SUPPORTING VOIP IN IEEE802.11 DISTRIBUTED WLANS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2013

By

Zuo Liu

School of Computer Science

Table of Contents

List of Tables

List of Figures

List of Abbreviations

GSM	Global System for Mobile Communications
IEEE	Institute of Electrical and Electronics Engineers
ITU	International Telecommunication Union
LAN	Local Area Networks
LLC	Logical link control
LTE	long-Term Evolution (Telecommunication)
MPDU	MAC Protocol Data Unit
MPSU	MAC Protocol Service Unit
MTU	Maximum Transmission Unit
MOS	Mean Opinion Score
MAC	Medium Access Control
Mb/s	Megabit per Second
MANET	Mobile Ad-hoc Network
MIMO	Multi Input Multi Output
NS	Network Simulator
NIC	Network Interface Card
PLC	Packet Loss Concealment
PESQ	Perceptual Evaluation of Speech Quality
PSQM	Perceptual Speech Quality Measure
PC	Personal Computer
PCF	Point Coordination Function
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RED	Radom Early Detection
RTP	Real-time Transport Protocol
RTS	Request to Send
RTT	Round-Trip Time
RTCP	RTP Control Protocol
SNR	Signal to Noise Ratio
SPAWN	Small Packet Aggregation for Wireless Network
STA	Station
TXOP	Transmit Opportunity
TCP	Transport Control Protocol
UMTS	Universal Mobile Telecommunications System
UDP	User Datagram Protocol
VoIP	Voice over Internet Protocol
VoWLAN	VoIP over WLAN
WLAN	Wireless Local Area Network

Abstract

Telecommunications is converging on the use of IP based networks. Due to the low cost of VoIP applications, they are being increasingly used instead of conventional telephony services. IEEE802.11 WLANs are already widely used both commercially and domestically. VoIP applications will also expand from usage over wired networks to voice communications over IEEE802.11 WLANs. This is known as VoWLAN. The use of VoWLAN may reach the maximum capacity of a wireless channel if there are many simultaneous VoIP calls operating close to each other.

There is much published research based on a single IEEE802.11 infrastructure WLAN concluding that packet loss, transmission efficiency and latency issues are the major challenges limiting the VoWLAN capacity. The VoIP service quality will drop sharply when the demands exceed the WLAN's capacity. This thesis demonstrates that these challenges also apply to distributed WLANs. To extend these findings from the existing research, the analysis in this thesis indicates that the capacity of a single IEEE802.11 WLAN channel is 12 VoIP calls. When the number of simultaneous VoIP calls is within the capacity, the WLAN can deliver more than 90% of the voice packets to the receiver within 150 ms (the lowest network performance for supporting acceptable VoIP service). However, as soon as the traffic loads are beyond the wireless channel capacity e.g. the number of simultaneous VoIP calls is greater than 13, the VoIP service quality catastrophically collapses. When the capacity is exceeded there are almost no voice packets that can be delivered to the receiver within 150 ms. Our research results indicate that the delay accumulation for voice packets in the transmitter's outgoing buffer causes this problem.

Our research also found that dropping ‘stale’ voice packets that are already late for delivery to the receiver can give more transmission opportunities to those voice packets that may still be delivered in time. This thesis presents a new strategy called Active Cleaning Queue (ACQ) which actively drops ‘stale’ voice packets from the outgoing buffer and prevents the accumulation of delay in congested conditions. When ACQ is applied in a saturated wireless channel the network performance for supporting VoIP traffic was found to gradually decrease proportional to the numbers of simultaneous VoIP calls rather than catastrophically collapse.

There is also published research suggesting that the aggregation of packets can improve the efficiency of WLAN transmissions. An algorithm called Small Packet Aggregation for Wireless Networks (SPAWN) is also presented in this thesis to improve transmission efficiency of small voice packets in WLANs without introducing further delay to VoIP traffic. The evaluation result shows that after applying the SPAWN algorithm, the VoIP capacity of a single wireless channel can be extended up to 24 simultaneous calls.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

First of all, I would like to thank my supervisor Dr. Nick Filer, who taught me a solid knowledge in wireless networking background and valuable skills in scientific research, academic writing and presentation. Furthermore, he also provided me invaluable technical guidance and support to my research such as, evaluating my ideas, suggesting solutions to me and commenting my research results. And moreover, he spent a large amount of effort and his valuable time for correcting and proof reading this thesis.

I would also like to thank my colleagues who exchanged knowledge and shared ideas with me.

Finally, I would like to thank my parents Mr Jianing Liu and Ms Cuilan Chen who gave me birth and built up my initial interest in computer science in my childhood. They also provided me an outstanding education opportunity to study abroad.

1. Introduction

The conventional telephony service is based on the Public Switched Telephone Network (PSTN) where the service provider offers dedicated network transmission to support high service quality. The rise of Voice over Internet Protocol (VoIP) applications has become an alternative solution for providing telephony service. Such applications are designed to use IP based networks. The coverage of the internet rapidly expanded worldwide and has become assimilated into people's everyday life. The cost for internet access has dropped to a level that is affordable for most people. Due to typically having no additional cost to the user, VoIP has become a strong competing technology to conventional service, especially for international telephony. Computer users are now increasingly familiar with applications for making VoIP calls such as Skype [1].

Wireless communication technology has been used in mobile telephone networks for telephony service over several decades. Cellular mobile telephony networks such as the Global System for Mobile Communications (GSM) and the Universal Mobile Telecommunication System (UMTS) provide mobility and flexibility in addition to telephony service. Similar to PSTN, such networks also provide dedicated transmission links over wireless channels to guarantee the service quality for callers. Soon after Wireless Local Area Networks (WLANs) emerged in the 1990's, they began to be used as a flexible and cheaper alternative to wired Local Area Networks (LANs) [2]. Although a WLAN also uses wireless communication technology, it is fundamentally different from mobile telephone networks. Firstly, links are not dedicated to particular users as with the traditional connection based telephone system. Conversations are normally carried over shared wireless channels. Secondly, WLANs are packetised networks that are not optimised for voice telephony. IEEE802.11 [3] WLANs, which are also known as WiFi, are now widely used for the sharing of internet access both commercially and domestically. The density of IEEE802.11 WLANs has been continuously increasing over the past decade especially in urban areas. VoIP applications are also expanding from being used over wired networks onto IEEE802.11 WLANs. VoIP over WLAN is often called VoWLAN.

1.1. Research Motivations and Objectives

1.1.1. Research Context and Scenarios

People now can easily get WiFi connections via either public or private WLANs. Therefore, they may use VoWLAN in various situations, for example:

1) In a large open plan office, wireless handsets are equipped on each desk to avoid cabling issues. People may either make external phone calls via the Access Point (AP) and the internet, or internal phone calls directly to their colleagues' handsets in an ad hoc mode without going through the AP.
2) When people carry their smartphones around their home, they may use smartphones to make phone calls through household wireless routers and the internet rather than through the mobile telephony network.
3) In a shopping mall or amusement park a group of friends may setup a wireless ad hoc network using VoWLAN to keep in contact with each other.

From the users' point of view, VoWLAN is beneficial in reducing the cost of using a mobile telephony service. From the service providers' point of view, there may be new commercial possibilities of offering new services to customers or improving the existing service coverage. They can extend their telephony service coverage to VoWLAN via their available WiFi hotspots near to users, for example, inside some buildings where the cellular signal is too weak. This may also reduce the number of conversations that are carried over their mobile cellular networks.

The topology of an IEEE802.11 WLAN can be either infrastructure or ad hoc. In an infrastructure WLAN there is a central AP through which all the network traffic is passed and the other wireless nodes are like satellite stations (STA). In an ad hoc WLAN there is no such central AP. The network traffic is transmitted directly between wireless nodes.

Most of the existing research on VoWLAN such as [2, 4-12] has been carried out based on the scenario of a single infrastructure WLAN. However, this scenario is not the case in the real world. The IEEE802.11 standard only specifies three non-overlapping channels in the 2.4GHz band where the IEEE802.11b and g devices work at [13, 14] and eight non-overlapping channels in the 5GHz band for IEEE802.11a devices [15]. However, the number of channels in each band may vary across different regions according to local regulations. Due to the popularity of people using WiFi, in a busy location there could be many independent WLANs operating on the same channel especially for those devices working in the 2.4GHz band. For example, in a big apartment building, each household may set up their own wireless router. In this case, the transmissions of several independent WLANs may be overlapped with each other and share a single wireless channel. Hence, VoWLAN might operate in such dense environments.

Even though the closely located WLANs in adjacent households are in infrastructure mode, the sharing of the wireless channel amongst them is without any central control. Furthermore, ad hoc networks may also exist amongst the WLANs. Because there is no central control for the transmissions over the wireless channel, the sharing of a wireless channel amongst transmissions relies on the active wireless nodes (including the APs) in all WLANs to cooperate with each other. These are called 'distributed WLANs' in this thesis in order to distinguish them from a pure ad hoc topology.

1.1.2. Research Motivations

Despite its rising popularity, there are still challenges that obstruct the further development and expansion of VoWLAN. The fundamental difference between IP based networks and telephony networks such as PSTN is that the transmission link is not dedicated and therefore the service quality is not guaranteed to the users. As shown by Goode in paper [16], variations in transmission latency, the bandwidth efficiency and packet loss have become major challenges for providing reasonable quality VoIP calls. If the transmission latency is too high, it may result in a large round trip delay and make the interactive conversation between two users difficult to maintain. A high packet loss rate may result in the users being unable to understand the voice speech.

The International Telecommunication Union (ITU) recommends in the document G.114 [17] that the one way transmission latency of a voice packet should not exceed 150 ms. In our research scenario, distributed WLANs, we focus on the one way transmission latency averaged amongst all the pairs of wireless nodes. To simplify our research, we assume that the uplink and downlink transmissions for each wireless node have similar performance. Hence, in our research, the round trip transmission latency is estimated to be approximately double the one way transmission latency. The maximum acceptable round trip transmission latency is therefore about 300 ms.

Furthermore, the physical and data link layer characteristics of IEEE802.11 WLANs intensify the challenges to VoIP quality. In modern wired networks, the point-to-point transmission is usually over dedicated cable and the medium is stable. However, the transmissions in WLAN may have to share the same wireless channels in which the medium conditions vary massively over time. Extra overheads are required to ensure that the data is delivered correctly and also that the transmissions do not interfere with one another over the shared wireless channel. There is a large amount of research such as [2, 4-12] showing that the transmission efficiency, packet loss and transmission latency issues are important limitations on the VoWLAN capacity.

As stated previously in this chapter, people's use of both VoIP and WLAN is rising. There could be many users making VoIP calls over different WLANs in the same area. The number of VoIP calls carried by some WLANs may reach the capacity limit of the wireless channels they provide. Research is required in order to fulfil the increasing demand on VoWLANs. The possible impacts of this research include:

1) Being able to expand the VoIP call capacity of each wireless channel;
2) Improving the VoWLAN quality in a busy shared wireless channel;
3) Making VoWLAN a better option for both customers and service providers.

1.1.3. Research Objectives

The objectives of this research are to:

1) Find out whether predictions of the VoIP capacity over a single infrastructure WLAN found in the existing research [2, 4-12] are also valid for distributed WLANs.

2) Extend the existing research findings and study further to discover what happens after the distributed WLANs reach the VoIP capacity in a wireless channel.

3) Propose some novel solutions that aim to improve the VoIP service quality and capacity after the distributed WLANs reach the VoIP capacity limit in a wireless channel.

The applicability of the proposed solutions is also considered as an issue in this research. There are a large amount of mobile handsets and other devices being used by users with WiFi (mainly the IEEE802.11b and g) enabled. The proposed solutions need to be applicable to those off the shelf devices. After deploying such solutions, the VoIP service quality could be improved and VoWLAN capacity could be expanded without the replacement of any existing devices, though updated software may be required. As a result, a constraint is applied to this research that the proposed solutions must be compatible with the existing IEEE802.11 standards.

This work focuses on the data link layer issues for VoWLAN to investigate the network performance when the VoIP demands exceed the capacity (in congested wireless channel) of WLANs. Hence, only the point-to-point transmissions at the data link layer are considered.

1.1.4. Research Hypotheses

In order to achieve the research objectives, two hypotheses are drawn at the beginning of this research:

> **Hypothesis 1** "The problems for limiting the VoIP capacity over a single infrastructure WLAN found by existing research also apply to the scenario of distributed WLANs" is drawn from research objectives 1) and 2).
>
> **Hypothesis 2** "The VoIP service quality and capacity of VoIP calls over distributed WLANs can be improved without modifications to the existing IEEE802.11 standards" is drawn from research objective 3).

1.2. Research Questions and Contributions

1.2.1. Research Questions

There are four research questions that need to be addressed in order to achieve the objectives of this research:

1) Are the factors imposing the upper bound of VoIP capacity for infrastructure WLANs the same for distributed WLANs?
2) How will the VoIP quality of service be affected in congested WLANs?
3) How can the collapsing in VoIP service quality be prevented in congested conditions?

4) What is the solution at the data link layer for distributed WLANs to improve the VoIP traffic efficiency?

In order to prove Hypothesis 1, question 1) needs to be answered. Several existing papers based on IEEE802.11 infrastructure WLAN such as [10, 18, 19] suggested that there are correlations between the service quality of VoIP and the configuration of the outgoing buffer. Paper [18] indicated that in congested conditions a sharp drop in network performance is expected from low-latency with a high-delivery rate down to high-latency with a low delivery rate. In order to study this further, question 2) extends the details from question 1).

In order to prove Hypothesis 2, this research will find some solutions without modifying the existing IEEE802.11 standards, which can prevent the catastrophic failure of VoWLAN under congested conditions and may also improve the transmission efficiency of small voice packets. Two research questions, 3) and 4) are specified to answer this.

