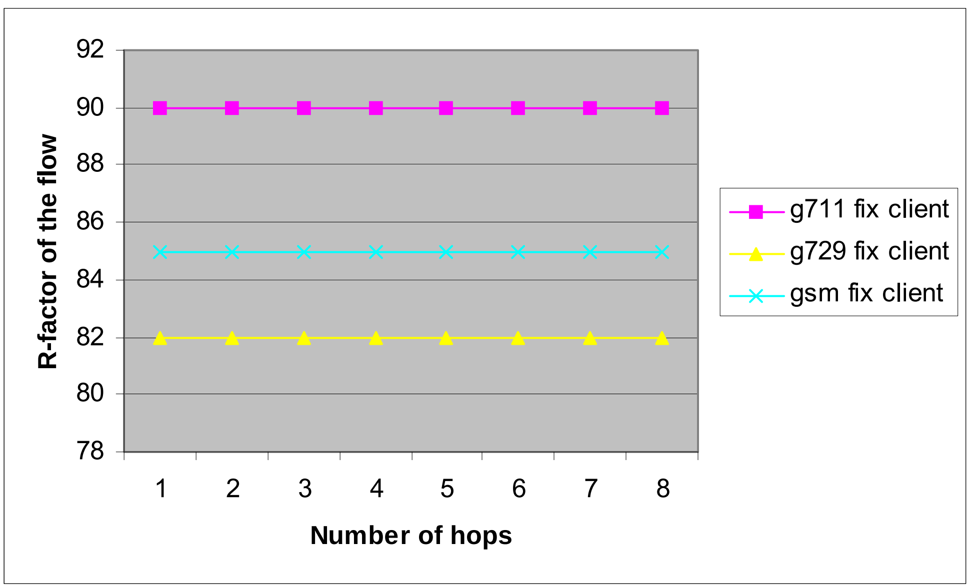


Figure 9: R-factor vs. number of hops traversed by a single flow as observed at the VoIP sink node

At any time, any node contends for channel access with at most four other terminals, so its available bandwidth will be divided, at most, by four. The above figure shows that a conscious interference-aware deployment of wireless multi-hop network can help, in the absence of other background traffic to increase the maximum number of hops that a VoIP flow can traverse without suffering any quality degradation.

Impact of number of hops on multi-flow VoIP quality

Now we study the impact on the quality of voice conversations when each one of the terminals present in the network starts a VoIP call with the VoIP sink node. The idea is to study the maximum number of terminals supported in such a scenario. This is a chain topology, so there is only one route possible from each one of the terminals and the VoIP sink node.

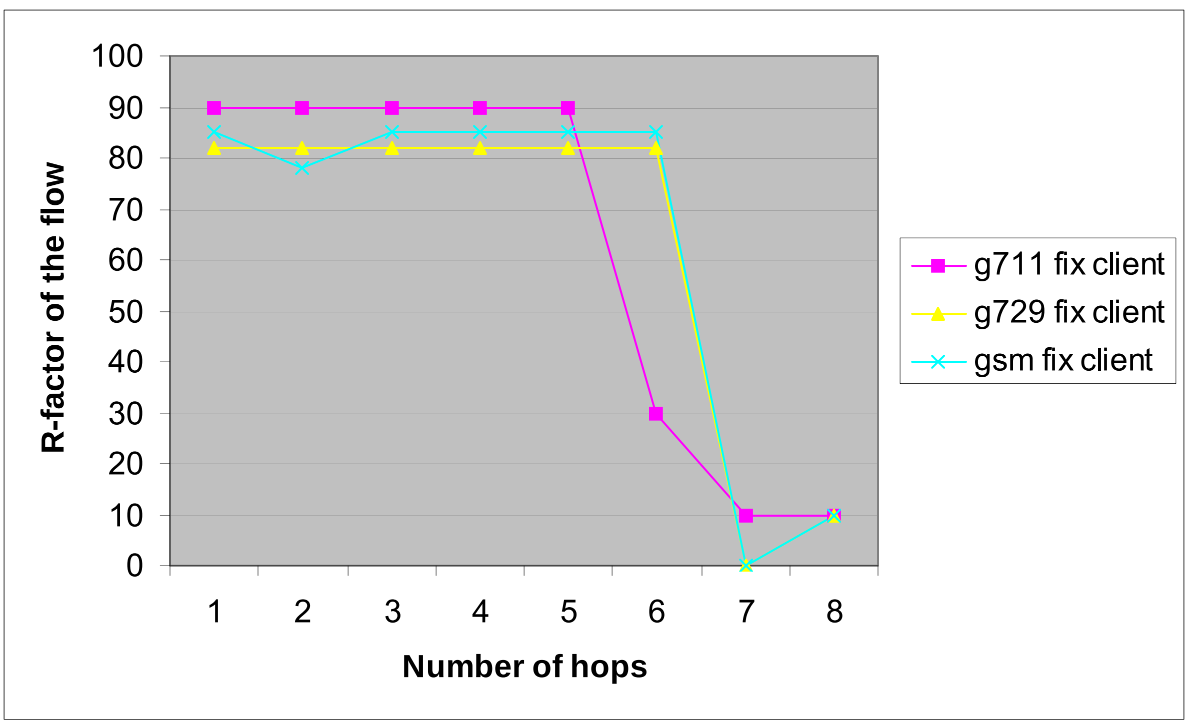


Figure 10: R-factor vs. number of hops traversed by a single flow as observed at the VoIP sink node

Figure 10 shows the quality of the VoIP call between the last terminal and the VoIP sink at the VoIP sink node.

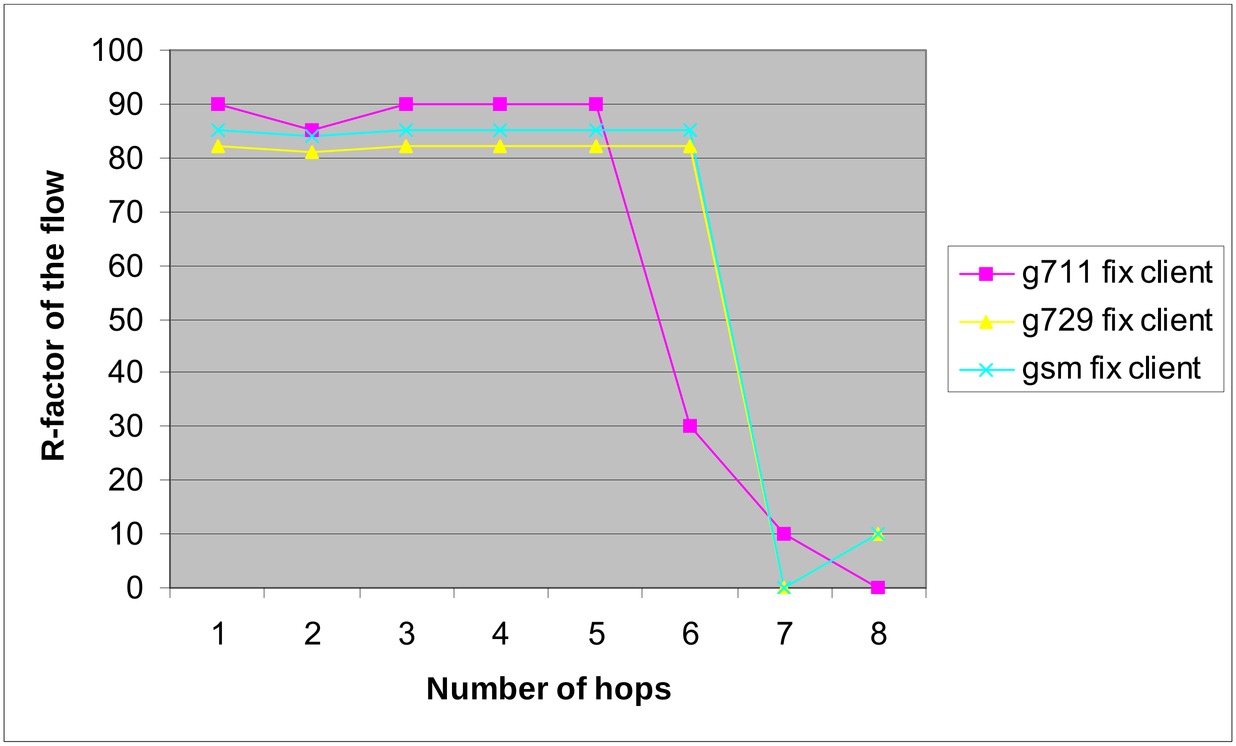


Figure 11: R-factor vs. number of hops traversed by a single flow as observed at the Wireless Terminal

Figure 11 shows the quality of the VoIP call between the last terminal and the VoIP sink at the VoIP sink node. The above figures plot the voice quality versus the total number of terminals (clients) maintaining a VoIP conversation with the VoIP sink node. The number of terminals is equivalent to the number of hops traversed by the VoIP flows going to and from the last wireless terminal.

The G.711 codec can only sustain up to 5 VoIP calls in the scenario described with an acceptable quality (R>70). The number of VoIP that can be sustained at an acceptable quality increases when using G.729 or GSM codecs as the system can sustain up to 6 calls. The observations that can be raised from the figures are that:

1) When the network saturates, the R-factor suffers a sudden breakdown preventing any communication between the last node and the VoIP sink.
2) When using different voice codecs, a different number of hops are reached. This suggests that a voice codec adaptation might be an efficient strategy to support a higher number of voice calls in saturated wireless multi-hop environments.

From the above figures, it shows that not only communications with the last terminal in the chain, but practically all the rest of the VoIP communication fall into unacceptable voice quality situation when the network enters saturation. This change is also abrupt, in a breakdown manner, which challenges the design of admission control mechanisms.

Proposed application layer aggregation of packets and its impact on the number of VoIP calls supported

In order to increase the number of VoIP calls supported in a mesh network, an aggregate strategy can be applied at the networking stack of each one of the wireless mesh nodes. The implications of this are that the networking stack of all terminals must be modified. This modification has to be done at the OS level, which increases complexity of the task. Also, considering that short buffering capacity is expected in the relaying terminals, not much forwarding opportunities might arise from packet aggregation.

The application can aggregate various voice packets into one, prior to the send process. This process reduces the amount of resources required to keep the communication. The number of packets sent is reduced and this leads to a reduced amount of overhead to send them. This strategy has an impact on the end-to-end delay of packets. In order to conduct aggregation, some packets are delayed on purpose. However, end-to-end delay is not an issue in the targeted scenario.

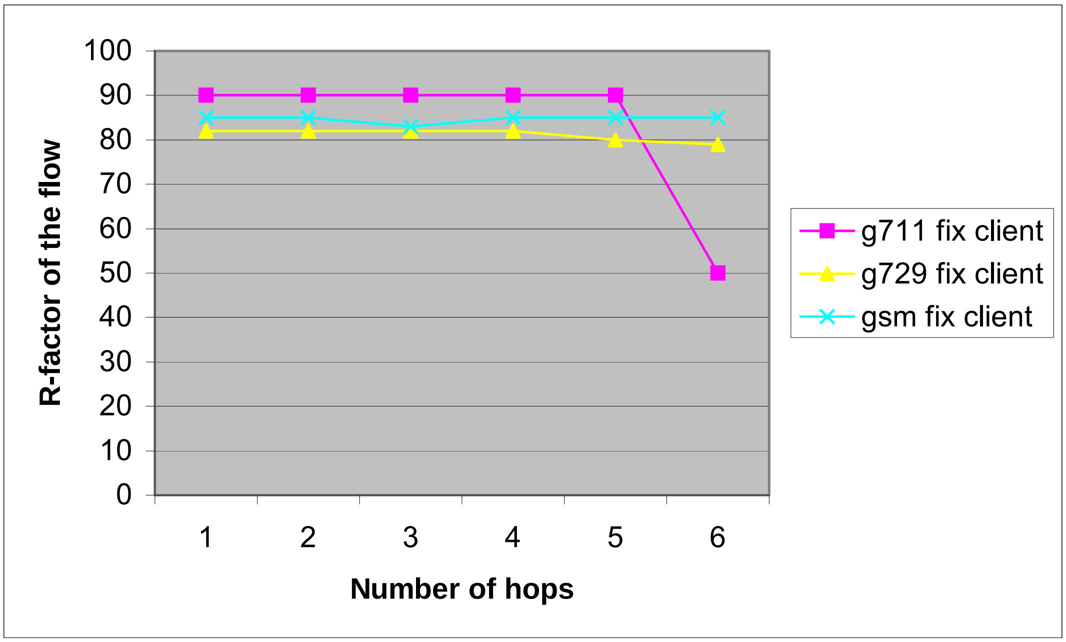


Figure 12: R-factor vs. number of hops traversed by a single flow as observed at the VoIP sink node

The above figure shows the plots when the stations are applying the aggregation strategy. Each one of the terminals aggregates at the application layer two VoIP packets into one and sends them together to the next hop towards the VoIP sink node. The extra delay suffered by some packets due to the aggregation process is accounted for in the computation of the R-factor value. The figure shows that the aggregation effectively serves the purpose of supporting a higher number of active VoIP terminals in the network chain. The results suggest a possibility of including aggregation strategies at the application layer instead of the lower layers as it extends the maximum number of terminals supported in the scenario.

Impact of Route Re-Discovery Latency Time on the Voice Quality

The impact of route re-discovery process on the quality of voice call as perceived by the user is analyzed below. The assumption is that packet losses are uniformly distributed over time. An effect is the recency, which is based on the fact that people tend to remember the most recent events.

The below figure shows the R-factor value perceived by a VoIP versus the time elapsed since the route re-discovery disconnection finished.

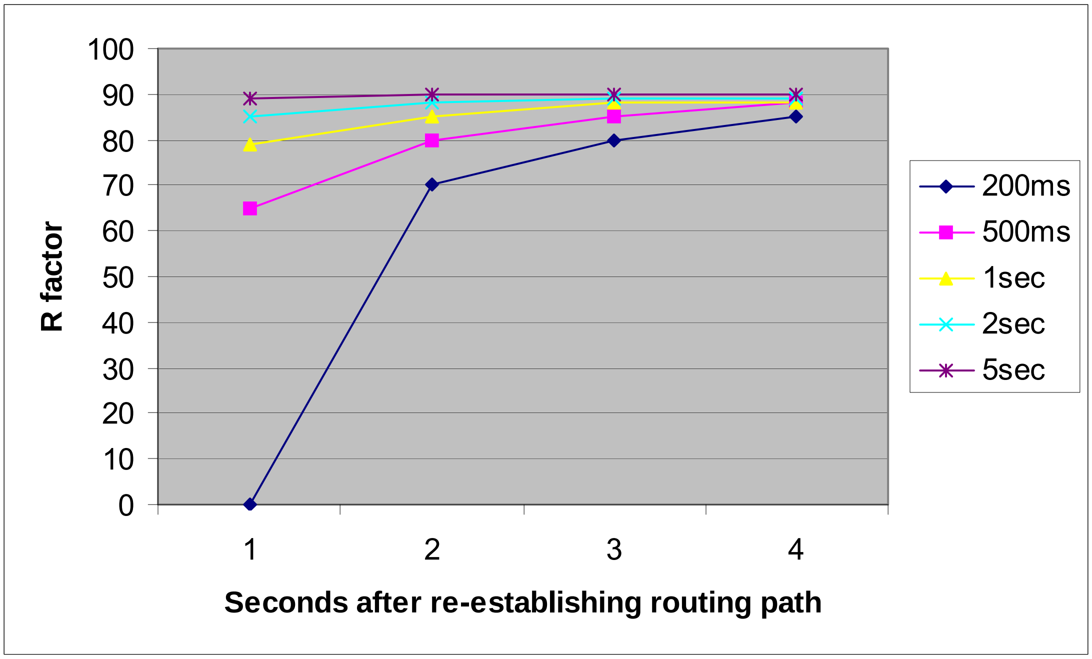


Figure 13: Impact of route rediscovery disconnections on the quality of VoIP calls when using the G.711 codec

From the above figure it can be observed that a long disconnection is preferable to several shorter frequent ones as the user may rapidly forget about the single disconnection but would not tolerate frequent shorter ones.