1.2.2. Research Contributions

Answering the research questions that mentioned previously leads to four possible research contributions:

Contribution 1 We demonstrate that limitations on VoIP service quality in infrastructure networks described in existing research publications also apply to distributed WLANs.

In order to answer question 1), this research needs to repeat the analyses for infrastructure WLANs described in existing publications such as [2, 4-12] on distributed WLANs. The effects caused by transmission efficiency, packet loss and transmission latency issues in distributed WLANs will be examined according to the network performance requirements of VoIP applications. Conclusions will be drawn from the observations. This leads to Contribution 1.

Contribution 2 We provide an analysis to extend the existing research finding on the VoWLAN performance collapsing in congested conditions and identify the causes of this phenomenon.

To answer question 2), the work of this thesis will extend the findings in paper [18], and investigate the causes of the collapse in network performance in congested conditions. This research will focus on the change in network performance caused by buffering in congested conditions in distributed WLANs. The effect of buffering in congested conditions on the VoIP service quality will be found out. This leads to Contribution 2.

Contribution 3 We propose a new queue management strategy for WLANs to prevent the sharp drop in VoIP service quality in congested conditions.

In order to answer question 3), a strategy for buffer (queue) management is developed based on the findings from the answers to Question 2. This strategy aims to eliminate the causes of the sharp drop in network performance in congested WLANs. By applying this strategy, the network performance is expected to remain in the low-latency high-delivery rate region as the number of calls increases beyond 12; of course, the performance for higher VoWLAN density still decreases proportionally to the increasing demands on WLANs. This strategy will be validated by extensive simulations. This leads to Contribution 3.

Contribution 4 We propose a new packet aggregation algorithm for WLANs to improve VoIP traffic efficiency.

Aggregating small packets into larger ones is expected to increase the efficiency of transmissions utilising wireless channels. However, there is an important characteristic that needs to be considered to apply this for VoWLAN. Aggregation requires some delay (forced delay) in forwarding the existing packets, waiting for the arrival of the latter packets to be aggregated with. The resulting delay may have a negative effect on VoIP applications. Furthermore, the new aggregation algorithm presented here does not need any change to existing wireless standards or to the majority of existing wireless network cards. Answering question 4) leads to Contribution 4.

1.3. Thesis Structure

The rest of this thesis is structured as follows and visualised in Figure 1.1:

Chapter 2 provides the background knowledge for this research. This includes an introduction to VoIP applications, the measurement of VoIP service quality and the VoWLAN. It also defines the research scenarios and problem scope for this research.

Chapter 3 defines and gives a detailed analysis of the research problems by means of numerical analysis, real device experiments and simulations. It demonstrates that the limitations on VoIP service quality in infrastructure networks concluded by existing research can also apply to distribute WLANs. To extend this research result, delay accumulation is found to cause the sharp drop in network performance for VoWLAN in congested conditions.

Chapter 4 presents a new queue management strategy to actively drop out the 'stale' voice packets from the outgoing buffer in congested conditions. The testing results show that this strategy can prevent delay accumulation in the outgoing buffer under congested conditions and maintain the network performance for VoWLAN in low-latency with high-delivery rate as the number of calls increases beyond 12.

Chapter 5 presents a new packet aggregation scheme that aggregates the outgoing packets without forced delay. The testing results show that packet aggregation can effectively improve the transmission efficiency of VoIP traffic. The aggregation scheme can double the VoWLAN capacity compared to that which is without aggregation.

Chapter 6 concludes and evaluates this research. It also suggests several possible further research topics and directions.

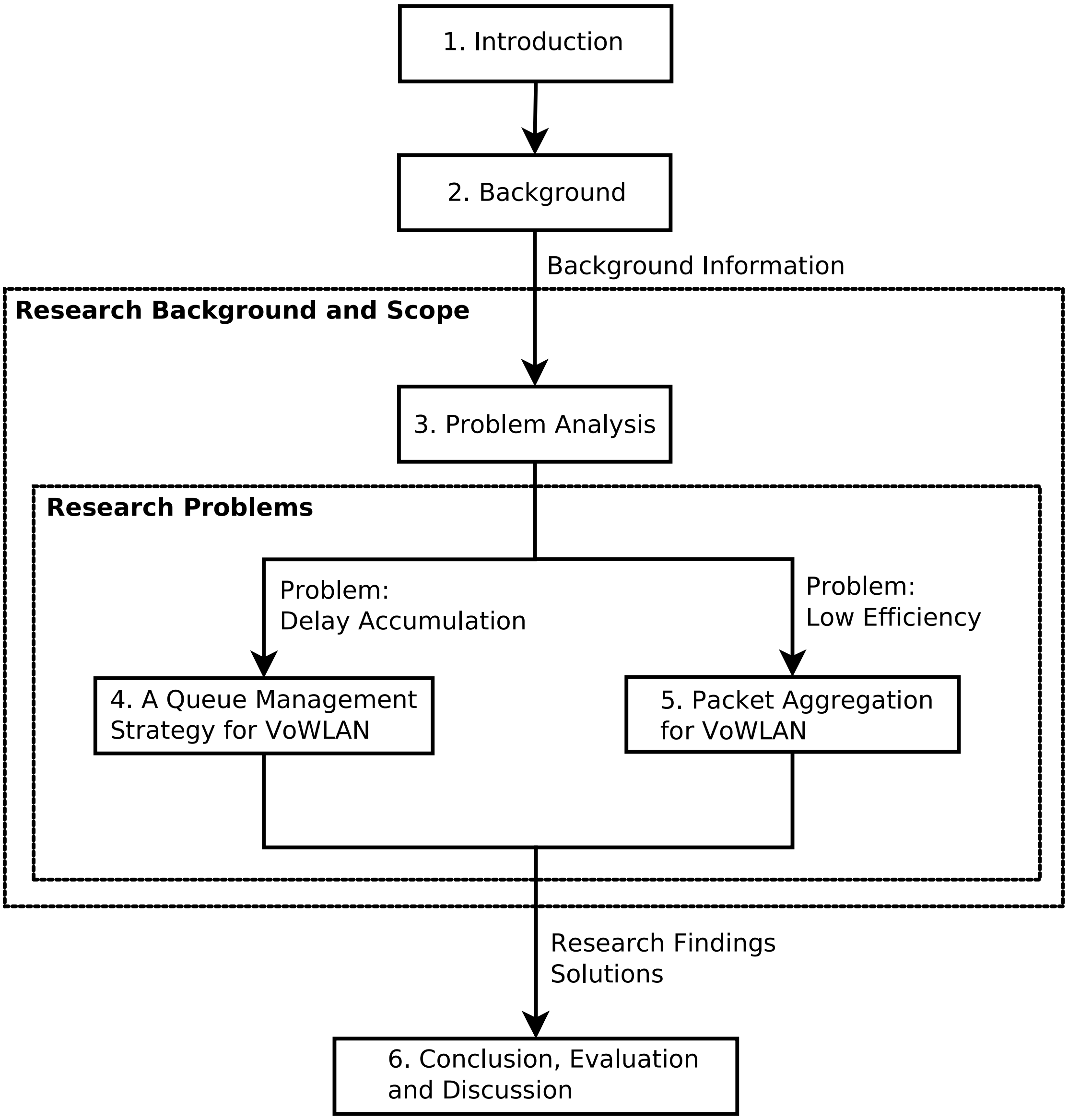

Figure 1.1Thesis Structure

2. Background

As mentioned in the introduction of Chapter 1 while computer networks, and especially the internet, are rapidly expanding, telecommunications is converging from a conventional PSTN on to an IP based network model. However, the IP network was not originally designed to carry real time applications such as voice services and there is a great deal of potential for the quality of the voice service to be impaired in IP networks due to transmission delays, packet loss and jitter, etc.

This chapter introduces the background knowledge required in order to present this research. Section 2.1 provides a brief introduction on VoIP applications including its protocol architecture, voice codecs, its network traffic patterns and how codecs deal with voice packet loss. Section 2.2 introduces different ways of measuring the VoIP service quality including subjective measurement, intrusive measurement and non-intrusive measurement. Section 2.3 introduces IEEE802.11 WLANs, the existing research on the capacity of VoIP in WLANs and the scenario of this research.

2.1. VoIP Application

VoIP is a family of applications that are implemented based on open protocols and packetised computer network architecture and provide interactive end-to-end voice services. The VoIP application initially emerged on a wired network, PC to PC, and was mainly used for long distance telephone services, especially over the internet. Nowadays, this technology tends to be moving towards handheld and mobile devices over wireless networks. The service coverage is also converging to local area networks such as within office buildings.

2.1.1. VoIP Application Protocol Architecture

Table 2.1 lists the major protocols used by VoIP applications at different layers. Just under the actual VoIP application Section Initiation Protocol (SIP) [20] and H.323 [21] are used to initiate the VoIP calls between users; Real Time Protocol (RTP) [22] is implemented to standardise the voice packet format; and RTP Control Protocol (RTCP) [23] is for monitoring the quality of service and report to the application. At the transport layer, User Datagram Protocol (UDP) [24] is used most of the time for VoIP applications because VoIP applications can tolerate a few packet losses, rather than a large delay caused by retransmitting the lost packets; however, the Transmission Control Protocol (TCP) [25] will also be used with less efficiency on some occasions such as when the UDP packets are filtered by the internet service provider. At the network layer, Internet Protocol (IP) [26] is used to route the voice packets over different networks to reach the end user.

Table 2.1 Major Protocols for VoIP at each Layer

Layers	**Protocols**
Application	RTP, RTCP, SIP, H.323
Transport	UDP, TCP
Network	IP
Data Link and Physical	IEEE802.3 (Ethernet), IEEE802.11 (WiFi), 3/4G, WiMax

The VoIP applications are not really concerned with the protocols at the bottom two layers, the data link layer and the physical layer. As mentioned above, the VoIP applications were initially based on the wired network and PC to PC, where the

Ethernet is used in most of the cases. As handheld and mobile devices such as smartphones and tablets become for VoIP applications, WiFi, 3/4G or WiMax are needed for wireless connections. This thesis focuses on the data link layer issues for supporting VoIP applications over distributed IEEE802.11 WLANs. Hence the other protocols mentioned above will not be studied in detail.

2.1.2. VoIP Codec

As shown in Figure 2.1, the VoIP application has both an encoder and decoder implanted at each end of the VoIP call. During a conversation between two users, the VoIP application uses an encoder to record and encode the analogue signal into digital data. The digital voice data is also called voice samples. They are encapsulated into RTP packets and transmitted to the decoders on the other end of the conversation. When voice packets are received, they may be out of order. The decoder puts them back into the original order, converts them back into an analogue signal and plays to the user.

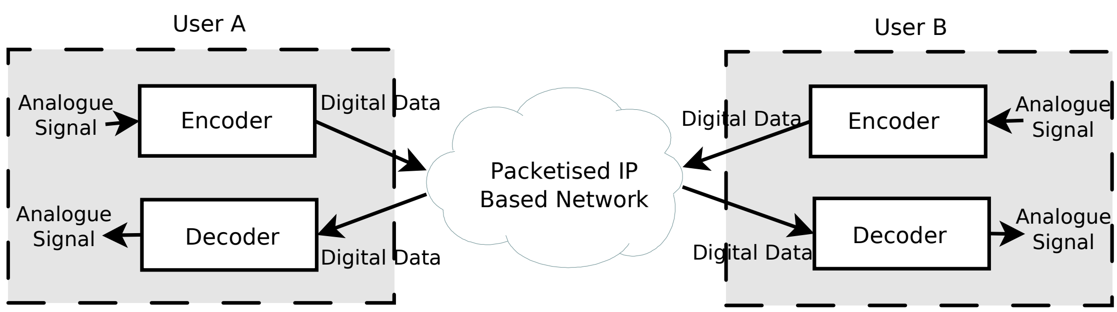

Figure 2.1 VoIP Encoding and Decoding

In order to maintain a low latency when delivering the voice samples, the encoder attempts to send out the voice data as soon as i.e. have ‘a few’ samples encoded. As a result, the voice encoder is constantly generating small RTP packets (voice packets) which need to be transmitted immediately by the network. The frequency of the RTP packets generated is called the packet rate.

The ITU published a series of voice encode-decode standards (codecs) for digitising voice signals. In addition, there are other series of voice codec standards, such as Speex [27]. Table 2.2 lists several well-known voice codec standards [12, 28] published by ITU. Most of the existing studies on VoWLAN used G.729 [29] and G.711 [30]. Firstly, the quality of compression between G.729 and G.711 are significantly different. This can be seen in the difference between their data rates. Secondly, they can operate at the same packetising interval (20 ms) for better comparison. Therefore, in order to be consistent with existing studies G.729 and G.711 are also selected for investigation in this research. In the rest of this thesis, the default packet interval is set as 20 ms unless stated otherwise.

Table 2.2 Popular ITU Codec Standards [12, 29-32]

Codec	Packet Rate	Packet Interval	Packet Payload Size	Data Rate
G.711	50 Hz	20 ms	160 byte	64 Kbit/s
G.729	33/50/100 Hz	10/20/30 ms	8/20/24 byte	6.4/8/6.4 Kbit/s
G.726-32	50Hz	20 ms	80 byte	32 Kbit/s
G.723.1	33 Hz	30 ms	20/24 byte	5.3/6.3 Kbit/s

2.1.3. VoIP Traffic Pattern

Table 2.2 above also indicates that the VoIP application generates a constant rate of network traffic. This is very different from the traffic pattern generated by non-real-time network applications, such as web browsing, file transfers and e-mails. For non-real-time network applications, the temporary demand on the network increases dramatically since a block of data needs to be transmitted. The size of the data block varies from a few kilobytes to many megabytes. For these applications, the network will typically use a series of large packet payloads to transfer the data as soon as possible, and then wait for the next cycle of data arrival. The typical duration of the next data arrival varies randomly from a few seconds to hours. This pattern is called bursty traffic. A burst of network traffic is generated when an application transmits files, such as web pages. For non-real-time applications, the data does not need to be delivered immediately. Users can tolerate the delays of transmission ranging from a few seconds (e.g. for web browsing) to hours (e.g. for file transfers).

An example of bursty traffic generated by a web browsing application model is given in paper [33] where each 'burst' of traffic arrives randomly with a mean interval of 10 seconds, and the amount of data within each burst is random generated with an average size of 7382 bytes.

On the other hand, the VoIP application transmits a stream of small packets at constant intervals. Compared to the bursty traffic, the voice data arrives much more frequently and the amount of data at each arrival is much less than a burst block (see Table 2.3). Each VoIP call may last from a few seconds to a few hours. In order to remain interactive for this real-time application the voice data must be delivered immediately. In a long term view, the VoIP application will generate much higher traffic demand than non-real-time applications.

Table 2.3 VoIP Traffic versus Burst Traffic

Applications	**Data Arrival Interval**	**Data Amount at Each Arrival**	**Bit Rate (Avg)**
G.729	20 ms	20 byte	8 Kbit/s
G.711	20 ms	160 byte	64 Kbit/s
Web Browsing[33]	10 second (average)	7382 byte (average)	0.74 Kbit/s

2.1.4. Packet Loss Concealment

The voice packets might be lost in transmission while a VoIP call is in progress. The causes of packet loss for VoIP include arriving too late to be decoded and played to the user; being dropped by the sender or intermediate routers usually due to a full buffer; or being discarded by the receiver due to bit errors. The loss of a voice packet usually results in the loss of several voice samples. If the decoder directly plays the next available sample, it will result in un-synchronisation (caused by the shift of sounds in the time domain) between both ends of the VoIP call and therefore considerable impairment to the original speech. In order to mask the effect of voice packet loss and reduce the impairment to the original speech, Packet Loss Concealment (PLC) techniques are usually used by codecs to generate synthetic voice samples to replace lost packets. Such techniques are sometimes also known as Frame Loss Concealment techniques.

One of the simplest PLC techniques is replacing all the missing voice samples by zero. Those zero samples will sound as silence when being played to the user. Most of the PLC techniques that are specified by existing codecs are more complex than this.

There is a type of PLC technique which generates synthetic voice samples based on some previous correctly received samples and then fills the missing voice segment with these synthetic samples. They are known as waveform substitution techniques [34]. Both G.729 and G.711 codecs have these type of PLC techniques built in [29, 30, 35]. G.729's waveform substitution technique firstly predicts whether the missing packet is voiced (with speech recorded in) or un-voiced (no speech recorded in). For a missing voiced packet G.729 will try to generate a segment of sound with similar pitch to the previous correct samples [29, 36]. If the missing packet is un-voiced G.729 will generate a segment of sound which gradually attenuates the previous correct samples [29, 36]. G.711 implements a more complex waveform substitution technique. It examines a much longer segment of previous correct voice samples to learn the patterns of the speech waveform. The synthetic voice samples are generated according to the predicted waveform based on these patterns [30]. The PLC techniques for both G.729 and G.711 codecs perform very well for a single packet loss but it is difficult to fix a large burst of missing packets [36, 37].

Another type of PLC technique the so-called model-based PLC techniques [38] uses mathematical models to predict the waveform of missing voice samples. Because neither G.729 nor G.711 considered in this thesis uses model-based PLC technique, they are not introduced in further detail.

2.2. VoIP Quality Measurement

In the past few decades, the telecommunications industry has developed several measurement methods to indicate the quality of the voice speech being delivered to the users. Engineers and academics also use these methods to evaluate the quality of VoIP service delivered over computer networks. These measurements are either subjective or objective. The subjective measurements are normally conducted using a large number of listeners in order to produce an unbiased averaged opinion. Each candidate evaluates the speech sample individually and the final evaluation is summarised by statistical processes. Objective measurements are more efficient, simpler and are classified as either intrusive or non-intrusive. The intrusive measurements require the use of the original speech sample for comparison with the perceived speech and then it calculates the deduction in quality. The non-intrusive measurements require the input of network parameters and use analytical models to estimate the quality of speech delivered by the network.

2.2.1. Mean Opinions Score an Subjective Measurement

Mean Opinion Score (MOS) [39] is a voice quality measurement method recommended by the ITU mainly designed for evaluating voice compression techniques and codecs. It is a widely recognised subjective measurement for voice service quality. In the ITU recommendations, the environments e.g. the room size, noise level for both recording and listening to the speech samples are specified. There are several standardised subjective test procedures defined in the recommendations for candidates to evaluate the quality of the speech sample. The well-known MOS score is defined in the Absolute Category Rating (ACR) test procedure [39]. In this procedure, the speech samples will be evaluated into five different categories by the candidates.

Table 2.4 presents these categories and their corresponding indications of the speech quality specified in the ITU recommendation [39]. The MOS score of a particular speech is the average of the evaluations over all listeners.

Table 2.4 Mean Opinion Scores [39]

MOS	Quality	Indications of the Speech Quality
5	Excellent	Complete relaxation possible; no effort required
4	Good	Attention necessary; no appreciable effort required
3	Fair	Moderate effort required
2	Poor	Considerable effort required
1	Bad	No meaning understood with any feasible effort

Because of its clear and simple indication, MOS scores have become a well-accepted metric for evaluating the voice service quality used by both engineers and researchers. However, conducting the subjective testing procedure is very complex and expensive. Objective speech quality measurements were developed to overcome these disadvantages and many of them are based on the MOS score metrics.

2.2.2. Intrusive Measurements

In order to measure the voice quality delivered by the telephone service without needing the subjective opinions from the listeners, engineers and researchers have developed some measurement methods to compare the original speech sample with the one delivered to the end user. These are so-called intrusive measurements.

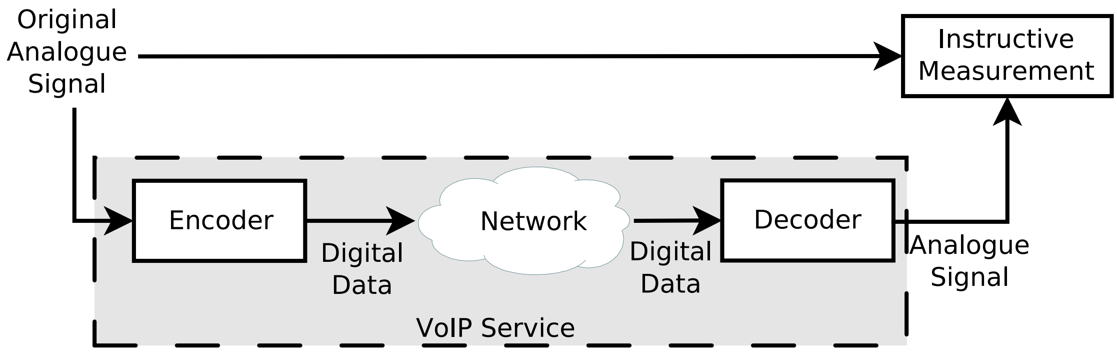

Figure 2.2 Intrusive Measurement for VoIP Service Quality

Figure 2.2, presents the way which the intrusive methods are applied for measuring the service quality of a VoIP application. The original voice (analogue signal) will be distorted during the encoding-decoding process and network transmission. After the voice is delivered and converted back to an analogue signal, it will be compared with the original sample and the reduction in quality calculated using analytical models based on their differences.

Perceptual Speech Quality Measure (PSQM) [40] and its replacement Perceptual Evaluation of Speech Quality (PESQ) [41] are examples of intrusive voice quality measurements recommended by ITU. PESQ can directly provide MOS scores as indications of quality and it is now widely applied for off-line evaluations. However, intrusive measurements are not designed for on-line evaluations because it is very difficult to pass the original samples to the receiver in real-time. As a result, the intrusive measurements are not considered in details in this research.

2.2.3. Non-intrusive Measurements

Apart from the encoding-decoding process of analogue signals using codecs, most of the impairment on the quality of voice service comes from network transmission. The original speech may be distorted when being delivered over the network due to factors such as transmission delay, jitter and packet loss. The effect of one way transmission delay and packet loss on the voice quality (MOS score) is investigated in some papers for example [42-44]; it is indicated that speech quality can be estimated based on these parameters of network performance. These types of methods are so-called non-intrusive measurements which enable the on-line evaluation of the voice service quality since no original sample is required for reference.

The ITU introduced a computational model, the E-Model in document G.107 [45] for predicting the voice service quality using the average transmission latency and packet loss ratio. In the E-Model, the voice service quality delivered over the network is estimated as:

$$\boldsymbol{R = R_0 - I_s - I_d - I_{e-eff} + A} \qquad \textbf{\textit{Equation 2.1[45]}}$$

Where, R is a value between 0 and 100 indicating the voice quality, R_0 denotes signal to noise ratio at the receiver, I_s denotes the simultaneous impairments to the voice, I_d denotes the effect of transmission delay, I_{e-eff} denotes the Impairment due to Equipment Effect including random packet loss, A is an advantage factor of user's expectation compared to the conventional wired telephony service.

The value of R indicates the estimated quality of speech delivered over the network (see Table 2.5) which also gives corresponding MOS scores for different ranges of R values. It is indicated that the quality of speech with R value below 50 (MOS score below 2.58) is not acceptable to any user.

Table 2.5 Indication of R Value [41]

R	Quality	MOS
90-100	Best	4.34-4.5
80-90	High	4.03-4.34
70-80	Medium	3.6-4.03
60-80	Low	3.1-3.6
50-60	Poor	2.58-3.10
Below may not possible to maintain the voice conversation		

The most concerning issues for a VoIP service over the computer network is delivering as many voice packets as possible with minimum delay. In the ITU recommendation G.107, after introducing a set of default values, the calculation of R values is simplified as:

$$\boldsymbol{R = 93.2 - I_d - I_{e-eff}} \qquad \textbf{\textit{Equation 2.2 [45]}}$$

By using simplification, the voice service quality estimation is only concerned with the network transmission delay and packet delivery rate. There is much research on voice service quality such as [46, 47] based on this equation. Paper [47] provided a list of corresponding delay impairments (I_d) calculated from selected one way delay conditions using the analytical models provided by G.107 [45] (see Table 2.6).

Table 2.6 One Way Delay with Corresponding Delay Impairment I_d [47]

One Way Delay (ms)	$\mathbf{I_d}$
0	0
25	0.9
50	1.5
75	2.1
100	2.6
125	3.1
150	3.7
175	5.0
200	7.4

The ITU documentation G.113 Appendix I [37] provided some provisional values of equipment impairment I_{e-eff} using G.729 (see Table 2.7) and G.711 (see Table 2.8) codecs in different packet loss conditions. G.113 assumes that both codecs use their built-in PLC techniques to mask the effects of packet loss. However G.113 has not provided the I_{e-eff} value using G.729 in bursty packet loss conditions or a clear definition of 'bursty packet loss' for G.711 codec [47].

Table 2.7 I_{e-eff} Values using G.729 in Random Packet Loss Conditions [37]

Packet Loss %	I_{e-eff} Values (Random Loss)
0	11
0.5	13
1	15
1.5	17
2	19
3	23
4	26
8	36
16	49

Table 2.8 I_{e-eff} Values using G.711 in Different Packet Loss Conditions [37]

Packet Loss %	I_{e-eff} Values (Random Loss)	I_{e-eff} Values (Bursty Loss)
0	0	0
1	5	5
2	7	7
3	10	10
5	15	30
7	20	35
10	25	40
15	35	45
20	45	50

Tables 2.7 and 2.8 taken from G.113 Appendix I [37] show that when the packet loss ratio increases the equipment impairment (I_{e-eff} value) to the perceived speech quality increases significantly for both codecs. Furthermore, G.729 which provides higher compression to the original voice samples is more vulnerable to packet loss in the same packet loss condition compared to G.711. Table 2.8 also shows that when the packet loss ratio is low such as less than 3% there is no noticeable difference in I_{e-eff} values between random packet loss and bursty packet loss for G.711. However when the packet loss ratio increases (i.e. greater than 5%) the G.711 built in PLC techniques cannot cope with longer segments of bursty packet loss. It results in the I_{e-eff} values for bursty packet loss being higher than for random packet loss.

Paper [47] derived numerical expressions to fit the I_{e-eff} values in Tables 2.7 and 2.8 according to the probability of packet loss as:

$$\boldsymbol{I_{e-eff}(G.729\ random) \approx 11 + 40\ ln(1 + 10\ e)} \qquad \textit{\textbf{Equation 2.3 [47]}}$$

for G.729 codec in random loss conditions and

$$\boldsymbol{I_{e-eff}(G.711\ random) \approx 0 + 30\ ln(1 + 15\ e)} \qquad \textit{\textbf{Equation 2.4 [47]}}$$

for G.711 codec in random loss conditions and

$$I_{e-eff}(G.711\ burst) \approx 0 + 30\ ln(1 + 15\ e) \qquad \textit{Equation 2.5 [47]}$$

for G.711 codec in burst loss conditions,

where e is the probability of packet loss.

Equations 2.3-2.5 allow the derivation of I_{e-eff} values by simply inputting one network performance parameter, the packet loss ratio. Having both I_d and I_{e-eff} values ready, the perceive speech quality can be estimated using Equation 2.2.

2.3. VoIP over WLAN

As mentioned in Chapter 1, the usage of VoIP applications is not only over wired networks but increasingly over wireless networks. This is due to the growing popularity of handheld and mobile devices and the maturity of the wireless communication technology. Wireless networks can be classified into different categories according to their network coverage. The Wireless Wide Area Networks (WWANs) such as GSM (2G), UMTS (3G), LTE (4G) and WiMax have large coverage areas normally up to tens of kilometres of range in a single hop. These types of networks are generally used for providing mobile telephony service and internet access to handheld and mobile devices. The Wireless Local Area Networks (WLANs) such as IEEE802.11 (WiFi) have a coverage range from tens of meters to a few hundred meters in a single hop. These types of networks are usually used for local data communication as the replacement for wired Ethernets. The Wireless Personal Area Networks (WPANs) such as BlueTooth and ZigBee have a coverage range within ten meters in a single hop. These types of networks are usually used for data transfer between mobile and hand held devices.