Empirical studies of wireless VoIP speech quality in the presence of Bluetooth interference.

Both Bluetooth and 802.11b wireless devices operate in the same frequency range [8]. VoIP traffic is being routed over wireless network segments where packet loss can be an issue. Real time voice communication has two issues: reliability and latency. Telephone service that cuts out frequently in the middle of a conversation is difficult to use and connections with an end-to-end latency of greater than 150ms are not recommended for high quality voice applications. There is no guarantee that the packets will arrive at their destination, and they can arrive in the wrong order and with varying delays.

Transmission Control Protocol (TCP) retransmits all dropped packets and is thus very reliable, but requires a large buffer that introduces an unacceptable latency overhead into the connection. VoIP calls use User Datagram Protocol (UDP) packets instead, because of their lower latency. Unfortunately, UDP has reliability problems, because lost UDP packets are never resent. Increasingly, VoIP calls are being routed over wireless network segments. However, wireless networks are prone to interference. Interference is especially widespread in medium range 802.11b networks because their frequency space – the 2.4GHz band, extending from 2400 to 2483.5 MHz in the US is quite crowded. Microwave ovens, cordless phones and other wireless networking schemes exist in this frequency range.

One of the most difficult challenges for 802.11b is operation in close proximity to Bluetooth devices. Bluetooth is a wireless networking standard that is intended for short-range cable replacement, rather than medium range networking like 802.11b. Since it is a complimentary standard to 802.11b, it is feasible that both Bluetooth and 802.11b devices would be active at the same time. Blue-tooth uses Frequency-Hopping spread spectrum (FHSS) to distribute itself over the 2.4 GHz band. It can cause a high level of packet loss on an 11Mbps 802.11b link using Direct Sequence Spread Spectrum (DSSS).

Below is a simple set-up containing one 802.11b equipped laptop associated with one access point. The access point is connected to a 10Mbps Ethernet hub. Also connected to the hub is a desktop computer.

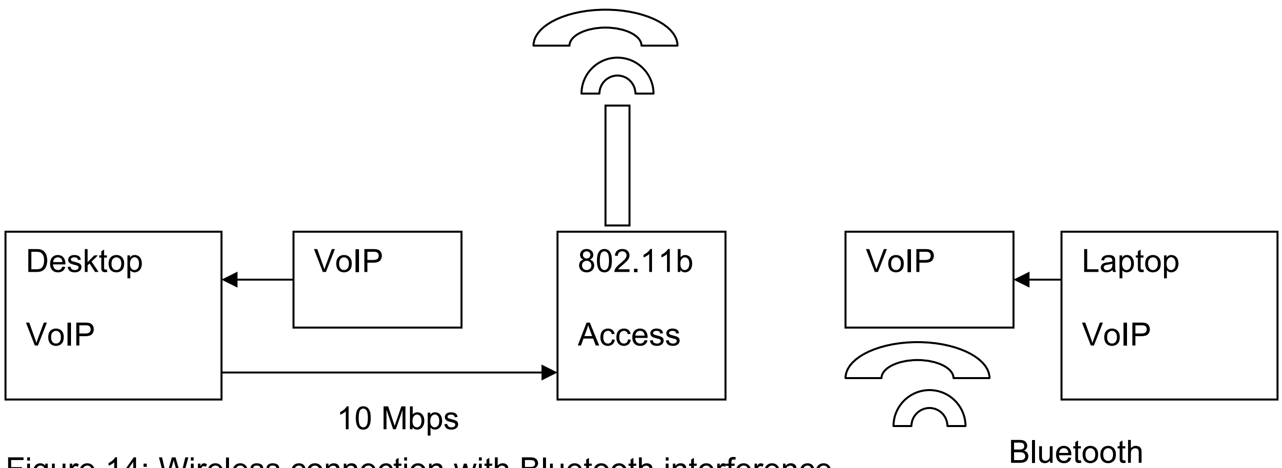


Figure 14: Wireless connection with Bluetooth interference

The study [8] took voice quality measurements under sixteen different conditions to characterize the usability of a VoIP call over the 802.11b link. The signal level of the 802.11b link was varied from –75dBm to –85dBm (in 5 dB steps) by using RF-absorbent material and metal objects to attenuate the signal from the access point to the laptop. The average noise floor was –95 dBm so the corresponding SNRs of the link were 20 dB, 15 dB and 10 dB.

The signal levels were low enough to be affected by Bluetooth interference, but high enough to support a stable network connection. At each signal level, voice quality tests were run without Bluetooth interference, and then with one, two, three, and four Bluetooth interferers active.

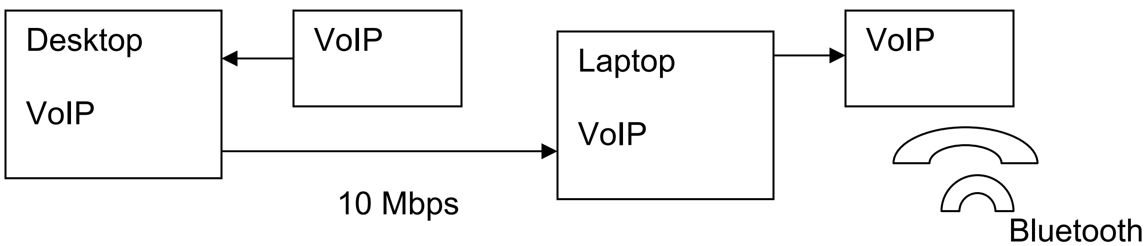


Figure 15: Directly connected laptop with Bluetooth Interference

Other tests were done with the Ethernet cable connecting the Desktop and Laptop directly. The devices used the H.323 set up protocol, and the G.723.1 vocoder at 5.3 Kbps. Although G.723.1 also specifies a 6.3 Kbps data rate, this mode was not supported by the hardware, and therefore was not tested. The vocoder used 30 ms frames and specified a method for frame-erasure concealment. The VoIP devices originally used low-quality headsets consisting of a condenser microphone and ear-bud speaker to record and play back sound. These headsets were replaced with connections to a third computer’s soundcard. Because the VoIP devices put a DC bias (~3 V) on the microphone signal pin, a 470 microF blocking capacitor was used in series with the input signal. This value was large enough so that only trivial signal attenuation occurred within the frequency range of interest. The capacitor and all cables used in the experiment were shielded. To test each VoIP link, forty different test sequences were transmitted and recorded. A third computer called the ‘audio computer’ provided input to the VoIP device attached to the desktop. Simultaneously, a recording of this input signal and the output from the VoIP device attached to the laptop was made. In this way, forty input-output pairs from the VoIP system were obtained. An in-house piece of software, Audio Play, Record, and Estimate (APRE) was used to perform this part of the testing.

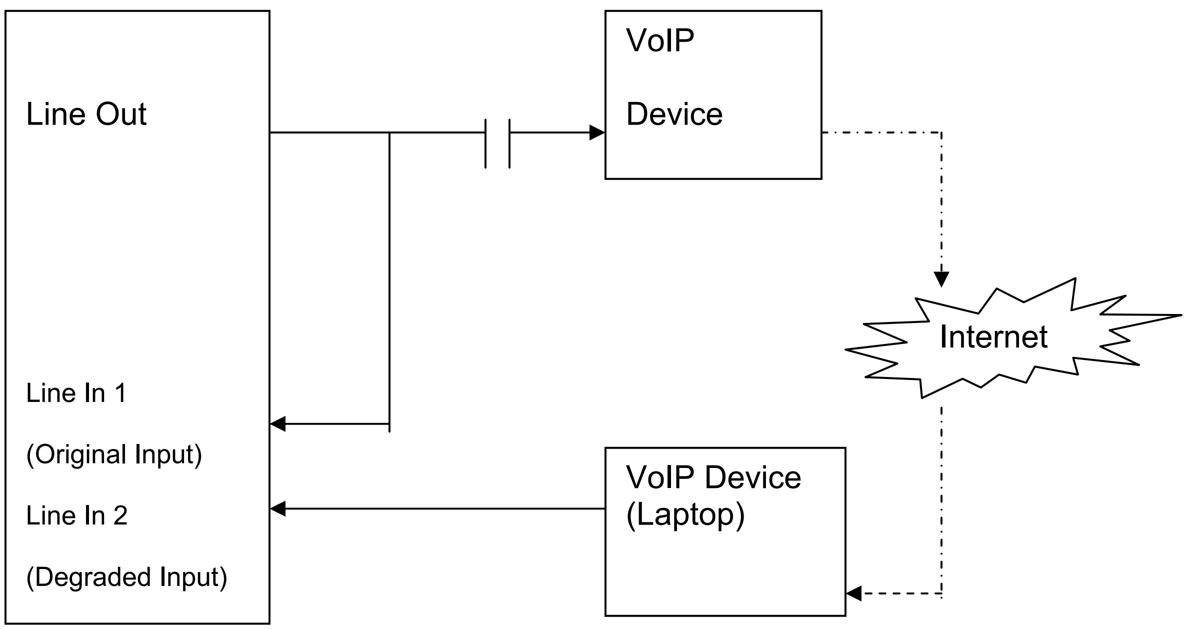


Figure 16: Audio signal block diagram

Each sentence was recorded monaurally with a sampling rate of 8 kHz. A speech quality scale called MOS (Mean Opinion Scale) was developed with a value from 1 to 5 with the corresponding English descriptions ‘bad’, ‘poor’, ‘fair’, ‘good’, and ‘excellent’. A score of 4 is considered ‘toll quality’ and is the goal of most commercial VoIP implementations.

To create Bluetooth interference, four pairs of Bluetooth transceivers were acquired and installed in other computers. A file transfer was set up between a pair of transceivers to simulate Bluetooth activity and to create a constant source of interference. The Bluetooth transceiver sending the file was placed 1 m away from the laptop’s 802.11b radio. Up to four interferers could be created in this manner.

The quality of a VoIP connection over a wired network was examined and the results formed the baseline to which other results were compared. The same measurements were taken when using the wireless network. The voice quality was observed when the level of Bluetooth interference changed. The change in Bluetooth interference affected the Signal to Noise ratio (SNR) of the network.

Speech quality over the wired network exhibited little variation with the variation in Bluetooth interference. However, the wireless network showed great variation in signal level with the number of active Bluetooth interferers. With no interference, the distributions of MOSs at 20dB and 15dB were almost identical and compared favourably with the speech quality from the wired network. Even without interference, the results at 10dB showed quality degradation. With one or two interferers active the voice quality was poor. When three or four interferers were active the 20dB SNR showed a clear advantage.

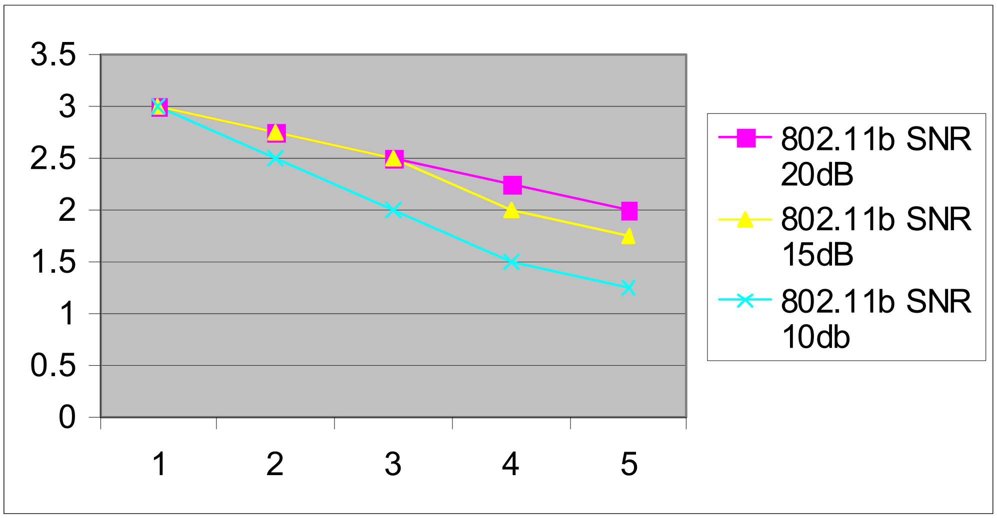


Figure 17: Wireless MOS Results

The experiment showed that 802.11b links could support VoIP traffic with very little quality loss. There was a drop in quality only when testing with a SNR of 10dB.

<u>Capacity Improvement of Wireless LAN VoIP using Distributed Transmission Scheduling.</u>

The below study [9] proposes a Medium Access Control (MAC) protocol that provides Quality of Service (QoS) in Voice over IP (VoIP) over Wireless Local Area Networks (WLANs) [9]. This MAC protocol minimizes packet collision and improves voice communication quality in both single-cell and multiple-cell cases. Introduction of the proposed protocol requires only VoIP STA software modification. There is no need of replacement of the Access Point (AP). The number of accommodated VoIP calls can be increased by approximately 50%.

There have been difficulties in QoS for real-time VoIP. Because of this, IEEE 802.11e adopted Enhanced Distributed Coordination Access (EDCA). EDCA defines four access categories (ACs) that provide packet prioritization for the delivery of traffic. Priority control on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is realized by setting the EDCA parameters to meet the QoS demands of application. However EDCA provides only the prioritized QoS for different ACs and does not take into consideration the QoS of traffics which belongs to the same AC. Hence it is difficult to meet the QoS demands of each application using only static EDCA parameters. Packet collisions increase as the number of simultaneous calls increase which results in voice quality degradation. The new proposed MAC protocol is based on a feasible and flexible scheduling method that takes advantage of the feature that VoIP packets are generated at intervals of voice codec period. Enhanced Distributed Coordination Access (EDCA) supports a prioritized mechanism on CSMA/CA. EDCA differentiates packets from an upper layer into four different ACs according to the applications and their QoS characteristics. A station (STA) with packets to transmit senses the medium before initiating a transmission. If the medium is sensed busy, the STA defers its transmission to a later time. If the medium is sensed idle for a pre-specified time, Arbitration Inter Frame Space

(AIFS), then the STA generates a random backoff period for an additional deferral time before initiating transmission.

AIFS = AIFSN * aSlotTime + SIFS

Where aSlotTime is the value of a slot time and SIFS is the value of a Short IFS. AIFSN is an integer value, which is set according to the AC. The higher prioritized AC has the smaller AIFSN value.

The random backoff period is derived as

BackoffTime = Random() * aSlotTime

Where Random() is a pseudo random integer drawn from a uniform distribution over the interval [0,CW]. CW is the connection window which is an integer in the range of values aCWmin <= CW <= aCWmax. The CW is set equal to aCWmin at the first transmission attempt and it increases in the following manner until it reaches aCWmax.

$$CW = 2^n * (CW_{min} + 1) – 1$$

Where n is the number of retransmissions.

The CWmin/CWmax values also vary for different ACs.

EDCA provides prioritized QoS for different ACs. However EDCA does not take into consideration the QoS of traffic, which belong to the same AC.