This research focuses on the VoIP application over WLANs (VoWLANs). The WLANs, especially the IEEE802.11 standards have developed rapidly for laptop usage. There is now another major jump in usage as smartphones with WiFi have become popular. Many domestic users are using wireless routers to share internet access between devices at home. Wireless hot spots are now distributed worldwide by service providers, providing internet access where available, with much higher data rates than cellular mobile telephony networks can provide. Furthermore, WLAN is now also considered as a networking solution to provide LANs service in larger scenarios such as open plan offices, conference halls and temporary event venues. A side effect of the popularity of WLANs is that the VoIP applications are moving from wired networks on to WLANs. The following text will briefly introduce VoWLANs.

2.3.1. IEEE802.11 WLANs

Most of the off the shelf WLAN devices implement IEEE802.11 standards in order to be compatible with the other WLAN devices manufactured by different vendors. IEEE802.11 standards, also known as WiFi, are a group of open standards that define physical and logical specifications in order to enable network devices communicating with each other to use wireless signals. The IEEE802.11 standard family mainly works at the physical and data link layers.

The topologies of the IEEE802.11 WLANs are either infrastructure or ad hoc. In infrastructure topology, there is an AP acting as a central point of control in the WLAN. The switching and routing functions of the infrastructure WLANs are built in the APs. All network traffics to or from WLAN STAs are via the AP. By contrast, there is no central control in ad hoc WLANs. All STAs manage themselves in order to participate into the network and directly communicate with other STAs.

The original IEEE802.11 standard defines the Point Coordination Function (PCF) for infrastructure networks and the Distributed Coordination Function (DCF) for distributed networks [3]. However, the DCF is also implemented in most of the off the shelf wireless routers in the market. In other words, such APs are also acting as a STA in an infrastructure network. Hence, the PCF is not considered in this research.

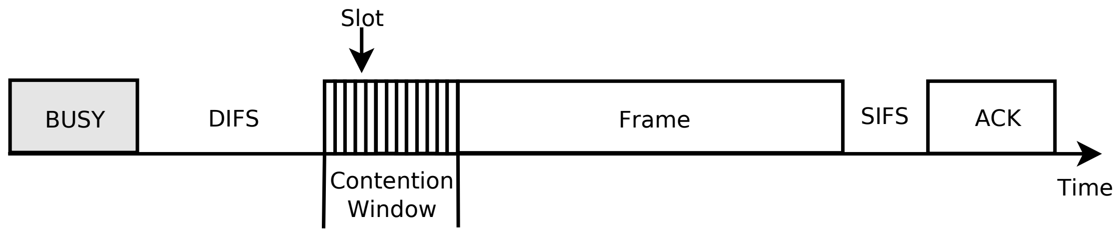

Figure 2.3 CSMA/CA Scheme

IEEE802.11 DCF uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) scheme to define how wireless nodes share the wireless channel [3] (see Figure 2.3). A wireless station can transmit immediately if, when a data packet (frame) arrives for transmission from the higher network layers, the medium has been idle for longer than the period called DCF Inter-Frame Space (DIFS). As shown in Figure 2.3 if the wireless channel is sensed as being busy when the frame arrived, the station must defer its transmission for a back-off time which is a random number of slots within the Contention Window (CW). If there are multiple stations deferring at the same time the one which first finishes the backing-off wins the transmission opportunity. The other deferring stations will pause their back-off counters whenever they sense the wireless channel being busy again. This mechanism avoids several stations starting transmission at the same time (collision). After a frame is successfully received, the receiver is going to wait for a period called Short Inter-Frame Space (SIFS) and send an acknowledgement (ACK) to the transmitter.

The most common IEEE802.11 standards are IEEE802.11a, b, g, e, n, s. IEEE802.11a, b, g [13-15] are the amendments of the original standard to use better modulation schemes at the physical layer in order to achieve higher data rates up to 54Mb/s. IEEE802.11a defines the wireless radio operates in 5GHz band. IEEE802.11b and g define the wireless radio operates in 2.4GHz band.

IEEE802.11e [48] defines some enhancements in the quality of service (QoS). The improvement to QoS is not achieved by enhancing network performance but by dividing the network traffic into categories according to different QoS requirements from the applications and prioritising the network traffic with higher QoS requirements. Table 2.9 shows the four QoS categories defined by the IEEE802.11e standard. Its Enhanced Distributed Channel Access (EDCA) function defines four buffers at the Logical Link Control (LLC) sub-layer within the data link layer. The higher priority traffic will have a smaller contention window and therefore it is more likely to win the transmission opportunity when the other contending nodes with lower priority traffic have larger contention windows. Furthermore, IEEE802.11e EDCA also introduces a scheme called Transmission Opportunity (TXOP) for higher priority traffic. After a node successfully transmits one high priority frame, it is allowed to transmit several subsequent frames from the same priority buffer straight away without contending for another transmission opportunity.

Table 2.9 IEEE802.11 Priority Categories [48]

Priority	Buffer Name	Example Applications
1	Voice (AC_VO)	VoIP
2	Video (AC_VI)	Video Streaming
3	Best Effort (AC_BE)	Web Browsing
4	Background (AC_BK)	Network Management

IEEE802.11n [49] makes some significant amendments to the original IEEE802.11 standard. At the physical layer, it further improves the data rate up to 600Mb/s (in theory) using Multiple Input Multiple Output (MIMO) antennas and a wider wireless channel. At the data link layer, IEEE802.11n introduces frame aggregation schemes to improve the transmission efficiency.

IEEE802.11s [50] defines the security and interconnection (routing) amongst WLANs over multiple hops to create a mesh network. Its security and routing protocols work between the data link layer and the network layer.

2.3.2. Existing Analysis on VoIP Capacity in WLANs

There is much existing research showing that the IEEE802.11 WLANs are not suitable for carrying VoIP services. The major finding in all of them is that the wireless channel utilising efficiency is extremely low for transmitting VoIP traffic. The number of simultaneous VoIP calls is very limited within IEEE802.11 WLANs although the data rate used nowadays can be very high. The text below provides a brief overview on the existing analyses related to this research.

Garg and Kappes in paper [2] provided a numerical estimation on the upper bound of VoIP capacity over WLANs using IEEE802.11a/b. Their results show that the maximum numbers of simultaneous VoIP calls are 12 using G.711 and 14 using G.729 for IEEE802.11b; 56 using G.711 and 64 using G.729 for IEEE802.11a. They also indicate that the VoIP capacity can increase if the packet interval is lower.

Garg and Kappes also implemented a real device experiment on real devices in [4]. The experiment was conducted in a single cell IEEE802.11b infrastructure WLAN containing one AP and eight nodes (PCs). The data rate they used was 11Mb/s. Their results show that this network can support up to 6 simultaneous VoIP calls using G.711 with 10 ms packet interval. This result conformed to their previous numerical analysis. They also found that the global network data throughput for running VoIP applications is very low due to the low wireless channel utilise efficiency.

In paper [5], Anjum et al. conducted a similar real device experiment as Garg and Kappes but concluding a maximum of eight supported VoIP calls (G.711, 10 milliseconds packet interval). They also developed a scheme to prioritise the transmission at the AP without backing off, called Backoff Control with Prioritised Queuing (BC-PQ). This scheme increases the maximum supported VoIP calls to ten.

Hole and Tobagi in their paper [6] analysed the VoIP capacity over IEEE802.11b infrastructure WLANs using simulation. Their results show that this network can support up to 12 G.711 or 14 G.729 VoIP calls with 20 ms packet interval. They also show that the wireless channel conditions such as the Bit Error Rate (BER) will affect the VoIP capacity over WLANs.

Wang et al. [12] studied the VoIP performance problem in both numerical analysis and simulation. They concluded the VoIP call capacity is about 11 for both G.711 and G.729 codecs. They proposed a solution to aggregate and broadcast the downlink traffics to all nodes by the AP, which can increase the capacity by up to 21.

Hegde at al. [7] conducted similar simulations on the IEEE802.11b infrastructure WLANs but concluded with the VoIP capacity in 14 G.711 calls and 17 G.729 calls. They also found that reducing Minimum Contention Window (CW_{min}) can significantly increase the VoIP capacity. Their simulation shown that the WLAN can support up to 22 VoIP calls if CW_{min} is set to eight.

Elaoud et al. [8] also conducted real device experiments on IEEE802.11b infrastructure WLANs containing 15 wireless nodes and found the VoIP capacity in this network to be 7 (G.711, 10 millisecond packet interval).

Cai et al. [9] used numerical analysis and real device experiments to find out that IEEE802.11b infrastructure WLANs can support 11 G.711 calls and 13 G.729 calls. They also indicated that the traffic loads between uplinks and downlinks in infrastructure WLANs are unbalanced.

Shin and Schulzrinne [10] conducted real device experiments on the IEEE802.11b infrastructure WLAN test-bed with the IEEE802.11e enabled which containing 64 wireless nodes and supported a maximum 15 G.711 calls simultaneously. They observed a significant increase in transmission latency and packet loss rate when additional calls are added.

Stoeckigt and Vu [11, 51] carried out both numerical analysis and simulations on IEEE802.11b infrastructure WLANs with IEEE802.11e enabled. Their results show that the IEEE802.11e EDCAT XOP scheme at AP in can significantly increase the VoIP capacity in the WLAN up to twice the capacity without this function.

There are few analyses [52, 53] of VoIP capacity on MANETs[1]. However, most of them were investigating the network layer issues such as, routing which will not be considered in this thesis.

Table 2.10 Existing Research on VoWLAN Capacity

Publication	Analysis Method	IEEE802.11 WLAN	Capacity	Major Findings
[2]	Numerical	b AP	G711 12 G729 14	N/A
[4]	Real Device	b AP 4 nodes	N/A	The transmission overhead is too large for voice packets.
[5]	Real Device	b AP 10 nodes	N/A	If the wireless channel becomes congested, the packet loss ratio for VoIP increases dramatically.
[6]	Simulation	b AP	G711 12 G729 14	BER affects strongly on the voice service quality.
[12]	Simulation	b AP	G.711 12	Multicasting aggregated downlink voice packets by the AP may improve the VoWLAN capacity.
[7]	Simulation	b AP	G.711 13	Using smaller CWmin may improve the voice service quality.
[8]	Real Device	b AP	G.711 7 (10 ms intv)	One way transmission latency increases when the number of call increases.
[9]	Numerical Real Device	b AP	G711 11 G729 13	The downlink and uplink VoIP traffics have unbalanced transmission opportunities in an infrastructure WLAN.
[10]	Real Device	b, e AP	G.711 15	If the wireless channel becomes congested, the one way transmission latency and the packet loss ratio for VoIP increase.
[11, 51]	Numerical Simulation	b, e AP	G.711 12 G729 16 (10 ms intv)	Using 802.11e TXOP and smaller CWmin may improve the voice service quality.
[52]	Simulation	b MANET multi-hops	G.711 8 G729 10 (1 Hop)	N/A
[53]	Simulation	a MANET	N/A	N/A

[1] Mobile Ad Hoc Network

Table 2.10 provides a summary of the existing VoWLAN capacity research mentioned in this section. In terms of the analytical methods used by the existing analyses, they can be classified as numerical analysis, simulation and real device experiments. Most of the existing analyses are based on IEEE802.11b. IEEE802.11a and e are also studied by some of the existing research. In terms of the WLAN topology, most of the existing research was based on infrastructure WLAN (labelled by 'AP') except those analyses on the network layer issues for MANETs. G.711 and G.729 are the most common codecs used as examples for studies. Therefore this research also focuses on these two codecs.

The overall major problems found include:

1) There is a large overhead for WLANs transmitting VoIP traffic;
2) The impairments such as packet loss, latency and jitter significantly increase after the WLANs become saturated;
3) There is unbalance between uplink and downlink in infrastructure WLANs.

2.3.3. Network Simulator (NS) 2

Network Simulator 2.33 (NS-2) [65] was used as the main platform to test and validate solutions developed in the rest of this research. NS-2 is an open source discrete event network simulator widely used in academic research on both wired and wireless networks. It was implemented using C++. The users use Object-TCL (OTCL) to configure and write the script of a simulation. The network components at each layer were implemented similar to real world cases. At the physical layer, NS uses various statistical models to simulate the signal propagation for wireless transmissions. All the simulations based analysis mentioned in Chapters 3, 4 and 5 were done using NS-2.

2.3.4. Research Scenario

Most of the existing research was carried out based on the infrastructure topology of WLANs which was discussed in the previous section (2.3.2). As introduced in Section 1.1.1, this research is concerned with supporting as many VoIP calls as possible over each wireless channel. Therefore we only focus on the number of voice packets that can be transmitted in time for the VoIP application over each wireless channel. Hence, a fully distributed topology is used in this scenario. As shown in Figure 2.4, VoIP connections (calls) are set up between pairs of wireless nodes. It needs to clarify that the reason why the term 'ad hoc' is not used in this thesis is to emphasis the assumption that pairs of VoIP connections are independent from each other.