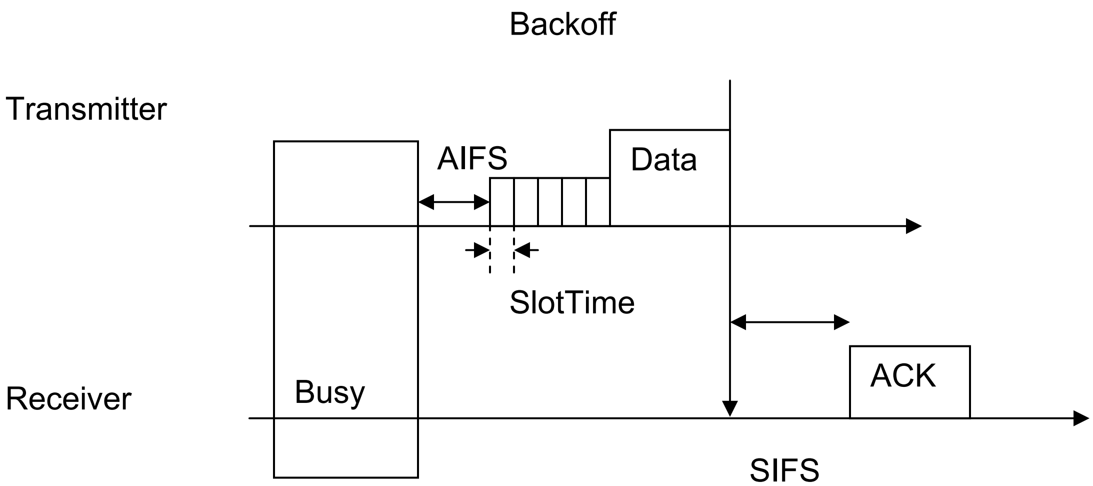


Figure 18: Access Schema in EDCA

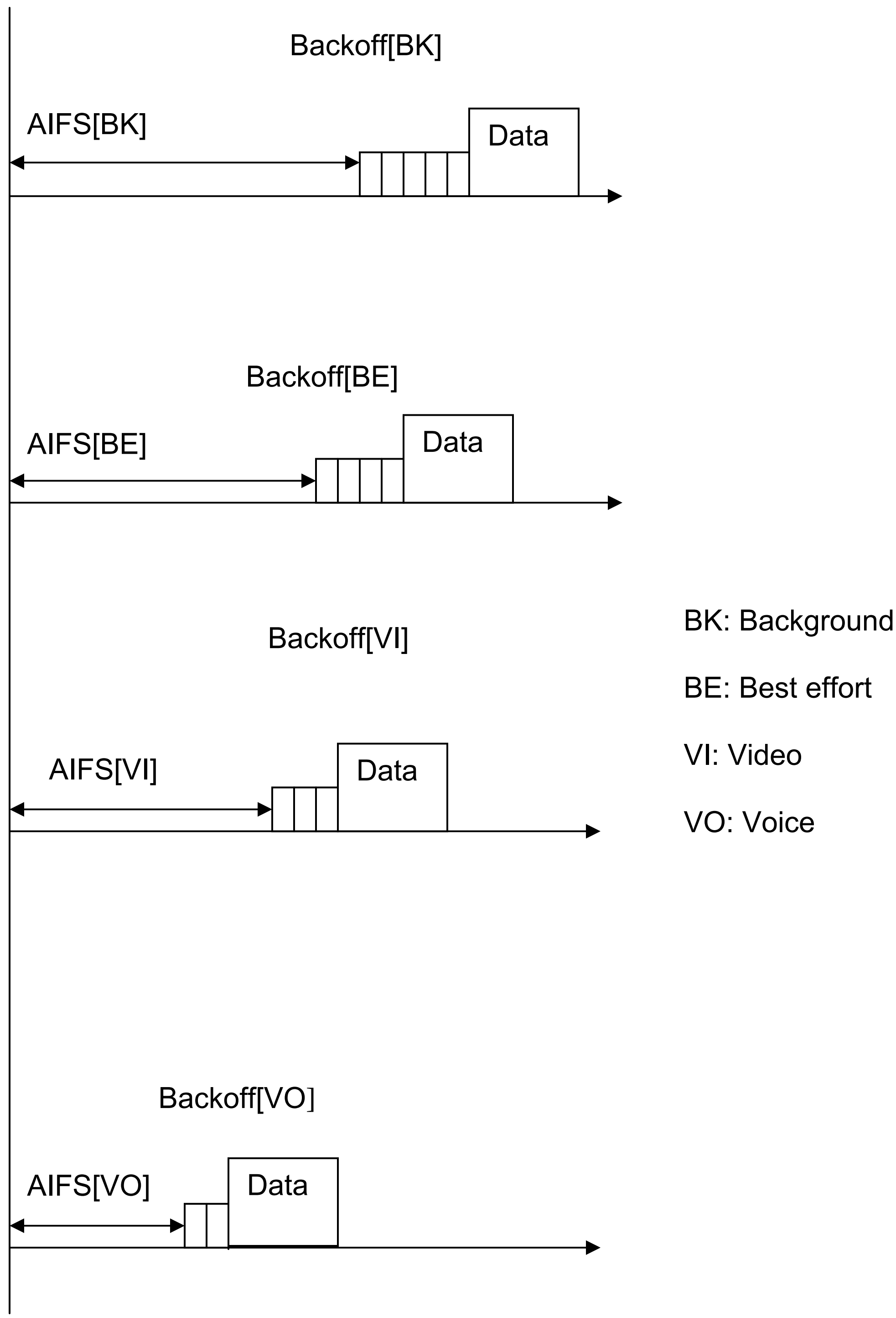


Figure 19: Prioritization in EDCA

The packet collision probability increases as the number of calls increases due to simultaneous transmission. Also, the delay tends to increase if more than one STA are in backoff states.

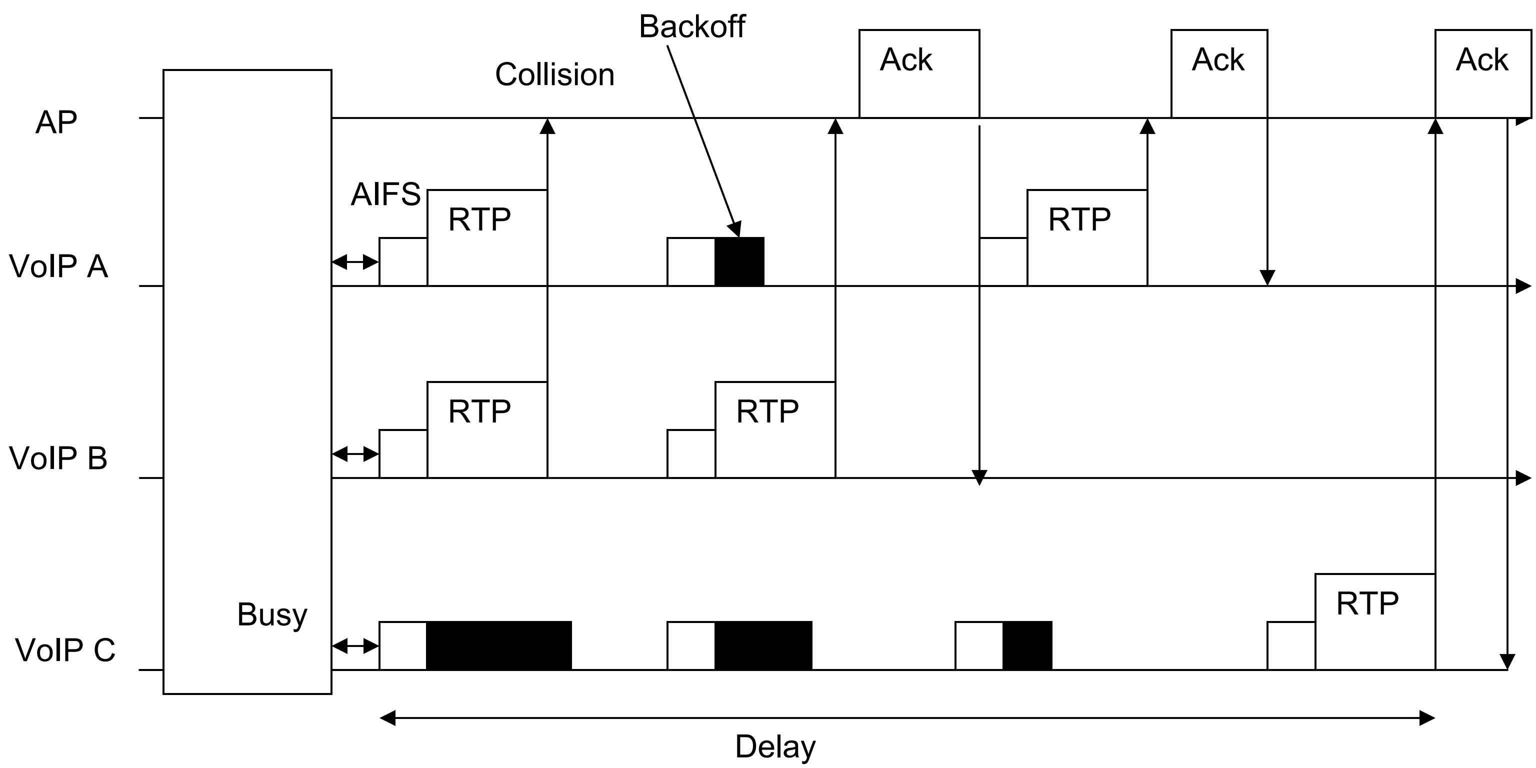


Figure 20: Simultaneous transmission in EDCA

A Novel AP for Improving the Performance of Wireless LANs Supporting VoIP.

Seamless handoff and balancing the traffic load in a wireless network can be achieved by using a Novel Access Point (AP) [10]. This is done by using a second transceiver to scan and find neighbouring STAs in the AP. The result is sent to the associate AP.

A mobile host (STA) that moves from one Access point to another while connected to the Internet has a delay time involved in the handoff process. Supporting Real-Time applications with continuous mobility implies that the total latency (layer 2 and layer 3) of the handoff must be small. Specifically, the overall latency should not exceed 50ms to prevent excessive delay and jitter. Unfortunately, the vast majority of Wi-Fi based networks do not currently meet this goal. Layer 2-related latencies contribute approximately 90% of the overall latency, which exceeds 100ms, while handoff-related latencies in layer 3 have an average of 15.37ms.

To accomplish this handoff latency time issue, a novel AP is used with two transceivers [10] that improves network efficiency in terms of reducing the latency time (layer 2 and layer 3) during the handoff process and supporting AP traffic load balancing.

The novel AP uses a second transceiver to scan and find neighbour STAs in the transmission range and later send the result to its associate AP. This information will be useful for the AP to control its associated STAs initiate the handoff process when a neighbour AP can provide higher quality and/or to sharing the traffic load with the neighbour APs. This does not require the customers to upgrade their wireless LAN devices. Instead, users are required to update their firmware in their devices in order to support the novel handoff scheme.

Handoff Procedure

In WLANs, the STA leaving its AP coverage area is required to initiate the handoff process for finding the next AP and establishing a link with that AP. The handoff is a function or process referring to the mechanism or sequence of messages exchanged by generally two APs and one STA, resulting in a transfer of physical layer connectivity and state information on the STA in consideration from one AP to another one.

In IEEE 802.11, the conventional handoff process is a kind of hard handoff. The process has been divided into three stages: 1) Detection phase, 2) Search phase, and 3) Execution phase.

1) Detection phase is a discovery of the need for the handoff; the STA detects a reason of frame loss among collision, radio signal fading, or the

STA being out of AP's transmission range in order to make a decision to start the handoff.

2) Search phase includes a set of actions performed by the STA to find information needed to perform the handoff. The IEEE 802.11 standard specifies two scanning nodes (Active and Passive mode) used by the STA to gather information of APs in range.
3) Execution phase is a two-step process. The STA sends a "reassociation request" frame to the intended new AP and later the new AP confirms the reassociation by sending back a response frame with a status value of "successful". The new AP needs to authenticate the STA before reassociation succeeds.

The STA assumes the current AP is always in range and tries to retransmit data for a period of time by reducing the bit rate if there is frame loss. If, after a period of time, the STA cannot get any response, the STA will transmit consecutively a certain number of RTS frames. If no CTS is received the reason of the frame loss is stated as: station is out of transmission range. The time elapsed until the reason of the frame loss is clarified is the delay time in this phase. After the reason of frame loss is proved, the STA will start the search phase by broadcasting a "probe request" frame in each channel and will wait for probe responses. APs in range will reply to the "probe request" frame in each channel and will wait for probe responses. APs in range will reply to the "probe request" frame by sending back "probe response” frames including their information. The STA receives the neighbour AP’s information, analyzes it, and decides to handoff by starting the execution phase. The delay time in the search phase corresponds to the amount of time necessary to send the probe request and then wait for the probe response from APs in range.

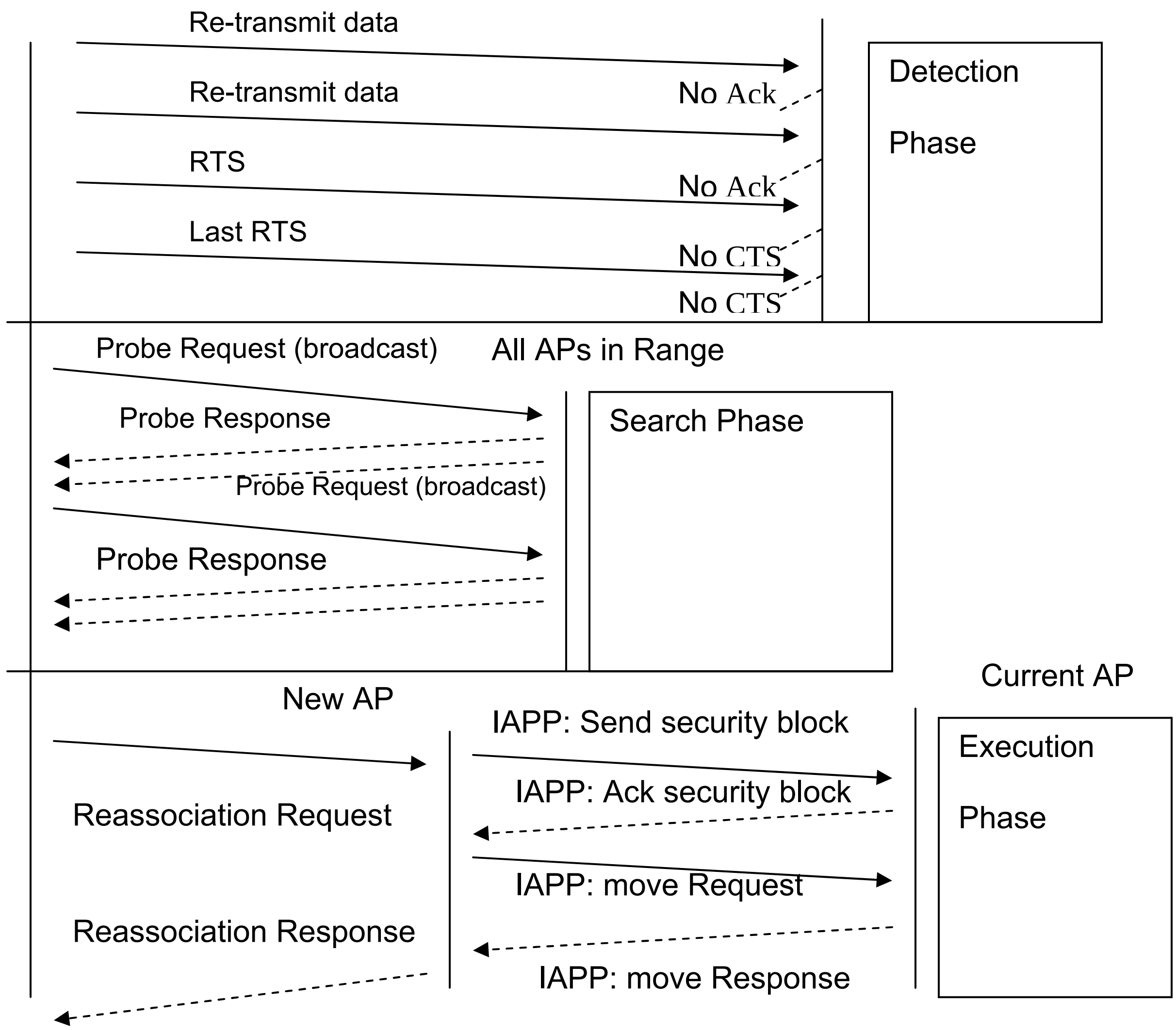


Figure 21: The Conventional Handoff Process

In the execution phase, the STA sends a re-association request frame to the new AP and later the new AP confirms the re-association by sending back a "re-association response" frame to the STA with a status of "successful".

	D-Link	Spectrum	Zoom Air	Orinoco
Detection	1630 ms	1292 ms	902 ms	1016 ms
Search	288 ms	98 ms	263 ms	87 ms
Execution	2 ms	3 ms	2 ms	1 ms
Total	1920 ms	1393 ms	1167 ms	1104 ms

Table 1: Handoff Times for Different WLAN Cards

According to the above table, the latency time of conventional handoff scheme does not meet the requirements of real-time applications such as VoIP (less than 50ms). The detection phase is the longest phase in all cases followed by search phase and then execution phase.

Proposed Handoff Procedure

The main difficulty in deciding whether to perform the handoff or not is to determine the reason for the frame loss among collision, the radio signal fading, or the STA being out of range.

Most WAN cards spend a lot of time in the detection phase to prove the AP is out of range before starting the search phase. In the search phase the STA also spends a long time broadcasting a “probe request” frame in each channel and waiting for probe responses from neighbour APs in order to find the information of neighbour APs. These high delay times prohibit the support of real-time services. By using the conventional handoff scheme, the STAs will not be able to start the handoff until there are frame losses because there is no information regarding neighbour APs and the handoff process is controlled by the STAs.

To improve the performance of wireless LANs and reduce the latency time in the handoff process, we introduce a novel AP with two transceivers. This AP will use the second transceiver to scan and find the information of neighbouring STAs in its range by using a fast passive scan mode. The fast passive scan mode is shown in the below figure. The first receiver works as a normal transceiver in the conventional scheme (sends and receives data).

The information on neighbour STAs received from a second transceiver will be sent back to its associated APs (the AP of those STAs), and then those APs will compare this information with that of other neighbouring APs and its own AP.

The AP controls the handoff process instead of the STA. The AP will send a “handoff request” message to its STAs to start the handoff to neighbour APs while providing better quality.

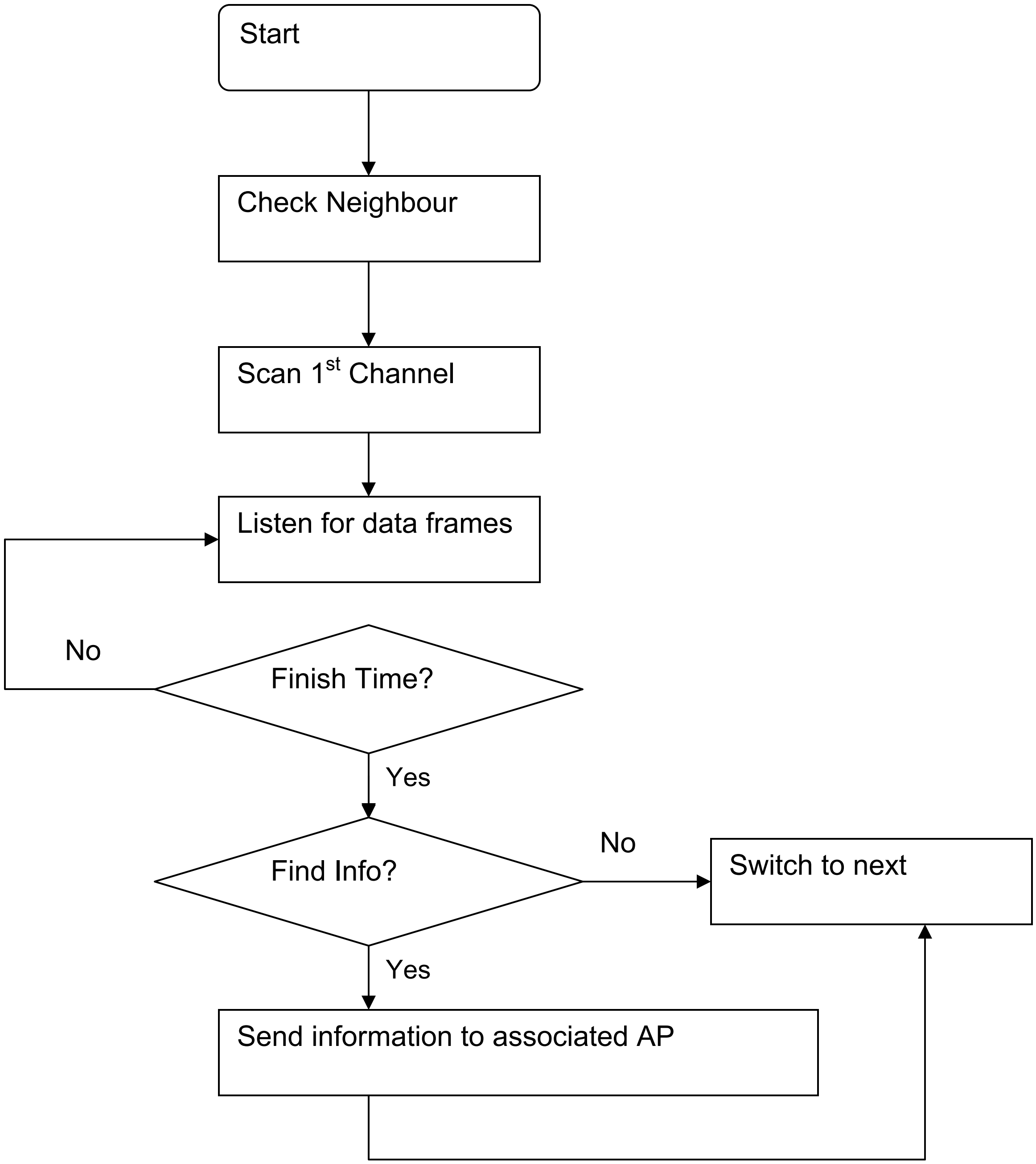


Figure 22: Fast Passive Scan Phase

The AP checks a list of channels that the APs in the network are currently using to scan nearby STAs in its range. The AP starts to scan from first channel in the list and continue to scan the next channel. It will not scan the channel that is currently used by its first transceiver. This scan phase will start by listening and waiting for data frames on each channel – a waiting time of 200 ms has been used based on the assumption that every STA is transmitting real time packets with a packet inter-arrival time smaller than 50ms. Upon receiving a data frame from any STA the AP analyzes it and then sends back the in-advance information including signal strength of the packet sent by the STA and its own traffic load conditions to the AP associated with the STA (the address of the AP associated with the STA can be directly extracted from the header of the data frame sent by the STA). This procedure is carried out separately for every channel.

In the proposed handoff scheme, the handoff process will be controlled by the network instead of the STAs. The APs will compare not only the signal strength, but also analyze the traffic load condition of neighbouring APs.

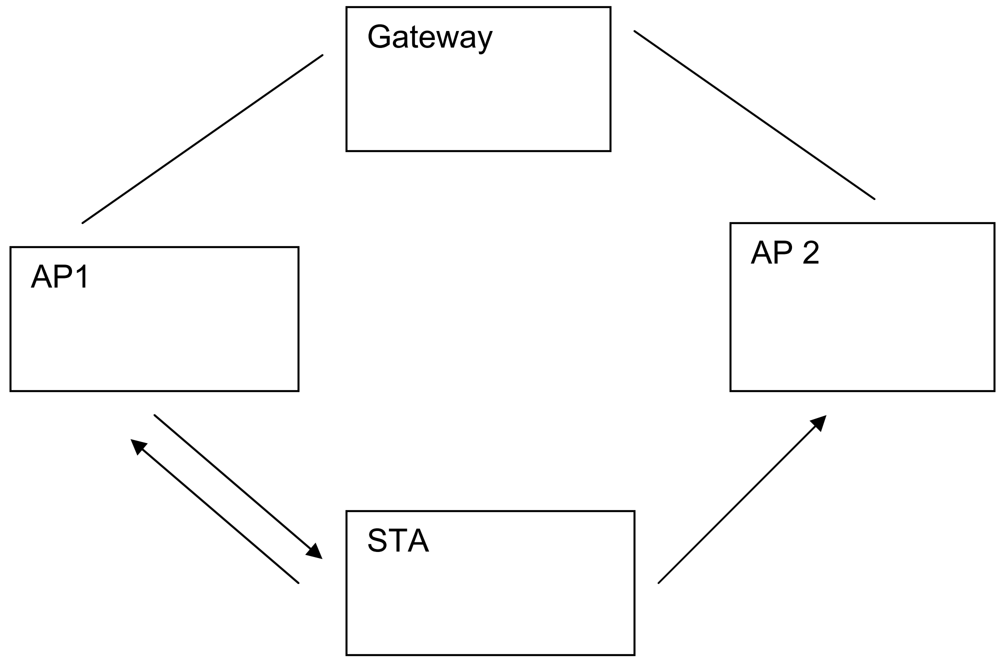


Figure 23: Handoff with New AP Module

The above figure shows a simplified model of our proposed scheme. AP2 will use the second transceiver to scan and find information of neighbouring STAs in its transmission range by using fast passive scan mode. AP2 will receive a data frame from the STA (by using the second transceiver) when the STA sends a data frame to AP1. AP2 analyzes this information and then sends back in-advance information to the associated AP (AP1) via a gateway (in-advance information includes the received signal strength of the STA and the traffic load condition of AP2). AP1 will analyze and compare this information with its own data, and could eventually decide to inform the STA to perform the handoff process. (The novel handoff scheme will use the Hysteresis factor to compare the signal strength between AP1 and AP2 in order to avoid the ping-pong effect).

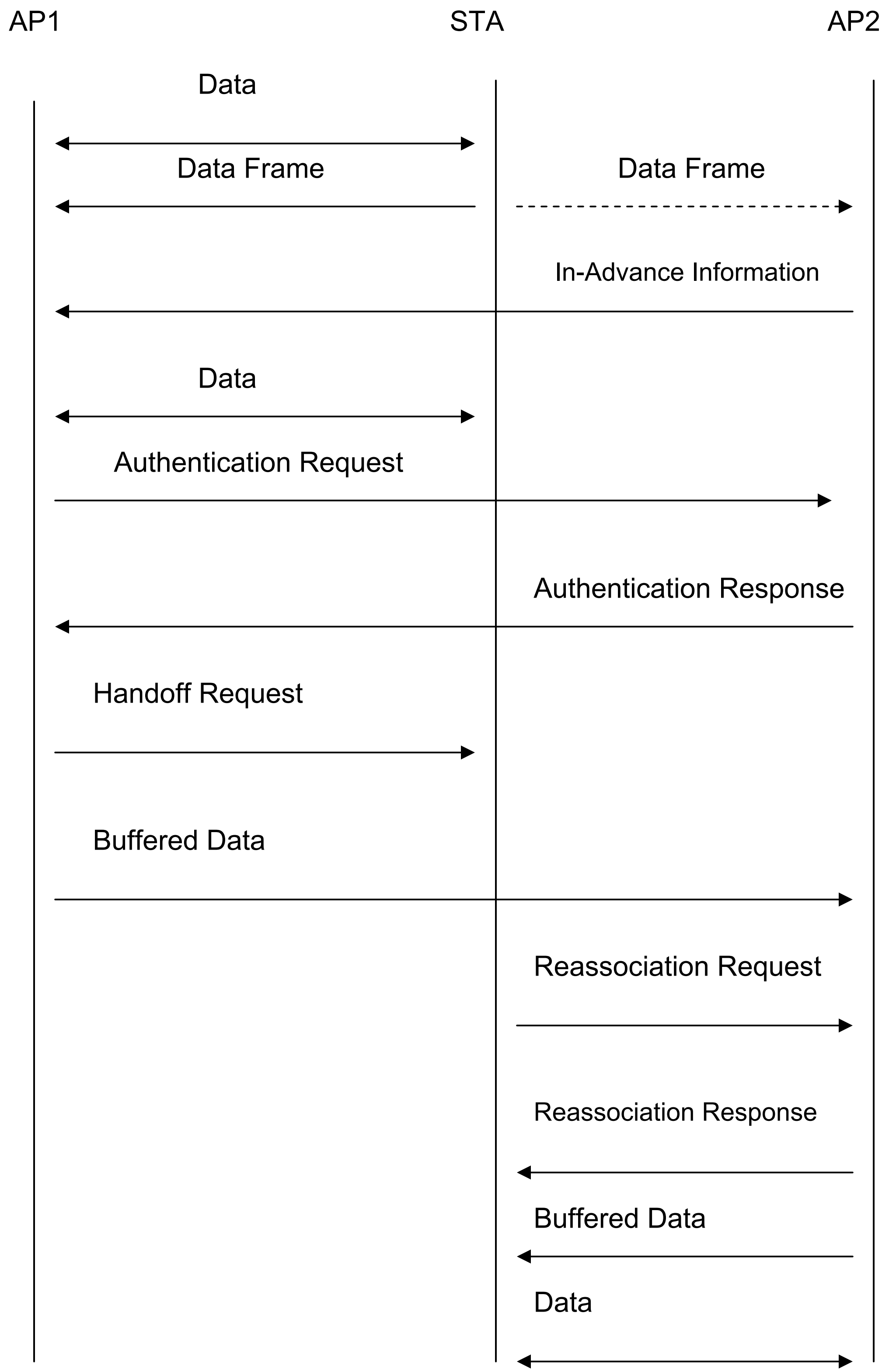


Figure 24: Handoff Process

Regardless the process to inform the STA, AP1 will first perform the pre-authentication process with AP2. AP1 sends an “Authentication Request” message to AP2. AP2 receives the message and performs an authentication procedure for the STA and sends back an “Authentication response” message to AP1. This message contains several pieces of information including the session key and WEP key. AP1 receives the response message and then sends a “handoff request” message, including the information of AP2 to the STA. At the same time, AP1 will send all buffered data (for the STA) to AP2. The STA starts the reassociation process to the new AP (AP2) by sending a “Reassociation Request” message to AP2 and waits for a reassociation response. Finally, AP2 replies with a “Reassociation Response” message to the STA and starts sending to it all buffered data previously sent by AP1.

The handoff algorithm is shown below. The Algorithm is used by an AP to analyze the data sent by neighbouring APs and compares it with its own data.

	From	Signal Power	Traffic Load
STA1	AP1	-87 dBm	42%
STA1	AP2	-73 dBm	38%
STA1	AP3	-98 dBm	43%
STA2	AP1	-56 dBm	42%
STA2	AP2	N/A	24%
…	…		…
STAn	Apn	X dBm	y%

Table 2: The example of Data

The above table shows an example of operation data. The data on traffic load at every AP allows us to balance the traffic load among neighbour APs by controlling the number of STAs in each cell. This is done to assure a higher Quality of Service (QoS) for real-time services.