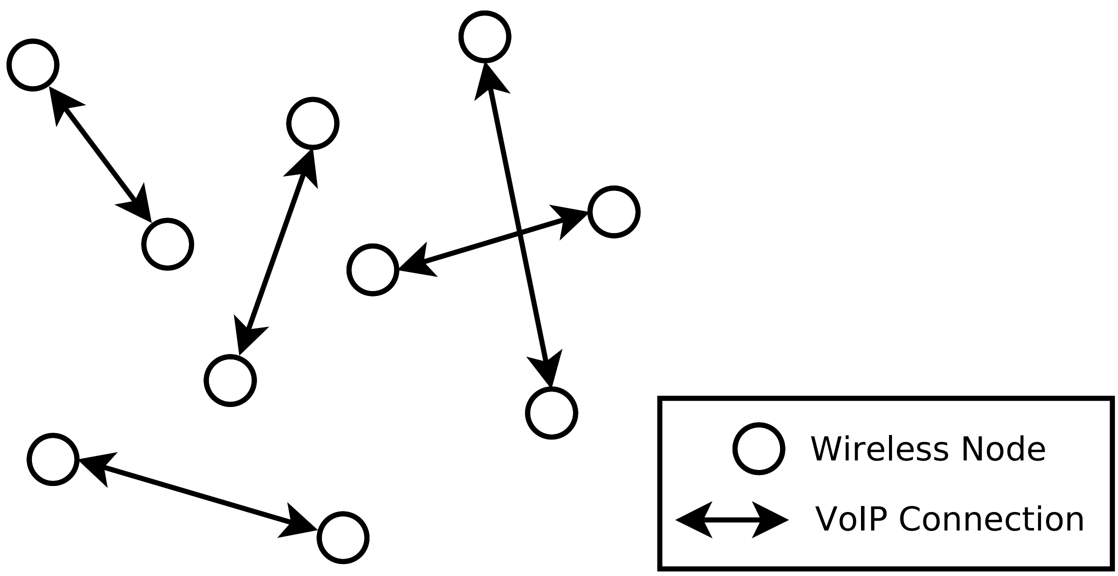

Figure 2.4 Fully Distributed Topology

Furthermore, the distributed WLANs scenario does not only represent the use of wireless ad hoc networks for VoIP in disaster places or battle fields where there could be no AP setup. It can also represent cases where there are VoIP calls which belong to different independent infrastructure WLANs. For instance, as shown in Figure 2.5, in an apartment building, each household may have a wireless internet router installed as the AP and there is a person in each household using a VoIP service on his/her smartphone over the internet via the AP. The WLANs and the VoIP connections in different households are independent from each other. This topology is also considered as 'distributed WLANs'.

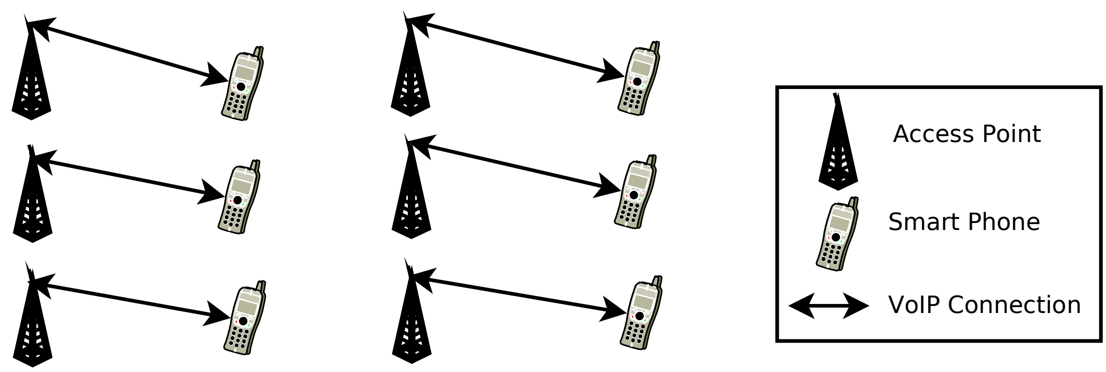

Figure 2.5 Independent WLANs in an Apartment Building Carrying VoIP Calls

IEEE802.11g using 11Mb/s and 54Mb/s data rates are selected as the WLAN configurations. The IEEE802.11a and n is delivering high network performance. However, most of this research is based on the saturation of the network performance. IEEE802.11a and n require significantly more network activities to saturate the WLANs. Due to the limitation of the computational power in simulation and the availability of real devices IEEE802.11a and n are not considered as examples in this research.

2.4. Summary

This chapter has provided the background knowledge which is required for this research. The VoIP application has a very different data traffic pattern compared to non-real-time network applications such as web browsing and emails. In order to reduce the round trip delay during VoIP conversations, the codecs send out voice packets as soon as they have a few samples recorded. This results in a constant stream of small voice packets being sent over the WLAN. When voice packets are being lost, the codecs use PLC techniques to produce synthetic voice samples to mask the effects of the packets lost.

The MOS scoring method recommended by the ITU is widely used to indicate the quality of speech heard by the end users. However, the quality of VoIP service is very subjective to users. The ITU also recommends some analytical models which can estimate the perceived speech quality transferred over the network by inputting some network performance parameters such as transmission latency and packet loss rate. These models allow people to quickly estimate the VoIP service quality without actually making voice calls.

Many existing studies have estimated that the VoIP capacity of a IEEE802.11 b/g WLAN is about 11 to 14 simultaneous calls. Most of them were carried out based on the scenario of a single infrastructure WLAN. There are also some studies using MANETs, but they focus on routing issues at the network layer and have not provided results that are directly useful for to this research. However, in for example, shared accommodation buildings such as offices and apartments there are more likely to be multiple independent WLANs which operate close to each other and share a wireless channel. Therefore, a fully distributed WLANs scenario is used in this research.

3. Problem Analysis

The previous chapter (Chapter 2) provides the background information required in order to carry out this research and reviews the major findings of existing research on VoWLAN capacity. Most of the existing research analysed a single infrastructure WLAN. The purpose of this chapter is to specify, recognise and analyse the problems need to solve in order to support better VoIP service over distributed WLANs. First of all, it will demonstrate that problems found using VoWLAN on a single infrastructure WLAN are also applicable to distributed WLANs. Furthermore, to extend the research finding by Malone et al. in paper [18] this chapter will also analyse the reasons why the network performance for VoIP collapse soon after the wireless channel becomes congested.

This chapter is structured as follows. Section 3.1 introduces the simulation parameter configurations used to represent the research scenario described in Section 2.3.4. Section 3.2 derives the voice service qualities (the R values introduced in Section 2.2.3) for using G.729 and G.711 codec based on different combinations of network performance parameters (packet loss ratio and one way transmission delay) using Equation 2.2-2.5 introduced in Section 2.2.3. The lowest network performance to support minimum acceptable voice service qualities will also be identified. The problem analysis will then be carried out in Section 3.3 to 3.5, which includes:

1) Section 3.3 follows the analytical approaches found in existing studies for infrastructure WLANs, performs a numerical analysis for distributed WLANs to estimate the efficiency of VoIP packet transmissions in IEEE802.11

distributed WLAN, calculates the capacity of VoIP calls in a single IEEE802.11 wireless channel and the impact when the number of simultaneous VoIP calls exceeds the capacity (in congested network conditions);

2) Section 3.4 presents an experiment on real 802.11b/g devices that was done to confirm the correctness of the numerical analysis and estimation for the transmission efficiency and channel capacity;
3) Section 3.5 presents the extensive simulation experiments that were designed and carried out to mirror the real device experiment and confirm the correctness of the numerical analysis on the impact of WLAN congestion.

3.1. Scenario Configurations

This section introduces the default parameter configurations used throughout this research. The analyses in the rest of this thesis are also based on these configurations in Table 3.1 unless otherwise specified.

A typical VoIP call in this thesis uses either the G.711 codec and generates 160 byte voice data packets [30] or uses the G.729 codec and generates 20 byte voice data packets [29]. Voice data packets are generated and sent in every 20 ms bidirectional between two end users (wireless nodes). A normal Ethernet data packet with typically 1,500 byte payload as Maximum Transmission Unit (MTU) will be used for comparison.

The WLAN parameters are configured according to the specifications from the IEEE802.11b and g standard DCF (see Table 3.1). The Minimum Contention Window ($\mathrm{CW_{MIN}}$) for IEEE802.11a, g and n should be 15. However, in order to compare with

IEEE802.11b which is considered in most of the existing research papers found [2, 4, 6, 7, 9-11, 51], the $\mathrm{CW_{MIN}}$ is also configured to 31 for IEEE802.11g here. Two data rates 11Mb/s and 54Mb/s are used for both voice and Ethernet data packets. The comparison of performance using the two chosen data rates will demonstrate whether significant improvement occurs by increasing the data rate used in transmissions. We assume for simplicity that the all wireless stations are covered by a single interference area and there is no other network or interference within this area.

Table 3.1 Parameters in IEEE 802.11 (Header Rate 1 Mb/s)

Parameters	Time (sec)	Size (bytes)
Preamble +PLCP Header(Data and ACK, 2 frames)	384	48
MAC Header + CRC	24.73	34
IP+UDP+RTP	29.09	40
T_{DATA} **(G.711 @ 11Mb/s)**	145.5	160
T_{DATA} **(G.711 @ 54Mb/s)**	29.6	160
T_{DATA} **(G.729 @ 11Mb/s)**	18.2	20
T_{DATA} **(G.729 @ 54Mb/s)**	3.7	20
T_{DATA} **(Ethernet Data @11Mb/s)**	1363.6	1500
T_{DATA} **(Ethernet Data @54Mb/s)**	222.22	1500
T_{ACK}	10.18	14
T_{SIFS}	10	
T_{DIFS}	50	
T_{SLOT}	20	
$\mathbf{CW_{MIN}}$	31 slots	
Outgoing Buffer size	50 packets	

3.2. Minimum Acceptable Network Performance

Equation 2.2 introduced in Section 2.2.3 provides a simplified estimation on voice service quality (the R value). For each given codec, there are two parameters that need to be provided, the one way transmission latency and the packet loss ratio. However, they are not directly used in Equation 2.2. As mentioned in Section 2.2.3, paper [47] provided the conversion methods from the one way transmission latency and the packet loss ratio to the terms I_d and I_{e-eff} respectively in Equation 2.2.

I_d in Equation 2.2 relates to the one way transmission latency over the network. Table 2.6 listed the I_d values provided by paper [47] corresponding to the one way transmission latency in milliseconds.

I_{e-eff} in Equation 2.2 relates to the packet loss ratio of the transmissions. The I_{e-eff} values according to some selected packet loss probabilities were calculated in Table 3.2 using Equation 2.3-2.5 respectively in Section 2.2.3.

Table 3.2 I_{e-eff} Values Calculated using Equation 2.3-2.5

Packet Loss Probability	$G.729\ random$	$G.711\ random$	$G.711\ burst$
0	11	0	0
%5	27.2	16.8	28.6
%10	38.7	27.5	39.5
%15	47.7	35.4	46.4
%20	54.9	41.6	51.4

We used Equation 2.2 and the available I_d and I_{e-eff} values to calculate the voice service qualities (the R values) corresponding to different combinations of network performance parameter values: the one way transmission latency and the packet loss ratio.

Table 3.3 represents the voice service qualities (the R values) calculated in this research using the G.729 codec with random packet loss. The numbers show that increases in packet loss have a more significant effect on the voice service quality than increases in transmission latency. It also shows using Table 2.5, that the lowest acceptable service quality (R=50.5 corresponding to poor quality and a MOS of 2.58 to 3.10) occurs when the one way transmission delay is 150 ms and the packet loss is 10%.

Table 3.3 G.729 R Values for various Network Performance Parameters (Random Packet Loss)

	One Way Transmission Delay							
Packet loss	**25ms**	**50ms**	**75ms**	**100ms**	**125ms**	**150ms**	**175ms**	**200ms**
0%	81.3	80.7	80.1	79.6	79.1	78.5	77.2	74.8
5%	65.1	64.5	63.9	63.4	62.9	62.3	61	58.6
10%	53.6	53	52.4	51.9	51.4	50.8	49.5	47.1
15%	44.6	44	43.4	42.9	42.4	41.8	40.5	38.1
20%	37.4	36.8	36.2	35.7	35.2	34.6	33.3	30.9

Table 3.4 represents the voice service quality calculated in this research using G.711 codec with random packet loss. The numbers show that using the G.711 codec is more robust to packet loss than using G.729 with R≈50 only occurring with packet loss of 20%. This may be due to the fact that compression by G.729 is eight times that of G.711. The loss of a voice packet is more significant using G.729 when compared to G.711. Furthermore, as introduced in Section 2.1.4 G.711 has more complex and therefore more robust PLC techniques built in. It can deal with up to 20% packets being randomly lost in the transmission.

Table 3.4 G.711 R Values for various Network Performance Parameters (Random Packet Loss)

	One Way Transmission Delay							
Packet loss	**25ms**	**50ms**	**75ms**	**100ms**	**125ms**	**150ms**	**175ms**	**200ms**
0%	92.3	91.7	91.1	90.6	90.1	89.5	88.2	85.8
5%	75.5	74.9	74.3	73.8	73.3	72.7	71.4	69
10%	64.8	64.2	63.6	63.1	62.6	62	60.7	58.3
15%	56.9	56.3	55.7	55.2	54.7	54.1	52.8	50.4
20%	50.7	50.1	49.5	49	48.5	47.9	46.6	44.2

Table 3.5 G.711 R Values for various Network Performance Parameters (Burst Packet Loss)

	One Way Transmission Delay							
Packet loss	**25ms**	**50ms**	**75ms**	**100ms**	**125ms**	**150ms**	**175ms**	**200ms**
0%	92.3	91.7	91.1	90.6	90.1	89.5	88.2	85.8
5%	63.7	63.1	62.5	62	61.5	60.9	59.6	57.2
10%	52.8	52.2	51.6	51.1	50.6	50	48.7	46.3
15%	45.9	45.3	44.7	44.2	43.7	43.1	41.8	39.4
20%	40.9	40.3	39.7	39.2	38.7	38.1	36.8	34.4

Table 3.5 represents the voice service quality calculated in this research using G.711 codec with bursty packet loss. These numbers show that bursty packet loss is more significant than random packet loss for G.711. Comparative figures for bursty packet loss with G.729 have not been found, so are not discussed here.