Xc: Signal strength measured by current AP

Xn: Signal strength measured by neighbour AP

Yc: Current APs traffic load

H: Hystersis

HOFc: Handoff factor of Current AP

HOFn: Handoff factor of Neighbour AP

HOFh: Handoff factor Hystersis

If Xc<Xn and Xn>Xth, then

start handoff

end if

if Xn>(Xth+H) and HOFn>(HOFc+HOFh) then

start handoff

end if

Table 3: Analyze and compare Data

In the above algorithm, we first calculate the strength of the packet sent by the STA in current AP (Xc) and the neighbouring AP (Xn) by using weighted average as shown by:

$$X_c = \frac{\sum_{i=1}^{n} w_i \, Xc_i}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad X_n = \frac{\sum_{i=1}^{n} w_i \, Xn_i}{\sum_{i=1}^{n} w_i}$$

$$\text{where } w_i = \frac{1 + (n-i)}{n+i} \quad \text{and } n = 4;$$

The weighted average shown above is intended to smooth the variable conditions of the signal strength along time. At the same time we calculate the traffic load of the current AP (Yc) and neighbouring APs (Yn) using:

$$Y_c = \frac{\sum_{i=1}^{n} w_i Yc_i}{\sum_{i=1}^{n} w_i} \quad \text{and } Y_n = \frac{\sum_{i=1}^{n} w_i Yn_i}{\sum_{i=1}^{n} w_i}$$

$$\text{where } w_i = \frac{1 + (n-i)}{n+i} \quad \text{and } n = 4;$$

In order to compare the current AP's and neighbour AP's capabilities, a parameter named as Handoff Factor (HOF) has been defined. The HOF allows two APs to be compared and takes into account both the signal strength and traffic load conditions.

$$HOF_c = \frac{X_c - X_{TH}}{X_{TH}} + \frac{Tr_{MAX} - Tr_c}{Tr_{MAX}}$$

$$HOF_N = \frac{X_N - X_{TH}}{X_{TH}} + \frac{Tr_{MAX} - Tr_N}{Tr_{MAX}}$$

In the case when the signal strength of the packet sent by the STA is lower than the threshold in the current AP, the STA will start the handoff process if and only if the signal strength measured at the neighbour AP is higher than a simple sum of threshold and hystersis, and the HOF of the neighbour AP (HOF_N) is higher than the sum of the HOF of the current AP (HOFc) and the HOF of the hystersis (HOF_N).

The novel AP with two transceivers can reduce the latency time in the handoff process to support real-time service in a wireless network such as Voice-over-IP (VoIP) and E-Conference. The STA can switch to a new AP with higher quality (comparing both the signal strength and traffic load conditions).

The handoff process is controlled by the AP, which makes it easy to manage the wireless network in terms of traffic load sharing (traffic load balance among neighbour APs). This feature can reduce the number of packets dropped due to traffic overload in one cell improving the performance of the overall network.

This handoff scheme is compatible with all commercial wireless LAN cards. Customers only need to update the firmware which can be downloaded from the vendor's website without changing or adding the hardware components.

The disadvantage of the handoff scheme is the two radio interfaces that every AP in the systems should have.

Performance Evaluation

Two simulation models were built to observe the throughput [10] and end-to-end delay time of the STA and also the traffic load of the network. The OPNET Modeler program was used to implement the simulations and compare results.

VoIP traffic based on G.711 codec standard (the simplest voice codec standard) was used so that a VoIP packet was generated every 20ms, with 160-byte data, 12-byte RTP header, 8-byte UDP header, and 20-byte IP header. The VoIP packet size at the IEEE 802.11 MAC layer became 200 bytes per packet and the data rate was 80kbps.

The below figure shows the simulation environment.

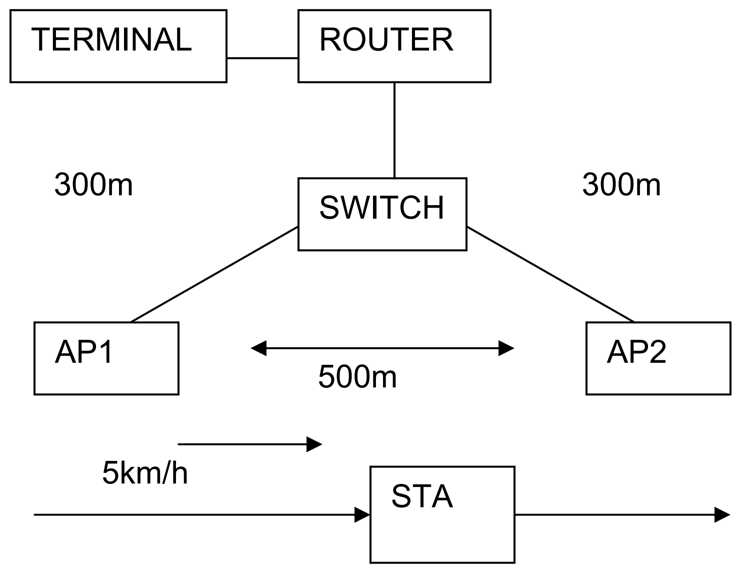


Figure 25: The Simulation Environment

Two APs with a cell radius of 300m and a transmission power of 200mW were connected to each other via a gateway. The distance between the centres of two neighbouring APs was 500m. The simulations consider a STA moving from AP1 to AP2 at a walking speed (5Km/h) using the conventional scheme, and later using the proposed scheme to compare the results. In the conventional handoff scheme the delay was between 1700 and 1900ms. In the proposed handoff scheme the delay was between 3 and 5ms. The proposed handoff scheme was so low because there was no delay time associated with the detection phase and search phase and the delay time in the execution phase was also reduced. It was also observed that there was a small latency time associated with the throughput drop during the handoff process in the proposed handoff scheme as compared to the conventional handoff scheme.

The AP2 receives the data frame from STA and sends the in-advance information back to AP1. AP1 compares that information with its own data through calculation of the respective HOFs and finally orders its own STA to start the handoff to AP2 because in this case AP2 can provide better quality than AP1.

The tests showed that the proposed handoff scheme did not effect the end-to-end packet delay time of the STA during a normal situation.

The below figure shows the simulation environment for traffic load balancing. A group of ten STAs were moving in the overlapping area of three neighbouring APs. Each AP had its own associated STAs moving freely in its cell. The proposed scheme was compared to the conventional scheme. The initial results [10] showed that the proposed handoff scheme provided a good level of traffic load balancing among APs. Every AP controls the handoff process of its own associated STAs to balance the traffic load of the overall network. The conventional handoff did not provide any load balancing since STA initiated handoff mainly depends on the signal strength measured at the STA. Also, the total number of dropped packets reduced using the proposed scheme. This was due to the fact that when the traffic load increased in any AP, it compared the traffic load with its neighbour APs and eventually commanded its own STA to handoff to a neighbour AP that was able to provide a higher quality (meaning lower traffic load in comparison with the current AP). In contrast the STA in the conventional handoff scheme initiated a handoff to a neighbouring AP basically based on the received signal strength, irrespective of the traffic load of its associated AP.

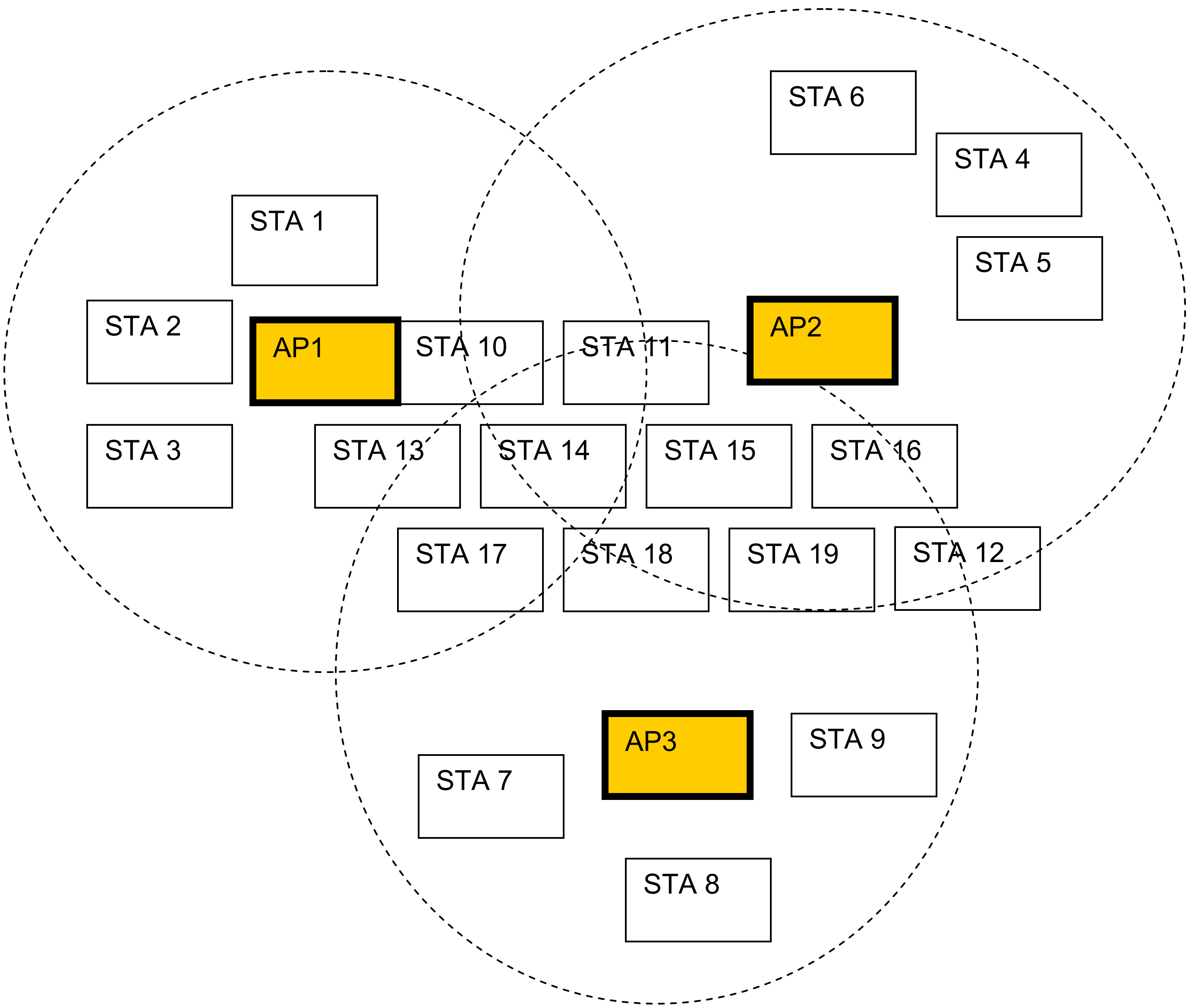


Figure 26: Network Architecture

Performance Optimizations for Deploying VoIP Services in Mesh Networks.

Wireless VoIP can provide the caller the convenience of using portable phones with limited roaming [11]. Providing the VoIP users with true mobile phone services while having the freedom of roaming requires wide area wireless coverage. IEEE 802.11-based multi-hop wireless mesh networks have been considered as a practical and inexpensive solution for providing such wide area coverage. The benefits of mesh network compared to wired LAN connecting WiFi access points are: 1) ease of deployment and expansion; 2) better and wider coverage; 3) resilience to node failure; 4) reduced cost of maintenance.

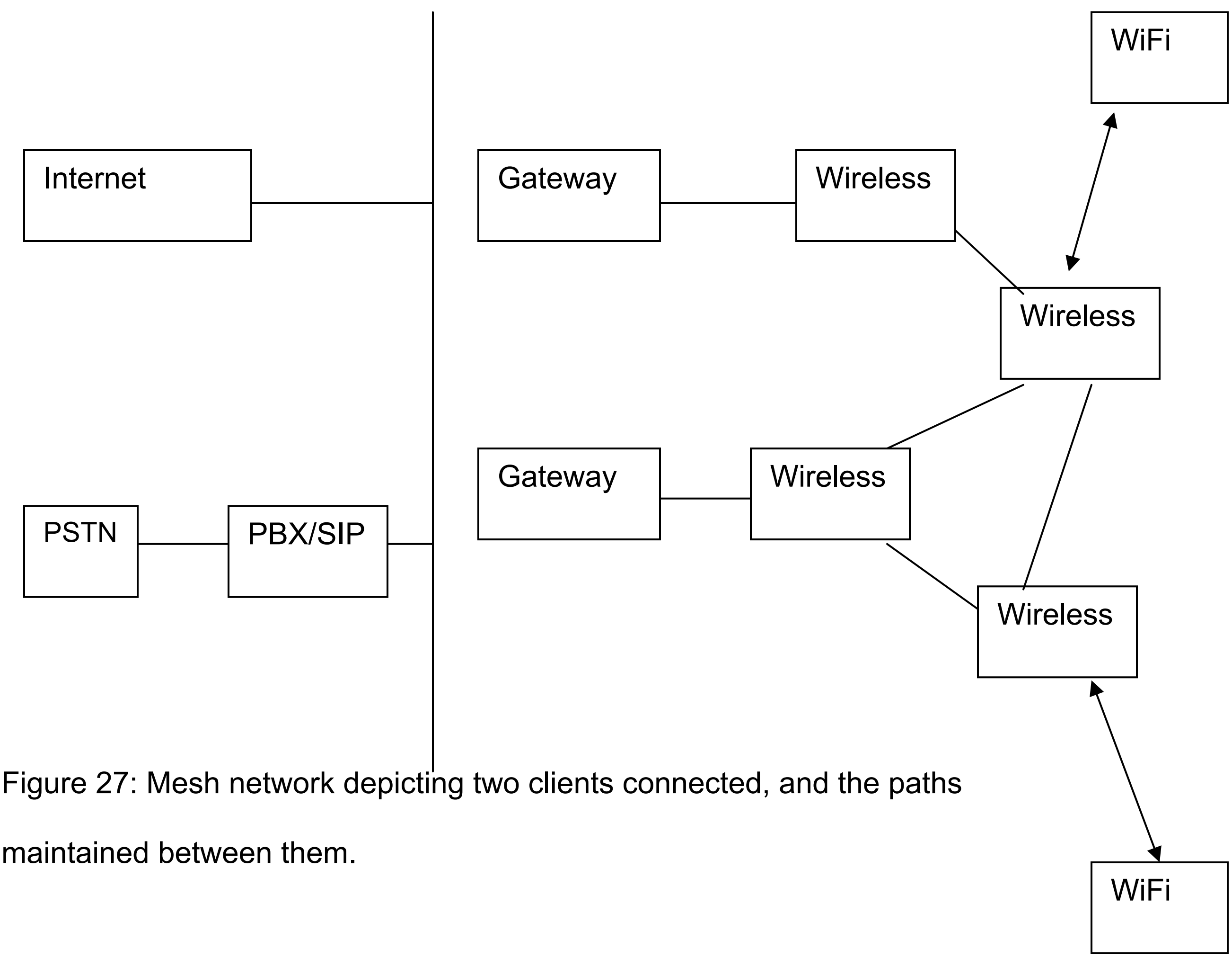


Figure 27: Mesh network depicting two clients connected, and the paths maintained between them.

One popular video encoder is G.729, which sends 50 packets per second of 20 bytes each. It is used by some available 802.11 VoIP phones (such as Zyxel Prestige). It is emulated in a test-bed using CBR UDP traffic. Quality is defined by the R-factor, which should provide a value above 70, for medium quality:

$$R = 94.2 - 0.024d - 0.11(d - 177.3)\,H(d - 177.3) - 11 - 40\log(1 + 10e)$$

Where:

1) $d = 25 + d_{jitter\_buffer} + d_{network}$

is the total ear to mouth delay comprising 25ms vocoder delay, delay in jitter buffer and network delay

2) $e = e_{network} + (1 - e_{network})\, e_{jitter}$

is the total loss including network and jitter losses

3) H(x) = 1 if x >0 otherwise is the Heavyside function

4) The parameters used are specific to the G729a encoder with uniformly distributed loss

The constants consider the delay introduced by the encoder for its look-ahead buffer, and the delay introduced by the jitter buffer. A jitter buffer of 60ms is considered. This has two contradictory effects. It increases the end-to-end delay, which reduces the quality but reduces the jitter, which is an overall better effect. The R score is computed from the loss and delay in the network. To emulate the behaviour of the jitter buffer, the play-out is assumed to start at the destination when packet 4 arrives (60ms jitter buffer = 3 packets).

All the deadlines for the packets at the receiving side are established at this point. Loss in the jitter buffer is computed as the fraction of packets, which do not meet their deadlines. To compute loss probabilities and average delay in the network, all packets from all flows in an experiment are considered together by macro-averaging.

The mesh in this experiment is a multi-hop extension of the access point infrastructure. A port is a mesh node, which has two interfaces: one in ad hoc mode for the backhaul in the mesh and one in infrastructure mode to connect to clients. The clients can be VoIP wireless phones, soft-phones running on laptops or hand-helds. These clients see the mesh as a switch or hub in the sense that they are not concerned with the internal routing of the mesh. The client stations are unaware of the mesh networking backbone. They view the network as a conventional wireless LAN spread out over an extended geographic area. Thus the clients still associate with an AP using a traditional association mechanism in wireless LANs. When the client moves and re-associates with a different AP, a layer-2 handoff event occurs that in turn triggers appropriate routing updates in the mesh network backbone. The handoff process involves both layer-2 and layer-3 procedures.

Test-bed Hardware and Software configuration

The test-bed [11] consists of 14 nodes based on the Routerboard 200 series processor boards with 256MB of RAM and 512MB of compact flash. Each node is equipped with two 802.11b wireless interfaces and has an open slot for a third one (PCMCIA 16 bit). The wireless cards are running at the fixed rate of 2Mbps. Each node operates with two interfaces, one that is used to get client traffic from the VoIP 802.11 phones and the other one for backhaul in the mesh.

Mesh node

A click router [11] is used on each mesh node. When a packet is received from the cards, it may get labelled if it a voice packet and if it needs to be routed over mesh network. To support VoIP traffic, packet aggregation service is implemented which encapsulates multiple small VoIP packets into a single large packet and forwards it. For each interface there is an aggregator and a de-aggregator to de-capsulate the aggregated packets into the original VoIP packets. The most important component in the mesh node is the classifier that decides a) whether a packet is destined for a local machine (signalling, aggregated packets); b) has to be routed (best effort, signalling); c) switched (voice). For aggregated packets, after the de-capsulation, the resulting packets are fed back to the classifier.

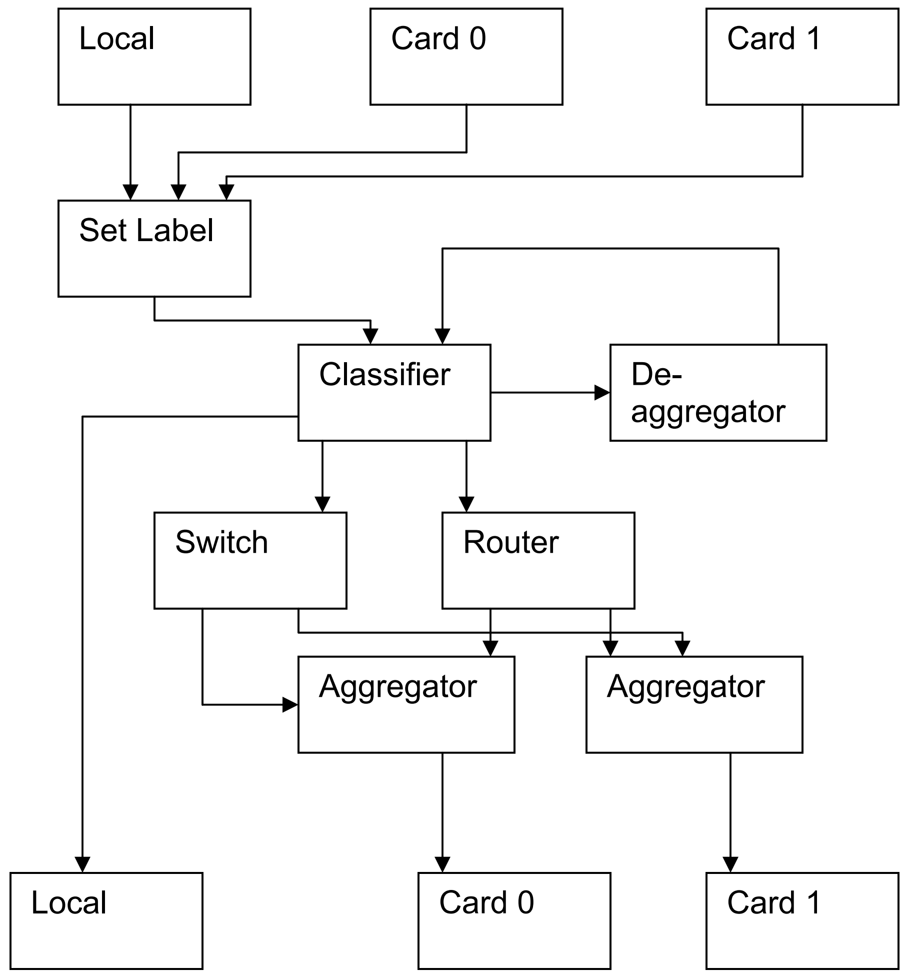


Figure 28: Click node component for label switching, routing and aggregation

Label based forwarding

For labelling the IP packets, the TOS field of each IP packet provides 255 labels at each node. For packets with label other than zero, label based forwarding is used. On the other hand, packets with label zero follows underlying routing protocol (DSDV) used in the mesh network. To perform label based forwarding, each node maintains an addition table. An entry in this table has the form: (in_label, out_label, interface, gateway).

Any packet that arrives on an interface and has a non zero label (in.label) is stamped with a corresponding out.label and sent to the interface to be delivered to the gateway which is the next hop in the path. Once the outgoing label is set (for the switching), and the outgoing interface is determined (for both routing and switching, by their respective lookups), packets are pushed to “pull” queues associated with each interface. These queues perform the aggregation of packets that have the same next hop (only voice and probe packets). The meaning of “pull” in Click terminology is that these queues are queried by the cards when the transmission is possible, so the waiting time in these queues is used for the purpose of the aggregation.

VoIP Packet Aggregation

As most vocoders use samples of 10-100ms, a mesh node is expected to get a large volume of small packet traffic. 802.11 networks incur a high overhead to transfer one packet, therefore small sizes of packets reduce the network utilization.

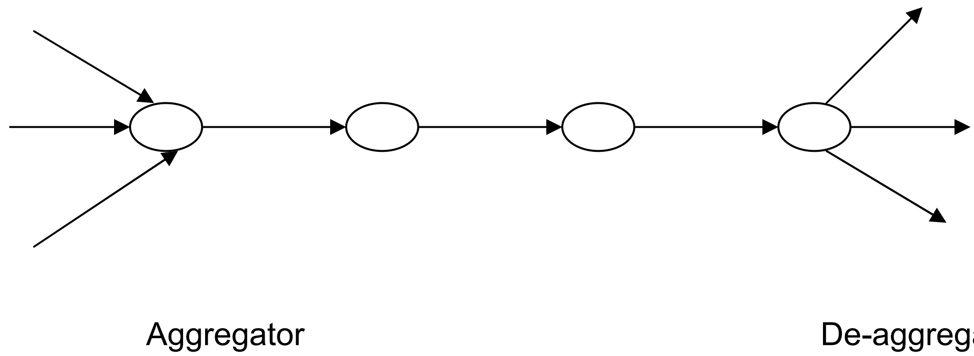


Figure 29: Aggregation merges small voice packets from different calls into larger packets to improve channel utilization

Aggregation combines together several small packets that arrive at the ingress node and forward them with one IP, MAC and PHY header across the air.

The problem with small payloads is that 802.11 MAC spends most of the time sending headers and acknowledgments, waiting for separations DIFS and SIFS and contending for the medium. Packet aggregation can significantly reduce the overhead and increase the number of supported calls. A common problem in packet aggregation is that it increases packet delay, which can possibly reduce its suitability for delay sensitive VoIP services.

In a mesh network scenario, one can consider two naïve mechanisms for aggregation: end-to-end and hop-by-hop. In the end-to-end scheme, the aggregation is done only at the ingress node for all flows routed for a common destination. A forced delay is added at the ingress to perform packet aggregation. However, its applicability is limited to availability of enough packets for aggregation under the forced delay budget. In the hop-by-hop scheme, packets are aggregated and de-aggregated at every hop by adding a forced delay at every hop. This scheme is oblivious to source or final destination of a given packet and can result in adding cumulative delay to all packets over multiple hops. The hop-by-hop scheme also suffers from extra complexity at each node for executing aggregation/de-aggregation on every incoming packet.

The accretion aggregation algorithm is a hybrid of the above two naïve algorithms. Forced delay is introduced at ingress nodes. In addition, at intermediate node, medium access queuing delay is used for possible aggregation resulting in no extra delay on any packet. At the ingress, the forced delay depends on the budget allowed by the probing of available paths. At forwarding node, the aggregation becomes more effective in high load, the case where the need becomes natural for saving bandwidth.

P-packet being queued at a node;

A-aggregation packet being prepared;

MinPackets-number of packets from the same flow that have to be aggregated at the ingress

MTU-maximum transmission unit, in voice packets

```
1: Find queue of P;
If size(queue) > MinPackets
      Add all packets from flow (P);
      If size(A) < MTU
            Find a queue with the same destination
            Go to 1;
      Else
            Send A directly to destination;
Else
      If size(A) < MTU
                  Add packets w same next hop as p;
         Else if aggregation timer is expired
                  Aggregate all packets from the queue
                     of which timer is expired also;
```

Figure 30: Algorithm showing Aggregation logic for ingress node

Evaluation of VoIP Packet Aggregation

A string of 6 nodes with and without aggregation [11] was simulated. The results were verified against a similar string in the test-bed. The non-aggregated traffic simulation matched the test-bed results in most points, but for aggregated traffic, the test-bed performed worse. The cause of this was identified in the fact that the capacity of some of the hops in 2Mbps mode was less than the optimal 1.7Mbps. The first hop was measured to provide a capacity of only 1.38Mbps.

One of the main claims of the accretion aggregation method was that it did not introduce additional delay by using the wait for the MAC availability to club together packets destined to the same next hop. To verify this claim, five longer calls indicated by A in the figure below were placed. Five short calls indicated by B below were placed. Aggregation was performed for each group of flows independently at the source by introducing a controlled delay of 80ms.

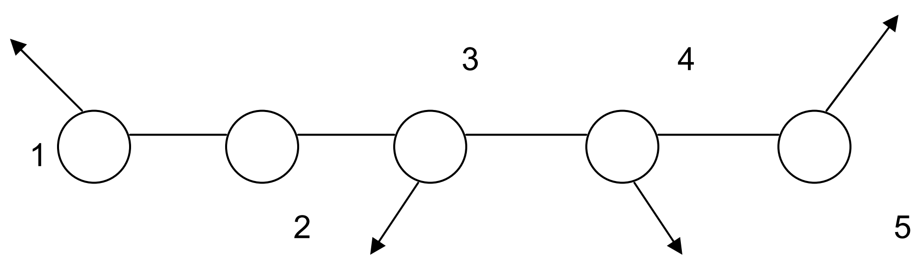


Figure 31: Aggregation introduces only controlled delay at the source of flows.

Intermediate nodes did not delay packets to improve aggregation, but use "natural" waiting required by MAC under load.

In the below picture we see that aggregation more than doubles the capacity if we consider an R-value of 70 as the threshold.

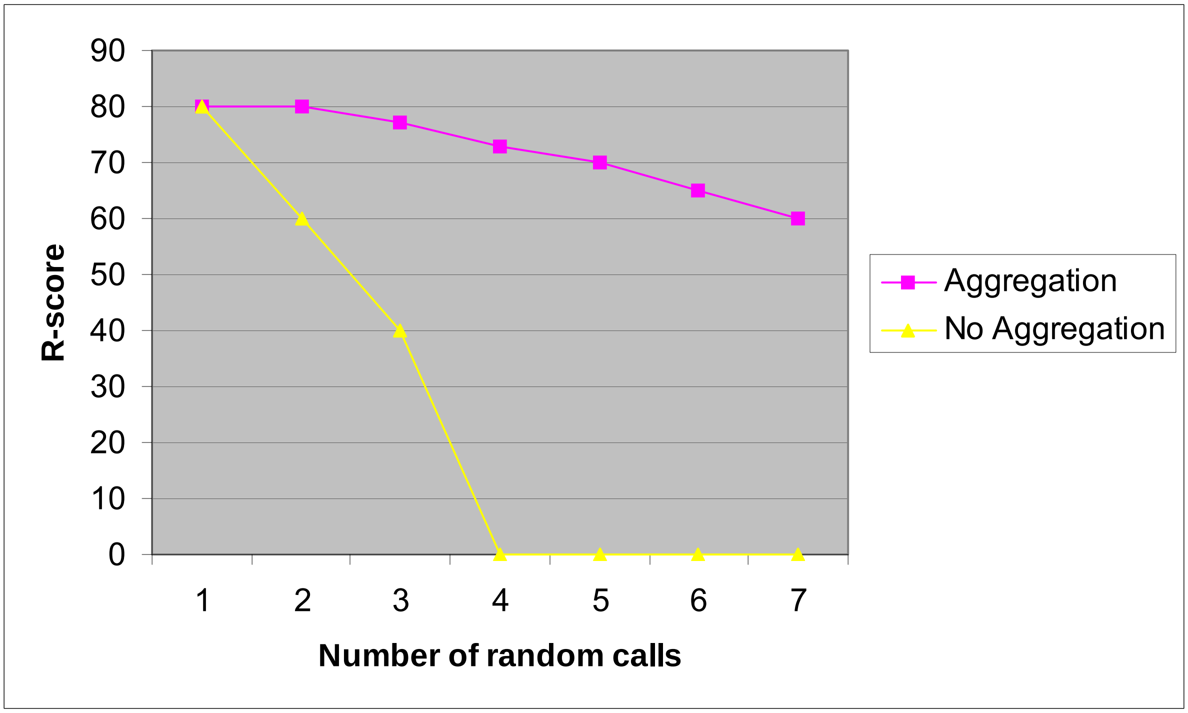


Figure 32: Aggregation performance for random calls on a string topology

Packet A is not merged with packets B during their common hops 3-4. The network time for flows A is about 61ms in the absence of B and increases to 89ms after B is added. This is normal, considering the increased interference for all the nodes and the longer queues at nodes 3 and 4. After we enable aggregation between A and B at hop 3-4 the network time for flows A decreases to 79ms. Not only the delay is not added to the long flow, but the creation of larger packets reduces the contention for the hop 3-4 thus reducing the load on the network.

1. The aggregation scheme only kicks in at higher load when waiting in queues can be used to group packets with the same next hop.

2. It is completely distributed in the mesh. The endpoints need to specify initial delays depending on their time budgets, but the intermediate nodes have the simpler task of looking for packets with the same next hop.
3. Short flows do not delay long flows for the purpose of aggregation. Short flows increases load and therefore causes delay for long flow, which is the inherent behaviour of the shared medium. In the network, distributed aggregation is a method that reduces load without impacting the network time of flows.

When mesh network is used as an extension of the access point infrastructure, a popular pattern is to have voice calls forwarded to the wired infrastructure or to the PSTN. In this case paths from the clients all lead to a common root, which provides access to the wired leg of VoIP. A tree with 15 nodes with a maximum 4 hops from root to leaf node was simulated. On this tree, it was found that the maximum number of supportable calls is 16, 16, 24 calls in End-to-End, Hop-by-Hop and Accretation algorithm.

The CDF on the R-score shows that proposed accretion mechanisms results in higher probability to meet the required R-score and lower packet loss. The number of packets maintaining near to zero jitter is higher for the proposed algorithm.

The results demonstrate that

a) End-to-End algorithm makes poor use of bandwidth with fixed aggregation packet size.
b) Hop-by-hop algorithm undergoes large packet jitter which increases jitter drop ratio.

Header Compression

Header compression is a complimentary scheme related to aggregation. The usage for header compression is motivated by the fact that

a) the VoIP payload is typically compressed at the application layer, which means another compression does not help reduce the payload size,
b) the headers occupies a large portion of the packet,
c) the headers have significant redundancy.

Packet headers with redundancy may be reduced through compression techniques as has been done with great success for cRTP or ROHC. The compressor and de-compressor should operate in synchronization, either implicitly (optimistic mode) or explicitly (acknowledged mode).