In this research, the simplification of the voice service quality measurement was derived from the above analysis. Furthermore, in the G.114 documentation [17] ITU recommended that the one way transmission latency should not be greater than 150 ms for the minimum acceptable VoIP service quality. Our calculations of R values show that in order to deliver voice service with the quality acceptable to users using

G.729 in random packet loss conditions or G.711 in bursty packet loss conditions, the network needs to deliver no less than 90% of the voice packets within 150 ms delay budget to the receiver. We use these criteria in this research to judge whether the usable quality VoIP calls can be maintained for users. In order to support better VoIP service over WLANs, the algorithms and protocols proposed in Chapters 4 and 5 are aiming to increase the packet delivery ratio within this delay budget.

However, the voice service quality is still subjectively judged by each individual user in the real application. The standard of judgment is varying according to each user's opinion and affected by many other factors such as the physical environment, languages used and talking speeds. In the real world, there is no hard boundary to indicate whether the voice service quality is acceptable to users.

3.3. Numerical Analysis

3.3.1. VoIP Packet Transmission Efficiency over 802.11

As introduced in Section 2.3.1, IEEE802.11 DCF specifies that wireless stations use the CSMA/CA scheme to access a shared wireless channel for transmission. The DIFS, SIFS, the time spent for transmit contention and ACK shown in Figure 2.3 in Section 2.3.1 are considered as the overheads caused by the CSMA/CA scheme.

Furthermore, there are also encapsulation overheads within an IEEE802.11 frame. The overheads are caused by the packet headers and trailers that are used to encapsulated the application data by the protocol at each network layer passing from the application down to the physical layer. Figure 3.1 represents the frame structure and its overheads in the frame used for transmitting a voice packet. The preamble and PLCP[2] header are the encapsulation overheads at the physical layer. It is important to note that in order to ensure the physical layer encapsulation information can be understood within the largest possible coverage area, they are transmitted using the lowest possible data rates, the so called base rates (1 Mb/s for IEEE802.11b/g [13, 14], 6 Mb/s for IEEE802.11a [15] and 6.5 Mb/s for IEEE802.11n [49]). The rest of the frame is transmitted using selected data rates. The MAC header and the Cyclic Redundancy Check results (CRC) are the encapsulation overheads at the data link layer. The IP, UDP and RTP headers are the encapsulation overheads at the network, transport and application layers respectively.

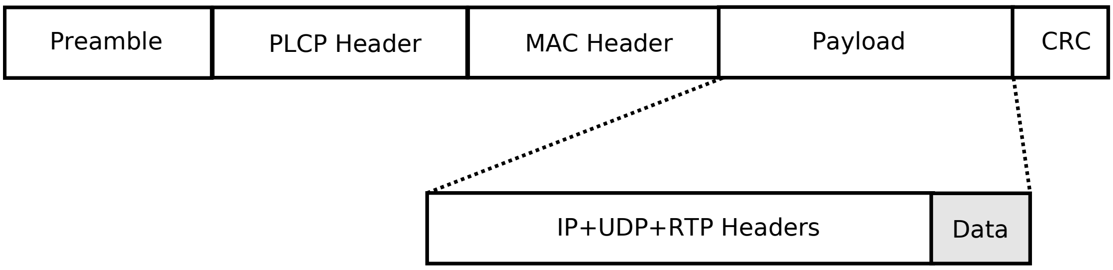

Figure 3.1 IEEE802.11 Frame Structure for VoIP

According to the definition of CSMA/CA scheme and IEEE802.11 frame structure, Equation 3.1 is derived here which uses the parameters listed in Table 3.1 to estimate the time (T_{PK}) to transmit a frame (packet) for one active wireless node using CSMA/CA scheme implemented in IEEE802.11 WLAN. Very similar equations can be found in papers [2, 4, 6, 7, 10, 12].

$$\boldsymbol{T_{PK} = T_{DIFS} + \left(T_{SLOT} \times \frac{CW_{MIN}}{2}\right) + T_{HEAD/TAIL} + T_{DATA} + T_{SIFS} + T_{ACK}} \qquad \textbf{\textit{Equation 3.1}}$$

[2] PLCP - Physical Layer Convergence Procedure

Where: $T_{HEAD/TAIL}$ *is the total time for transmitting PLCP Preamble, PLCP Header, MAC Header + CRC and IP+UDP+RTP;* $\frac{CW_{MIN}}{2}$ *is the average slots selected by the transmitter;* T_{DIFS}*,* T_{SLOT}*,* T_{DATA}*,* T_{SIFS} *and* T_{ACK} *are specified in Table 3.1.*

The calculation results are shown in Figure 3.2 and Table 3.6. It requires 2181.6 µsec and 1040.22 µsec to transmit a packet with 1,500 byte payload at 11Mb/s and 54Mb/s data rates respectively. This result shows that increasing the data rate from 11Mb/s to 54Mb/s can significantly reduce the transmission time (about 50%) for large data packets. However, the performance improvement is not significant for transmitting voice packets using a higher data rate, because the proportion of the overhead is too large (818 µsec). For example, for transmitting a G.729 packet requires 836.2 µsec using 11Mb/s data rate. If the data rate increases to 54Mb/s it requires 821.7 µsec. The improvement in total transmission time required is very miniscule (less than 2%).

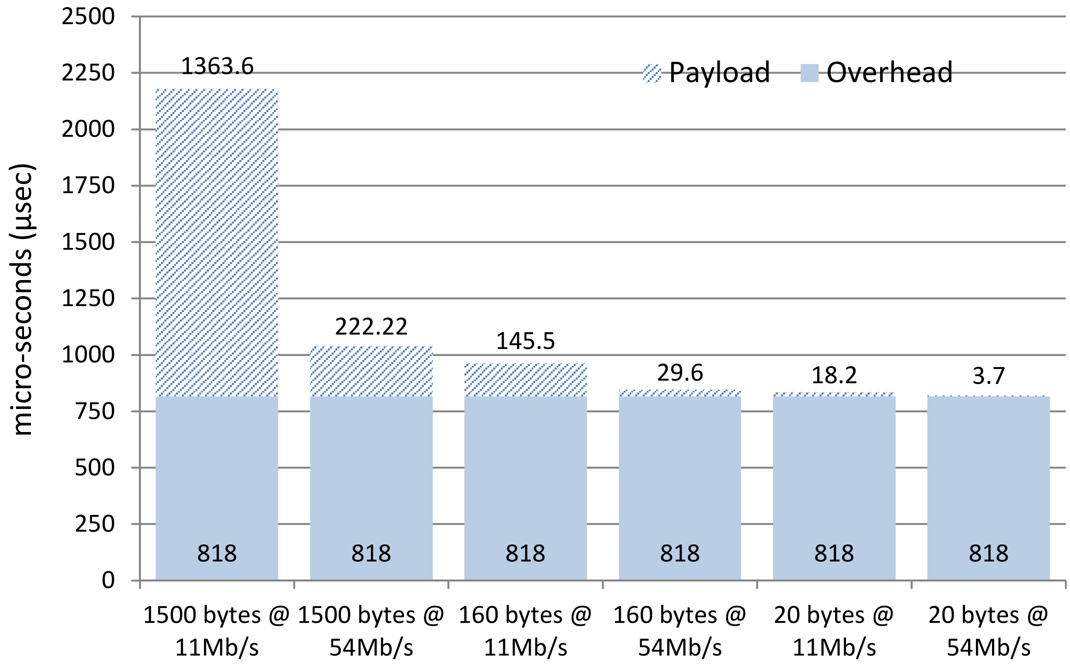

Figure 3.2 Durations for Transmission: Payloads versus Overheads(µsec)

Table 3.6 Calculated T_{PK} (μsec)

	1500 bytes @ 11Mb/s	1500 bytes @ 54Mb/s	160 bytes @ 11Mb/s	160 bytes @ 54Mb/s	20 bytes @ 11Mb/s	20 bytes @ 54Mb/s
T_{PK}	2181.6	1040.22	963.5	847.6	836.2	821.7

To transmit a packet with only a 160 byte payload, it requires 963.5 μsec and 847.6 μsec respectively at 11Mb/s and 54Mb/s. To transmit a packet with only a 20 byte payload it requires 836.2 μsec and 821.7 μsec at 11Mb/s and 54Mb/s data rates respectively. The results also show that if the voice sample is more compressed, e.g. using G.729 instead of G.711 does not significantly reduce the transmission time. In these transmission durations, 818 μsec is overhead. It indicates that it is much more efficient to transmit large payload packets compared to small payloads. These analysis results indicate that neither increasing the data rate nor reinforcing the voice data compression can provide a significant improvement for VoIP packet transmission efficiency.

3.3.2. VoWLAN Capacity Analysis

As discussed in Section 1.1.1, in situations where people gather, crowds as well as environments such as offices and shopping malls, calls are often close to each other. When there are few VoWLAN callers and very little other wireless traffic, there will be very little contention for access to the wireless channel (transmission opportunity) and therefore both the latency and the jitter should be low. If contention does occur in this low usage environment the latency for the frames delayed by contention for the transmission opportunities will be much higher. This will show up in a larger jitter for the contention frames compared to frames where no contention took place.

As the number of VoWLAN callers or other uses of the wireless medium increases and the medium gets busier, more stations will have to contend to transmit whenever they have data waiting for transmission. This contention will increase the transmission latency and reduce the number of frames being delivered in time, resulting in reduction to the VoIP service quality. When the demand for use of the wireless channel exceeds the capacity of the medium with the given configuration, the transmission latency and the packet delivery rate within the delay budget will result in effectively unusable VoIP service quality.

Equation 3.2 was derived in this research to estimate the maximum number (C_{MAX}) of simultaneous VoWLAN calls over distributed IEEE802.11 WLANs per wireless channel:

$$\boldsymbol{C_{MAX} = \frac{T_P}{2T_{PK}}} \qquad \textbf{\textit{Equation 3.2}}$$

Where: the voice packetizing interval T_P *is set to 20 ms for both G.711 and G.729 codec, in Equation 3.2.*

The calculation results of T_{PK} from Table 3.6 for both codecs and data rates are used and applied to Equation 3.2. The estimated theoretical VoIP capacity (C_{MAX}) for both data rates and codecs are listed in Table 3.7. The IEEE802.11 WLAN is estimated to support maximum of 10-12 simultaneous calls. The results also show that varying the data rates or using a higher compression ratio codec does not significantly improve the VoWLAN conversation capacity much.

Table 3.7 Calculated C_{MAX}

	(G.711) 160 bytes @ 11Mb/s	(G.711) 160 bytes @ 54Mb/s	(G.729) 20 bytes @ 11Mb/s	(G.729) 20 bytes @ 54Mb/s
C_{MAX}	10.37883	11.79802	11.95886	12.16989

3.3.3. Impact of Wireless Channel Congestion

As the number of VoWLAN callers or other uses of the network transmission increases and the wireless channel gets busier, more stations will have to contend for a transmit opportunity whenever they have packets waiting in the outgoing buffer. The contention for transmit opportunity will increase both the transmission latency due to waiting in the local and nearby node's outgoing buffer and the amount of jitter due to random back-off and the queue length in the outgoing buffer. Furthermore, the packet loss for VoIP application may increase due to more packets which may be delivered beyond the delay budget which will result in much reduced service quality. When the demand for use of the medium exceeds the capacity of the wireless channel with the given configuration, the probability increases that by the time the next frame arrives, the previous frame is still in the buffer waiting for transmission. The occupation of the outgoing buffer then grows by one. This situation may be caused by the wireless channel being congested by the traffic demands that exceed the capacity. This section will analyse what the effect will be for buffering under congested conditions. The analysis conclusion is expected to explain the observations of collapsing in network performance under congested conditions indicated in paper [18] that it goes from a low-latency high-delivery rate down to high-latency low-delivery rate.

Firstly, the wireless channel transmission capacity in packets per second will be calculated. Equation 3.3 is derived here to estimate the maximum number of voice packets ($P_{MAX/S}$) that the IEEE802.11 distributed WLAN can transmit per second.

$$\boldsymbol{P_{MAX/s} = \frac{1}{T_{PK}}}$$ ***Equation 3.3***

If there is no congestion in the wireless channel, the occupation of the outgoing buffer remains very low or empty. However, under congested conditions in distributed WLANs such as when the number of simultaneous VoIP calls exceeds 12, the wireless channel capacity ($P_{MAX/S}$) will be shared evenly amongst all the VoIP sources (active wireless stations) in the long term. The growth in occupancy of the outgoing buffer mentioned previously will happen. Equation 3.4 derived here estimates the growth rate ($G_{/S}$) in occupancy of the outgoing buffer for each VoIP source under congested conditions.

$$G_{/S} = \begin{cases} \boldsymbol{0}, & (\boldsymbol{N_{call} < 12}) \\ \boldsymbol{\frac{1}{T_P} - \frac{P_{MAX/s}}{N_{call} \times 2}}, & (\boldsymbol{N_{call} \geq 12}) \end{cases}$$ ***Equation 3.4***

Where: $\boldsymbol{\frac{1}{T_P}}$ *calculates the number of voice packets arriving at a node's outgoing buffer per second;* $\boldsymbol{N_{call}}$ *is the number of simultaneous nearby VoIP calls in the distributed WLANs; there are two VoIP sources (*$\boldsymbol{N_{call} \times 2}$*) for each VoIP call;* $\boldsymbol{\frac{P_{MAX/s}}{N_{call} \times 2}}$ *calculates the average amount of wireless channel capacity shared by each VoIP source.*

Figure 3.3 shows the estimated buffer growth rates with a different number of simultaneous G.711 VoIP calls in the congested network situation. It shows that the greater the number of the simultaneous VoIP calls, the more traffic demand exceeds the wireless channel capacity and hence, the faster growth of the number of packets in the outgoing buffer.

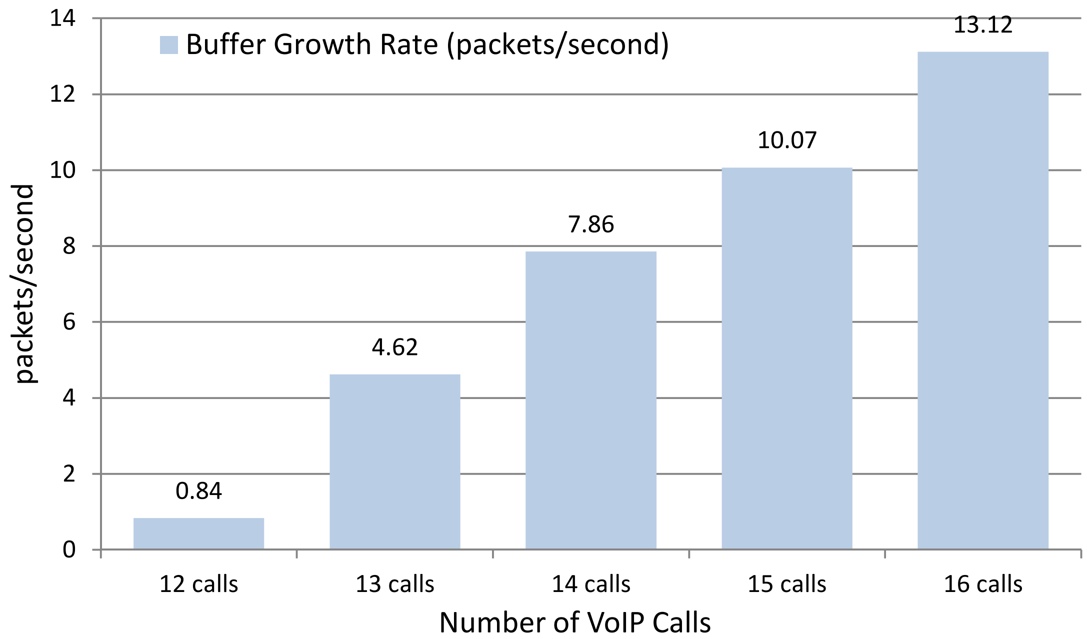

Figure 3.3 Estimated Buffer Growth Rate under Congestion (G.711 @ 54/MB/s)

Although the buffer growth rate can be different, the occupancy of the outgoing buffer will stop growth when the buffer is full. In this case, additional arriving packets will be dropped. As a result, the packet loss rate increases. If there are multiple packets in the outgoing buffer, the later arriving packets can only be transmitted once the earlier arriving packets are successfully transmitted. The queuing delay also contributes to the transmission latency. Equation 3.5 derived here estimates the queuing delay (D_{buf}) in seconds of a packet in an outgoing buffer under congested conditions:

$$D_{buf} = S_{buf} / \frac{P_{MAX/s}}{N_{call} \times 2}, \qquad (N_{call} \geq 12)$$ ***Equation 3.5***

Where: S_{buf} is the buffer occupation.

The occupancy (S_{buf}) of the outgoing buffer is growing under the congested conditions by the buffer growth rate ($G_{/S}$). The queuing delay (D_{buf}) is also increasing proportionally to the buffer growth rate ($G_{/S}$) until the outgoing buffer is full. Furthermore, if an infinite outgoing buffer is used, the queue delay can increase infinitely as long as the congestion exists. In this thesis, this phenomenon is the so called 'delay accumulation' during the growth of the outgoing buffer under congested conditions.

Table 3.8 Calculated D_{buf} (when S_{buf}=50, N_{call}=14)

	(G.711) **160 bytes @ 11Mb/s**	**(G.711)** **160 bytes @ 54Mb/s**	**(G.729)** **20 bytes @ 11Mb/s**	**(G.729)** **20 bytes @ 54Mb/s**
D_{buf}	1.35 sec	1.19 sec	1.17 sec	1.15 sec

Table 3.8 lists the queuing delay (D_{buf}) for an outgoing buffer with the size of 50 packets when are there 14 simultaneous VoIP calls active in the WLAN. The results show that there will be no voice packets that can be delivered within the delay budget of 150 ms in this thesis. It also explains why the network performance for VoIP traffic drops sharply under congested conditions.

3.4. Real Device Experiments

The results above are theoretical. To check the accuracy of the theory some real device experiments were carried out. Experiments on a real IEEE 802.11 distributed WLAN were carried out to estimate the real wireless channel capacity and observe the phenomenon of delay accumulation predicted in the Section 3.3.3. In these experiments, pseudo VoIP streams were generated according to the parameter configurations in Table 3.1 between pairs of nodes in a wireless network. Various wireless network performance metrics were recorded while the number of simultaneous pseudo VoIP streams was increased until the wireless channel becomes congested.

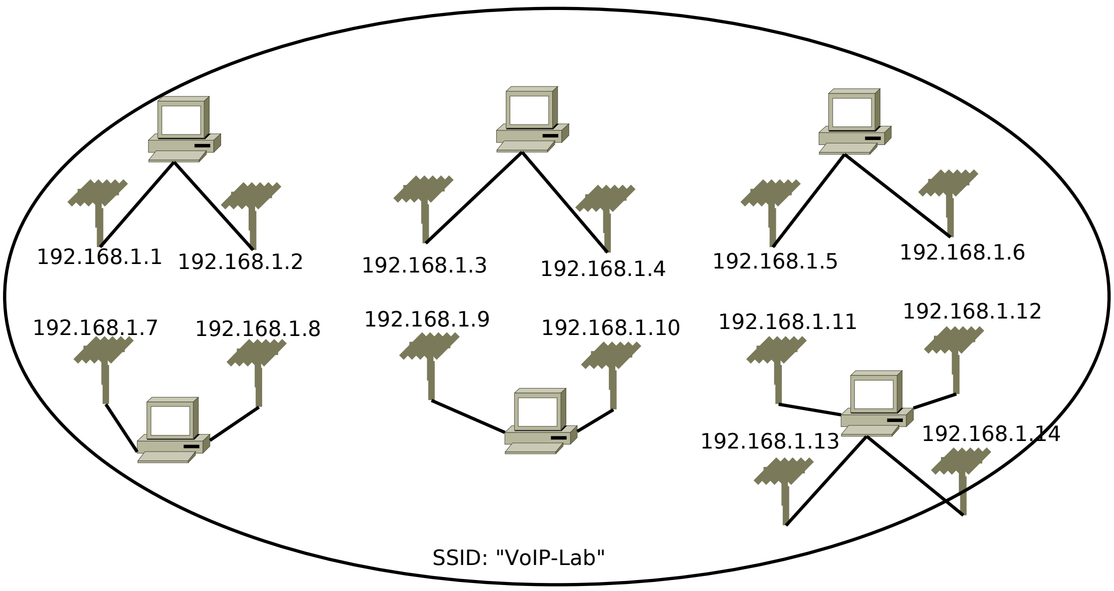

Figure 3.4 Real Device IEEE802.11 Distributed WLAN Test Bed

A test bed was built with 6 desktop computers (PCs) as shown in Figure 3.4. Each PC was running the Ubuntu 9.04 [54] operating system. Because of the limitation on the number of available PCs, multiple wireless Network Interface Cards (NICs) are installed in each PC. There were 14 wireless NICs in total. 5 PCs were installed with two wireless NICs each. The other PC was installed with four wireless NICs. Each wireless NIC was equipped with a complete network protocol stack as shown in Table 3.9. As a result, all NICs were working independently of one another. We assume that the computational resources (e.g. memory space and processing power) on the PCs were sufficient to not introduce extra latency or packet loss to the transmissions.

Table 3.9 Protocol Stack in Real IEEE802.11Test Bed Devices

Layers	Configuration
Application	JPerf 2.0.2 [55] and VoIP-RTT-Estimator running on Ubuntu 9.04
Transport Layer	UDP
Network Layer	Manually assigned IP address and static routes
Data Link Layer	Ad hoc mode, SSID: 'VoIP-Lab' IEEE802.11 DCF supported by Belkin F5D7050 WinXP driver wrapped by NDISwrapper [56]
Physical Layer	Belkin Wireless G USB Adapter F5D7050 [57]

At the physical layer and data link layer, a wireless ad hoc network with SSID 'VoIP-Lab' was created. We assume that, the use of wireless channel by nearby WLANs did not significantly affect the experimental results. The RTS/CTS[3] function was disabled. At the network layer, each node (NIC) was manually assigned with an IP address. The network was small so each source node was guaranteed to reach the destination within one hop. Static routes were provided in the IP table of the Ubuntu kernel on each PC. At the transport layer, each simulated stream was assigned with a unique UDP outgoing or incoming port at the source or destination respectively. Two applications were used and run on Ubuntu 9.04 in the experiments. JPerf [55] is an open source front end implementation in Java for IPerf [58]. VoIP-RTT-Estimator is a Java application implemented as part of this research for measuring pseudo VoIP stream round trip delay. Both of these experiments will be introduced in detail in the following subsections.

3.4.1. Wireless Channel Capacity Estimation using JPerf

In order to estimate the maximum capacity of a wireless channel, JPerf was used to generate bidirectional pseudo UDP VoIP streams between pairs of wireless nodes. The packet payloads generated by JPerf are 1500 byte (Ethernet MTU), 172 byte (as a 160 byte G.711 packet + 12 byte RTP header) and 32 byte (as a 20 byte G.729 packet + 12 byte RTP header). In each stream, 1500 packets were generated per second to guarantee the traffic load is significantly over the wireless channel capacity. Both the 11 Mb/s and the 54 Mb/s data rates were tested separately in these experiments. The number of packets transmitted per second by each stream is recorded and summed up for all the wireless nodes as the wireless channel capacity for each experimental configuration. Each experiment was repeated ten times with

[3] Request-to-Send/Clear-to-Send

the results presented being the average of these repeated experiments. In each experiment, the number of transmitters is gradually increased by 2 until all 14 nodes are assigned 2-way pseudo VoIP streams to transmit. Then duplicated streams start being added on existing transmitters until each node has 2 streams to transmit.

Figure 3.5 presents the maximum number of packets transmitted per second measured over a wireless channel in a real IEEE802.11 distributed WLAN. Figure 3.6 presents the maximum data throughput measured over a wireless channel in a real IEEE802.11 distributed WLAN. The results conformed to the numerical analysis in Section 3.3.1 and 3.3.2 that:

1) The results using different experiment configurations in both Figure 3.5 and Figure 3.6 show that increasing the data rate from 11 Mb/s to 54 Mb/s result in a clear gain (about 25%) in both channel capacity and throughput for large packet (1500 byte) transmissions. However, the gain in throughput for small (VoIP) packets is not significant though noticeable in Figure 3.6.
2) Figure 3.5 shows that the maximum number of packets that can be sent per second over a wireless channel are just under 1200. Figure 3.5 also indicates that varying the voice compression rates from G.711 to G.729 in order to reduce the VoIP packet payload is not going to make much difference to the performance (less than 5%).
3) The transmission of small packets significantly reduces the wireless network performance in data throughput. Figure 3.6 presents the calculated data throughput from the experiment results. The wireless channel can only transmit less than 50 Kbytes/s G.729 data and 200 Kbytes/s G.711 data in

maximum. It clearly shows that transmitting larger packets achieves higher data throughput (over 1300 Kbytes/s) than using smaller packets.