However, the 802.11 wireless media shows high error rate, the links fluctuate, the paths may change frequently between two nodes. If the compressor and de-compressor are out of synch, they need to exchange the information to recover their inconsistency. Such recovery requires time, complexity and bandwidth and can lead to data loss.

For a VoIP flow RTP/UDP/IP headers take 40 bytes, but only 12 bytes of them change when the packets get routed. CRTP or ROHC aim at compressing the 40 bytes into a 2 byte connection ID, but they are appropriate for one link only.

To emulate a simpler scheme that only transmits the changing fields we reduce the header from 40 to 14 bytes so that VoIP protocol overhead coming from large header size can be relieved.

A compression logic is proposed [11] called the zero-length header compression (ZHC) algorithm for redundant header elimination algorithm that leverages the VoIP packet aggregation mechanism. ZHC does not require context synchronization between two nodes.

The compressed headers from the compressor are needed to restore the original packet headers index at the de-compressor. The de-compressor has enough information to restore all headers from the first header.

TCP Window Control for Variable Bandwidth in Wireless Cellular Environment

The bandwidth of the bottleneck links, do not remain constant in wireless networks. When there are limited resources, the bandwidth is controlled by the radio condition over time. The varying bandwidth can cause network congestion or underutilization. A new window control algorithm can be used to improve TCP performance in wireless cellular networks with variable bandwidth.

A base station can have a bottleneck due to the difference of network bandwidth between wired and wireless links. If the bandwidth of a base station is C. C is a piecewise constant with jumps where the duration between two consecutive jumps is larger than the settling time of the system [12].

If M is the number of active flows routed through a base station and a sender updates its window size at the end of every round trip time (RTT). The Eifel algorithm [12] is used to detect timeout due to abrupt bandwidth decrease. The timestamp of acknowledgement (ACK) packet is compared with the timestamp of the retransmitted data packet to detect a timeout. The timestamp is also used to calculate the RTT at the receiver. The sender writes the timestamp into the header of each data packet and the receiver echoes the timestamp in the corresponding ACKs. The receiver can estimate RTT by comparing the timestamp of the ACK with the timestamp of the corresponding data packet.

The available bandwidth of the bottleneck link is estimated using the inter-arrival gaps of returning ACKs. This measurement is done both at the sender and the receiver because in wireless cellular networks the bandwidth of the downlink at the base station is much higher than that of a mobile user.

The inter-arrival gaps are taken at the receiver as follows:

$ABW_{ik} = L/D_{ik}$

Where $D_{ik} = t_{ik} - t_{i(k-1)}$, t_{ik} is the arriving time of the kth data packet of flow i at a receiver, and L is the packet size.

The above can be smoothed out by defining the average available bandwidth:

$$ABW_i = \frac{L}{\sum_{2}^{K} D_{ik} / (K-1)}$$

Where K is the total number of arriving packets at the receiver during RTT. This results in an analysis which is less affected by reverse link conditions and also able to detect spurious timeout.

To avoid network congestion, the queue length is small. A feedback value FV = $\alpha (q_{th} - q)$ where α is a positive control parameter, q_{th} is the queue threshold and q is a queue length of the base station. The feedback value is computed every predefined value, Ts where Ts is less than RTT of each flow. However, there is a heavy overhead for transmitting the feedback value which results in trade-off in choosing Ts.

The advertised window AWND is set every RTT as below:

AWND = (Average ABW_i + FV) * Ti

Where Ti is the RTT of flow i. This allows a guaranteed fairness among flows with different RTTs.

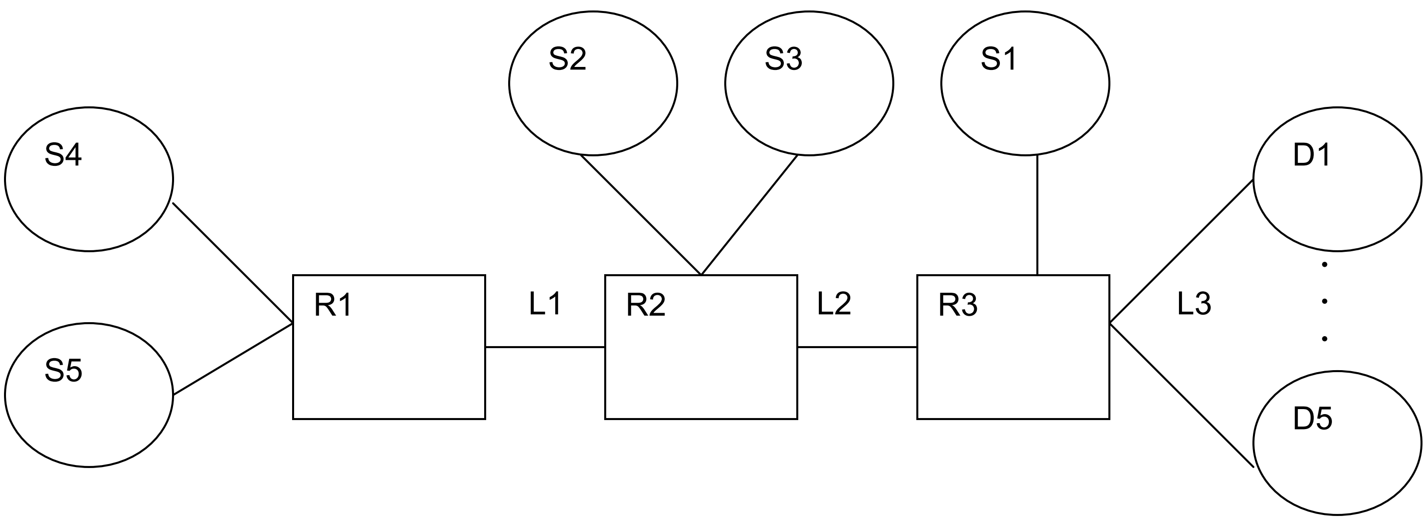


Figure 33: Network configuration

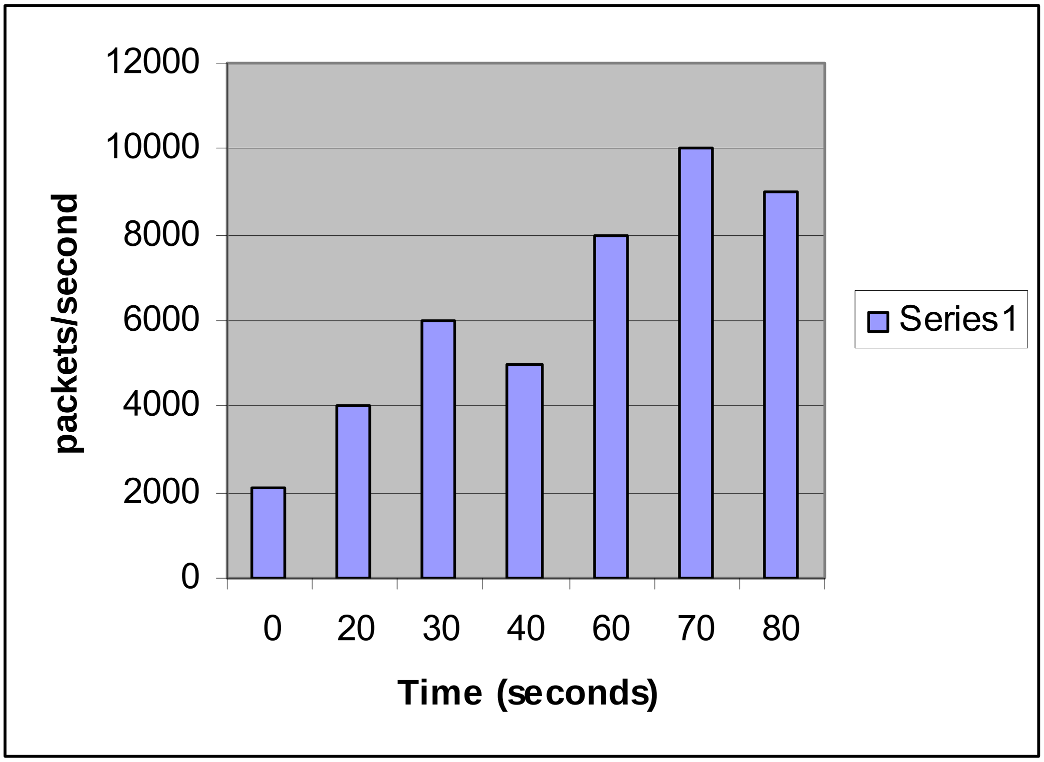


Figure 34: Bandwidth of link 3 (L3)

The TCP New Jersey is compared with the proposed configuration. Figure 33 shows three routers connected via two (L1, L2) wired networks and one wireless link L3. L1 and L2 have constant bandwidths of 10,000 packets per second. L3 has a variable bandwidth as described in the above figure. There are five flows with different RTTs in the network. Link L1 is shared by S4 and S5. Link L2 is shared by S2, S3, S4 and S5. Link L3 is shared by S1, S2, S3, S4 and S5. The third link L3 is a bottleneck link. TCP flow 1 had a minimum RTT of 0.05 seconds. TCP flow 2-3 RTT is 0.1 seconds and RTT of TCP flow 4-5 is 0.25 seconds. With Ts = 0.05 seconds and q_{th} = 60 packets, and α = 1 for stability the experiments were run [12] for a simulation time of 100 seconds and repeated 10 times.

The proposal provided the bandwidth fairly with flows, but each flow in TCP New Jersey did not share the throughput fairly as shown in Table 4.

Algorithms	TCP1	TCP2	TCP3	TCP4	TCP5
TCP New Jersey	2032.5	1171.4	1177	888.5	880.21
Proposal	1298.6	1317.1	1324.4	1291.4	1300.5

Table 4: Average Throughput of Links

CHAPTER III

METHODOLOGY

<u>Overview of the main ideas of the technologies/models/methodologies</u>

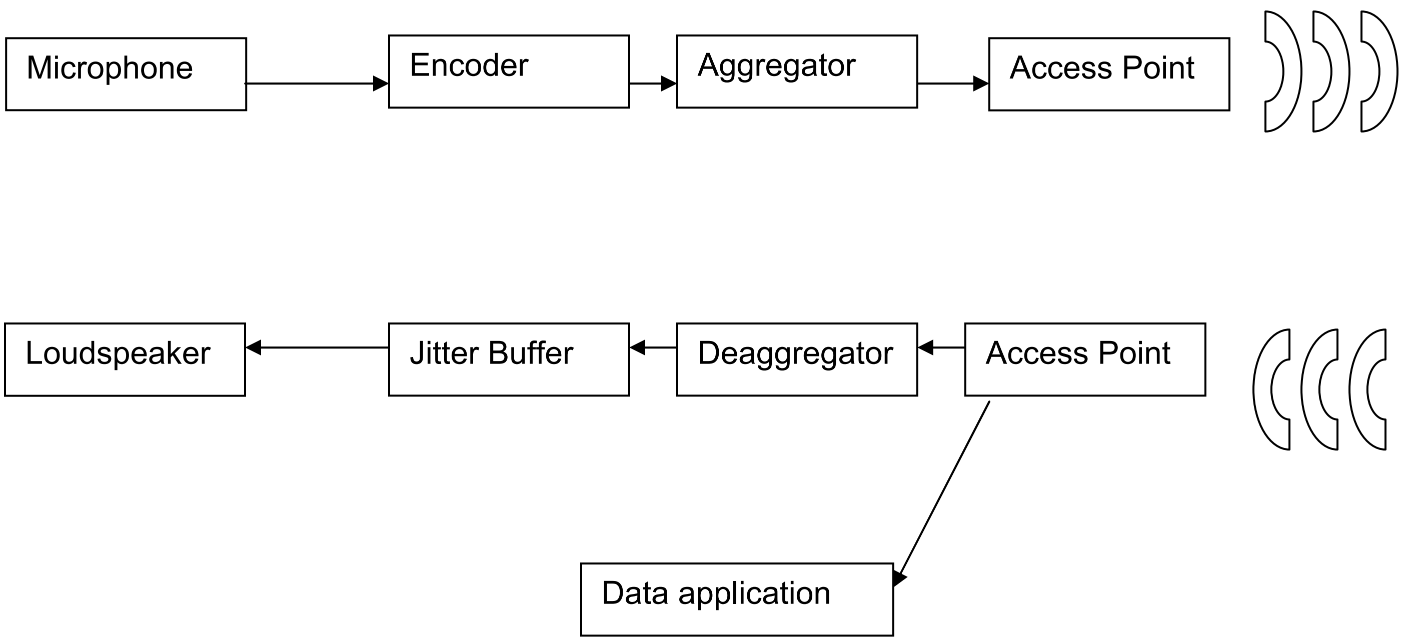


Figure 35: Proposed scheme Front end

The above figure has the topology for the proposed scheme. At one end of the topology there is a microphone which converts voice into electrical signals. G.711 codec is used to encode the signals to digital signals. The aggregator aggregates the packets into 80 Byte payloads. Access point priority gives higher priority to VoIP packets and lower priority to Data packets.

Aggregation increases the amount of data that can be transmitted. The transmission time for a 30B payload at 11Mb/second is 30*8/11 = 22μs [1]. The transmission time for a 80B IP/UDP/RTP header is 80*8/11 = 58μs. The 802.111 MAC/PHY layers have additional overhead of more than 800μs attributed to the physical preamble, MAC header, MAC back-off time, MAC acknowledgement (ACK), and intermission times of packets and acknowledgement. The overall efficiency increases from less than 3% to 6% if the payload increases from 30 Bytes to 80 Bytes. The time to transmit the data remains almost the same. The time to transmit a 30 Byte load is 822μs while the time to transmit a 80 Byte load is 858μs. We can still have about 10 sessions running on a 11Mbps IEEE 802.11b wireless connection.

However this aggregation can only be done for a limited number of packets as the jitter buffer limits the amount of aggregation that can be done.
If the loss rate is 1-2% loss rate and tolerable delay is 250 ms [4], the value of Np = 3 is preferable for a jitter buffer of up to 100 ms. This corresponds to (250 ms – (120 ms + 30 ms)) of delay jitter. This means that 3 out of order packets, each of 10 ms of voice data, can be stored in the jitter buffer and replayed. The redundancy factor Nr = 3 * Np is permissible and yields a loss rate of 1-2%. If voice data is generated at Nyquist optimum sampling rate of 8 kHz [5] and each sample has 8 bits, we have 64 kbps of data. 10 ms of voice data corresponds to 80 bytes of data.

We can have 3 out of order packets each of 80 bytes of aggregated data packets which will prevent dropping of packets which have arrived late. The aggregation of packets will improve traffic and service load. The optimum Nyquist optimum sampling rate will prevent too many packets from being generated as aggregation and de-aggregation of packets also takes time.

The access point should use 20dB SNR as it has been shown [8] to provide a clear advantage with various numbers of Bluetooth interferers.

The data application access point can have a different AIFS (Arbitrary Inter Frame Space) for data application packets.

AIFS = AIFSN * aSlotTime + SIFS

Where aSlotTime [9] is the value of a slot time and SIFS is the value of a Short IFS. AIFSN is an integer value, which is set according to the AC. The higher prioritized AC has the smaller AIFSN value.

This increases the priority for VoIP packets.

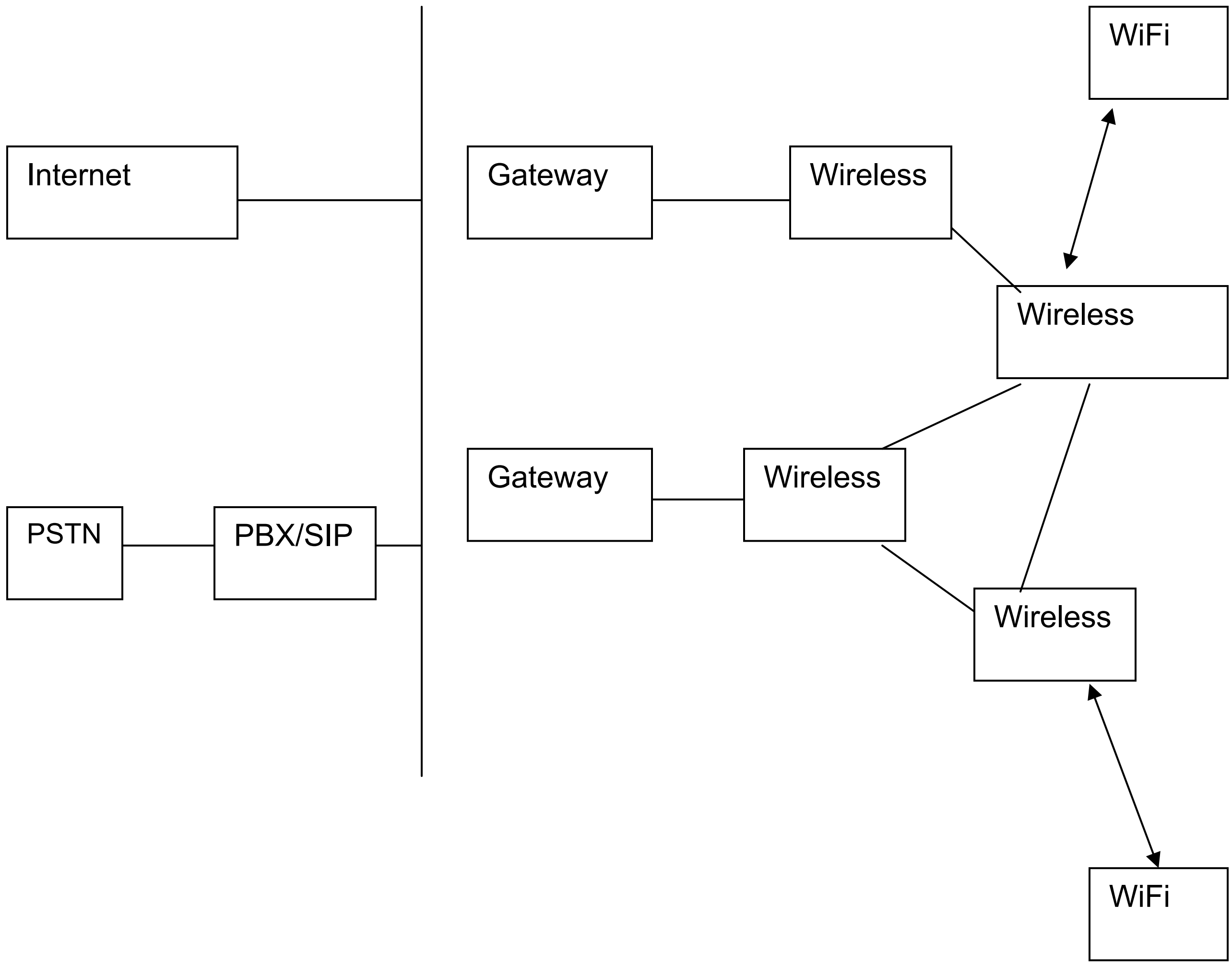


Figure 36: Proposed backend schema

The above proposed schema consists of Zyxel Prestige VoIP phones (with the microphone and loudspeaker set up shown in the front end scheme of Figure 35) which access the backend system [11]. The number of hops should be kept to a maximum of five. It has been found that when using a G.711 codec, five VoIP calls produce an acceptable quality (R>70). If G.729 or GSM codecs are used, 6 calls can be sustained [6].

	D-Link
Detection	1630 ms
Search	288 ms
Execution	2 ms
Total	1920 ms

Table 5: Handoff Times for D-Link

Instead of using a normal Access point, the proposed scheme could use a Novel Access Point [10] which balances loads among Access points and also transfers STAs in less time between Access points. This reduces the handoff time and improves the QoS for mobile VoIP users. As can be seen from Table 5, the handoff time can be quite large and cause unacceptable delays.

The mesh network can use labelling. A voice packet will get labelled if it is to be routed through the mesh network [11]. The TOS field of each IP packet can have a label specifying that it is a VoIP packet. All VoIP packets going to the same destination are aggregated at the ingress. Queues collect the VoIP packets going to the same destination. When the number of voice packets reaches a threshold, they are aggregated and sent to the destination. Header compression can be used to reduce the headers from 40 bytes to 14 bytes. Only 12 bytes are used and a 2 byte connection ID is required [11].

Strengths and limitations of them

The proposed scheme can provide coverage over a large area using mesh networks. The strength of Capacity improvement of Wireless LAN VoIP [9] is that it only requires VoIP STA software modification. There is no need of replacement of the Access Point. Also the compression of the header and aggregation in the mesh network reduce the amount of data transmitted by reducing the overheads. This increases the number of VoIP sessions that can be accommodated on the network. The jitter buffer improves sound quality at the receiver.

The disadvantage of the novel AP [10] is that the AP requires two transceivers. One transceiver is used to communicate with the mobile host. The other transceiver is used to communicate with the neighbour AP. The strength of the novel AP is that, from the user's stand-point, they can keep their mobile devices. They just need to upgrade their mobile device firmware.

The end-to-end scheme was chosen over the hop-by-hop scheme because packets are aggregated and de-aggregated at every hop by adding a forced delay at every hop. This adds a cumulative delay to all packets over multiple hops as well as adds an extra complexity at each node for executing aggregation/de-aggregation on every incoming packet.

Comparisons and contrast

The proposed scheme is different from the other schemes specified in the literature because it uses a combination of the schemes to improve QoS. There is a need to aggregate packets however aggregation is limited by the capacity of the jitter buffer. Also, mesh networks can be used with Novel Access Point to provide greater coverage for a wireless network as well as improve load balancing among access points and improve time of transfer of an STA from one access point to another.

The proposed scheme divides the setup into the front end and back end. The front end contains the microphone and loudspeaker which is normally found in the STA. The aggregator and de-aggregator can be incorporated in the front end. The back end contains the Novel Access point and Wireless mesh network.

The novel AP [10] and mesh network [11] both deal with mobile hosts connecting to different nodes. The novel AP covers the handoff between various APs. The mesh network deals with aggregation and header compression issues when routing messages between two points.

CHAPTER IV

ISSUES, CHALLENGES, AND TRENDS

A challenge that the Access point priority based Capacity scheme might pose is that when the AP is handling non real time traffic along with the VoIP calls, the downlink traffic will get priority over uplink leading to possible delays in the uplink real time traffic [1]. Despite these challenges, increasing the AP priority increases voice call capacity in a WLAN.

In Capacity improvement of Wireless LAN VoIP [9] the delays increase if more STAs are in the backoff states. The probability of delays increases with the number of STAs. Introduction of the proposed protocol requires only VoIP STA software modification. There is no need of replacement of the Access Point (AP).

In the Novel AP proposal, there is the situation where a ping-pong effect can occur between transceivers handing over control to each other. The OPNET Modeler program was used to simulate a handoff of APs, which were 500 meters apart. In a normal handoff, the delay was between 1700 and 1900ms. In the proposed handoff scheme the delay was between 3 and 5ms.

A simulation was done with a group of ten STAs was moving in the overlapping area of three neighbouring APs. The proposed handoff scheme provided a good level of traffic load balancing among APs. The total number of dropped packets is reduced using the proposed scheme.

In Wireless Mesh networks, aggregation increased [11] packet-delay as there were a need for processing. In hop-by-hop scheme, delay was added at every node due to the time needed to aggregate and de-aggregate the packets at each node. There was also extra complexity added at each node to perform the aggregation. One of the challenges of the decompression mode [11] was that all the nodes had to be synchronized and if they got out of synch, there should have been a method to synchronize all the nodes.

A common problem in packet aggregation is that it increases packet delay, which can possibly reduce its suitability for delay sensitive VoIP services. Aggregation is done only at the ingress node for all flows routed for a common destination. Aggregation more than doubles the capacity if we consider an R-value of 70 as the threshold in a mesh network. Short flows do not delay long flows for the purpose of aggregation. To emulate a simpler scheme that only transmits the changing fields; we reduce the header from 40 to 14 bytes so that VoIP protocol overhead coming from large header size can be relieved.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The TCP and VoIP packet interference [1] is reduced by using priority scheme at the access points. VoIP packets are given a higher priority increasing VoIP quality.

Multiplex-multicast (M-M) scheme [2] is performed by multiplexing packets. Results showed that 80%-90% more packets can flow in VoIP over WLAN with the M-M scheme than the regular VoIP over VLAN.

The Adaptive Jitter Buffering (AJB) and latency management algorithm [4] was introduced to manage the jitter buffer. It was found that skipping packets that have been lost prevents the jitter buffer from becoming too large and degrading the voice quality. Those packets that the system feels would arrive on time are waited for by introducing delays in the system and storing packets in sequence in the buffer. Older frames are piggybacked on new ones to reconstruct the audio. The end to end delay that the system can have with a good quality signal was incorporated into the AJB algorithm and allows for the system to know when to discard a packet, and when to wait for a packet to arrive.

It was found in the study [5] that transmitting voice packets with small amount of data prevented loss of a large amount of the conversation with the loss of voice packets. However, when the voice data packet was too small, the overhead to data ratio increased as the overhead for each packet was to be transmitted. Larger packet sizes tended to create saturation at the jitter buffer when the large packets got delayed. A higher sampling rate of the voice signal caused more packets to be transmitted. Nyquist theorem suggests a sampling rate of 8 kHz is enough for digitization of speech. The lowest possible sampling rate was taken to reduce the number of packets that needed to be transmitted.

The number of hops in a mesh network and the number of disconnections in a network affect the VoIP signal quality [6]. The study [6] showed that controlling the mutual interference between multi-hop nodes helped increase VoIP signal quality and increased the number of hops that a device could have. The study shows that codec adaptation increases the number of clients that can be supported in a VoIP network. It also shows that aggregation at the application layer can increase the VoIP flow.

The end-to-end VoIP quality [7] investigation was carried out to find out the quality of the signal from the sender to the receiver. It took into account not only the wireless VoIP portion of the network but the entire network. It found out that there were operational tradeoffs associated with the chosen voice frame size, the number of frames per IP packet and the transport network operational capacity.

Bluetooth devices were used to simulate interference with wireless devices. It was found that the strength of the bluetooth device did not degrade the quality of the wireless device as much as the number of bluetooth devices. The experiments showed that only with a SNR of 10dB was the signal attenuation of the wireless device affected [8].

Capacity improvement of Wireless LAN VoIP [9] can increase the number of VoIP calls by up to 50% using prioritization techniques for different ACs.

The novel AP [10] balances the load on APs and also reduces the time of handover of a mobile device from one AP to another. In the novel AP [10] approach, the AP largely improves the performance of wireless LANs in terms of latency time during the handoff and also provides traffic-load balance among APs in the network. The STAs in the proposed handoff scheme will start a handoff faster and smoother than in the conventional scheme. The handoff delay time of both the detection phase and the search phase are omitted, and the delay time of the execution phase is largely reduced. Furthermore, as the handoff process is controlled by the AP instead of the STA, the network is able to provide traffic-load balance among neighbour APs, thus the proposed scheme can reduce the number of dropped packets due to traffic overload. The simulation results show a large reduction of the delay time and a fair traffic-load sharing among neighbour APs. The disadvantage of the method is that APs must be equipped with two radios and users need to update the firmware of their wireless LAN cards. The seamless handoff scheme provided compensates these disadvantages by improving the support of real-time services and distribution of traffic load inside a WLAN.

I think that the above paper covers most of the topics in VoIP performance improvements. However, there are still a lot of methods, which have not been used to improve VoIP performance. VoIP performance improvement is a work in progress. VoIP is a relatively newer technology and has a lot of potential. The QoS is still quite low. A high bandwidth 11Mb/s line could in principle support more than 500 VoIP sessions however it can only support 10 VoIP sessions.

The available solutions like aggregation, use of a jitter buffer and sampling of data help reduce overheads, improve QoS and reduce the number of packets. However, these solutions have limited success in certain situations and cannot be applied widely. They require firmware of devices to be upgraded or purchase of additional equipment.

Video phone technology is a new technology which uses a video chip to stream low-bandwidth video over a portable satellite telephone. The equipment is inexpensive when compared to equipment currently used for similar purposes and does not require an expensive communication link. Video phony has the same issues as VoIP. Bandwidth capacity and routing delays have an effect on the picture quality.

Both VoIP and Video phony are getting more popular and need a higher QoS than currently available. Increase in the bandwidth provided to VoIP and Video phony consumers will increase the QoS. Smarter routers, which do not require too much header information, will reduce overheads. I think the answer to increase QoS of VoIP lies with the Internet Service Providers and Router designers’ ability to offer cost effective new innovations to their products increasing network bandwidth and reducing overheads for VoIP.

The purpose of this study was to find ways of improving QoS of VoIP. The present study found many ways of improving QoS of VoIP by using aggregation, adaptive jitter buffer and access point priority.

REFERENCES

[1] Hashmi, D. & Kiran, P.V. & Lall, B. (2006). Access Point priority based Capacity Enhancement scheme for VoIP over WLAN*. Proceedings of the Annual India Conference,* pp. 1-4, 2006

[2] Wang, W. & Liew, S.C. (2005). Solutions to Performance Problems in VoIP over 802.11 Wireless LAN. *IEEE Transactions on Vehicular Technology,* Issue 1, Vol. 54, pp. 366-384, Jan. 2005

[3] Kim, D.H. & Jong, K.K. & Kim, D.Y. (2006). Periodic medium reservation timer mechanism for VoIP in IEEE 802.11 wireless LAN. *Advanced Communication Technology*, vol. 3, pp. 4, 2006

[4] Madhani, S. & Shah, S. & Gutierrez, A. (2007) Optimized Adaptive Jitter Buffer Design for Wireless Internet Telephony. *Global Telecommunications Conference,* pp. 5248-525, Nov. 2007

[5] Ergen, M. & Ergen, S.C. & Varaiya, P. (2006). Throughput Performance of a Wireless VoIP Model with Packet Aggregation in IEEE 802.11. *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE vol. 4,* pp. 2235-2239

[6] Comeras, M.P. & Bafalluy, J.M. & Suriol, M.C. (2007). Performance issues for VoIP call routing in a hybrid ad hoc office environment. *Mobile and Wireless Communications,* pp. 1-5

[7] Hernandez-Valencia, E.J. & Chuah, M.C. (2000). Transport delays for UMTS VoIP. *Wireless Communications and Networking Conference,* vol. 3, 23-28, pp. 1552-1556, Sept. 2000

[8] McKay, C. & Masuda, F. (2003). Empirical studies of wireless VoIP speech quality in the presence of Bluetooth interference. *IEEE International Symposium on Electromagnetic Compatibility, vol. 1, 18-22,* pp. 269-272, Aug. 2003

[9] Igarashi, K. & Yamada, A. & Ohya, T. (2007). Capacity Improvement of Wireless LAN VoIP using Distributed Transmission Scheduling *Personal, Indoor and Mobile Radio Communications 3-7,* pp. 1-5

[10] Manodham, T. & Miki, T. (2006). A Novel AP for Improving the Performance of Wireless LANs Supporting VoIP. *Journal of Networks*, vol. 1, no. 4, pp. 41-48, Aug. 2006

[11] Ganguly, S. & Navda, V. & Kim, K. & Kashyap, A. & Niculescu, D. & Izmailov, R. & Hong, S. & Das, S. (2006). Performance Optimizations for Deploying VoIP Services in Mesh Networks, *IEEE Journal on Selected Areas in Communications,* vol. 24, Issue 11, pp. 2147-2158, Nov. 2006

[12] Ho-Jin, L, Hee-Jung B. & Jong-Tae L. (2007). TCP Window Control for Variable Bandwidth in Wireless Cellular Networks, *Communications Letters, IEEE,* vol. 11, Issue 2, pp. 152-154, Feb. 2007