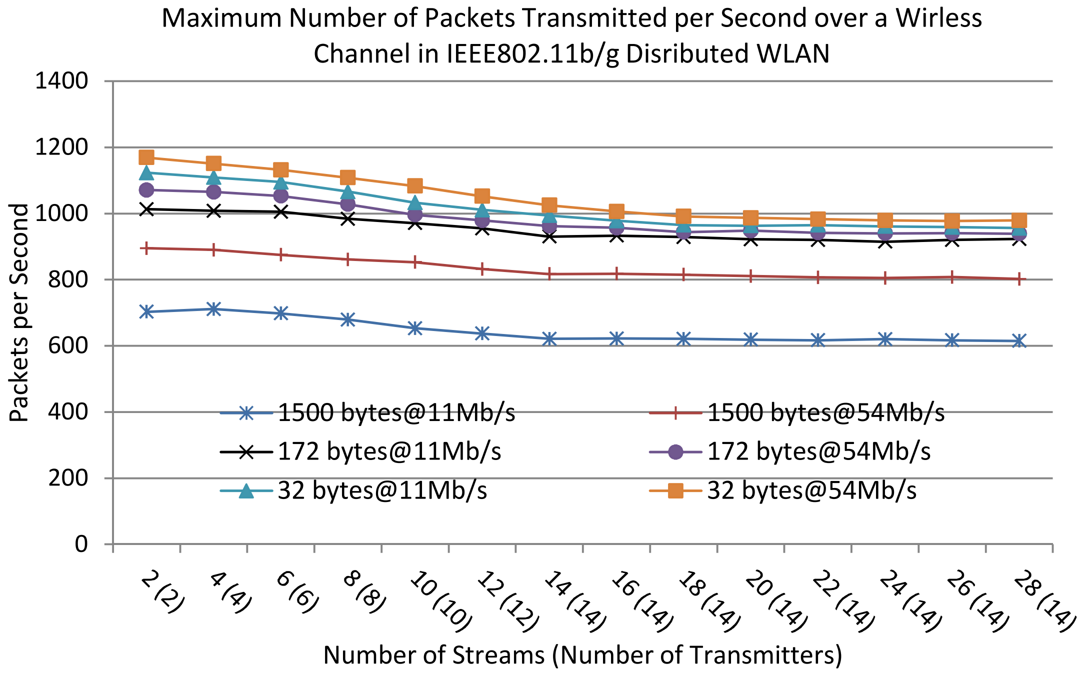

Figure 3.5 Wireless Channel Capacity Measured in Real Device Experiments

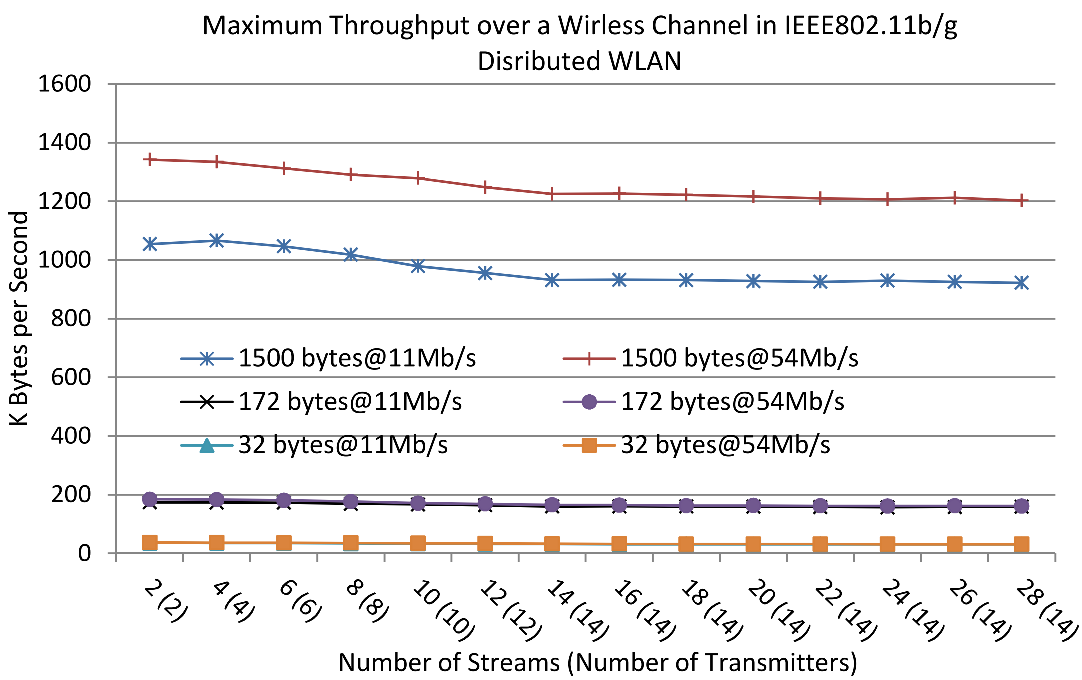

Figure 3.6 Data Throughputs (Kilo Bytes per Second) Measured in Real Device Experiments

It is also observed from the experimental results that increasing the number of simultaneous transmitters will reduce the wireless channel capacity. This reduction in capacity is caused by the increasing probability of contention collisions. Bianchi [59, 60] presented a well-known numerical analysis using a Markov Chain model on the correlations between contention window size, number of contending transmitters and collision probabilities. Both his analytical and simulation results indicated that in a congested WLAN if the number of transmitting stations increases from 5 to 15 the overall throughput will drop by about 10%. This analysis is used by many researchers on WLAN capacity issues including [61-64]. All these analyses concluded that the collision probability increases proportional to the number of contending transmitters in the WLAN. Xu et al. in paper [61] suggested with a minimum contention window size of 31 the overall WLAN throughput will reduce more than 20% when the number of contending transmitters increase from 2 to 14.

A reduction in the number of packets transmitted per second was found in the experimental results which might be due to contention collisions. However, the reduction found is slightly less than stated in paper [61], only about 15%. A possible reason is that optimisations on contention window selection may be implemented by the wireless NIC vendor. However, the contention window size and the collision probability are not included within my research scope. For simplicity, this research is not going to consider this issue.

3.4.2. VoIP Round-Trip Time Measurement using VoIP-RTT-Estimator

In this second set of experiments, due to the difficulty of clock synchronisation, round trip transmission latency for VoIP streams was measured using my VoIP-RTT-Estimator instead of using one way delay. VoIP-RTT-Estimator was implemented in this research so that pseudo UDP VoIP streams could be used to measure the round trip latency between two wireless nodes without using real voice traffic on applications.

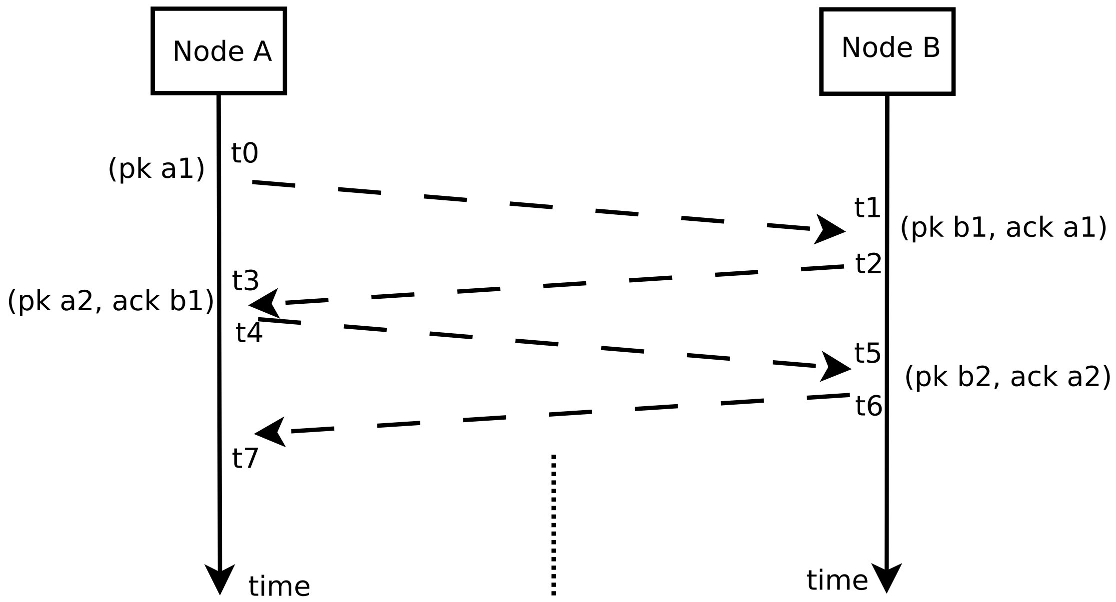

Figure 3.7 VoIP-RTT-Estimator Packet Exchange Scheme

Figure 3.7 presents the packet exchange scheme used by the VoIP-RTT-Estimator to measure the round-trip time. In this example, at time t0 Node A starts sending packets to Node B with constant packet interval. Node B starts sending packets at t0 at which there is an un-acknowledged packet from Node A in its un-acknowledged log. Node B embeds the acknowledgements for packet ID a1 within its packet sent to

Node A. At t3, Node A has received the packet b1 with the acknowledgement of a1. Then Node A will calculate the round-trip time for a1 as t3 - t0. It is important to note that it is impossible for Node A to know the accurate time stamp at Node B as it is lacking a time synchronisation mechanism here. If there is a gap between t1 and t2, it means that the additional delay is caused by e.g. the packet is waiting in the buffer before being sent out. If there are multiple packet IDs in an un-acknowledged log, they will all be acknowledged at once by the next available packet sending to the other end. In this case, different results will be calculated for the round-trip times of these packets at the other end.

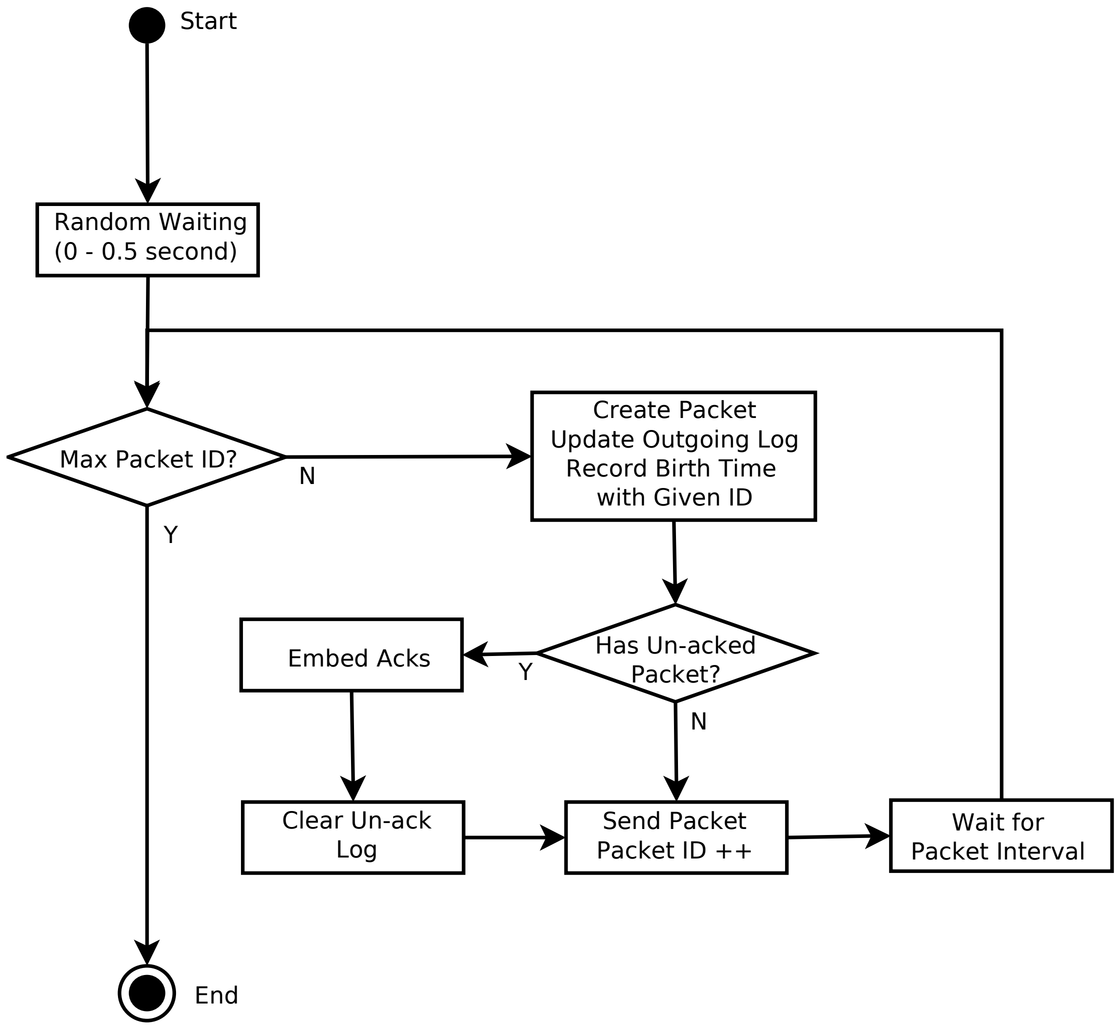

Figure 3.8 Pseudo VoIP Packet Generation Process in VoIP-RTT-Estimator

Figure 3.8 represents the pseudo VoIP packet generation process used in the VoIP-RTT-Estimator. When a packet is generated, the birth time is recorded in an outgoing log. The sender also checks if there is any un-acknowledged packet ID in the unacknowledged log. If so, all the IDs of those unacknowledged packets will be appended within that outgoing packet. Then the unacknowledged log should be cleared and the outgoing packet will be sent down to the UDP socket for transmission.

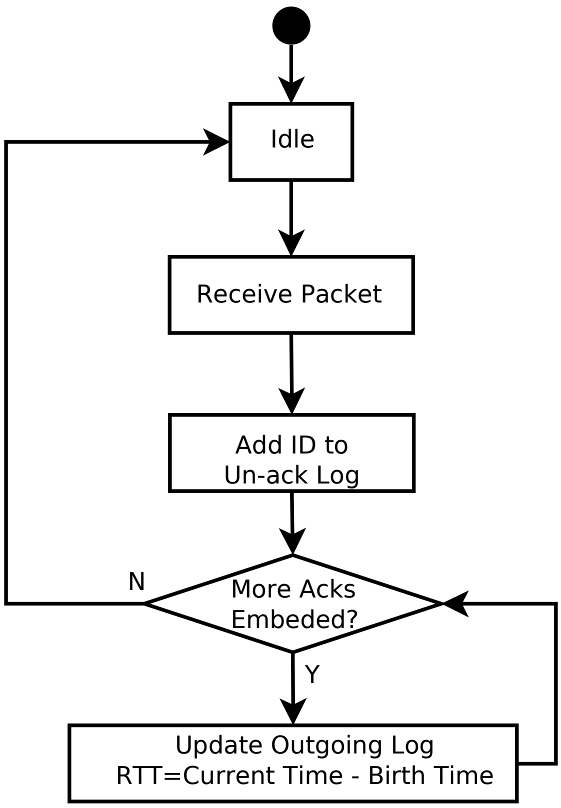

Figure 3.9 Receiving a pseudo VoIP packet by VoIP-RTT-Estimator

Figure 3.9 presents the process for receiving a pseudo VoIP packet by the VoIP-RTT-Estimator. After an incoming packet is received by the UDP socket, its ID is added to the unacknowledged packet list. Then the receiver checks if there is any acknowledgement accomplished with this incoming packet. If so, the round trip time of the acknowledged packet will be derived as current time – birth time.

In order to tell all nodes to start exchanging pseudo VoIP packets at roughly the same time, the experiment processes of VoIP-RTT-Estimator on each wireless node are triggered by a broadcasted signalling packet. However, in order to prevent collision, because calls in the real world are asynchronous the packet generation process at each node does not start until a uniform pseudo random waiting time between 0 and 0.5 seconds has elapsed.

According to Table 3.1, in this experiment the pseudo VoIP is configured with 172 byte (as a 160 byte G.711 packet + 12 byte RTP header) and 32 byte (as a 20 byte G.729 packet + 12 byte RTP header) random generated payload. The packet interval was set to 20 ms. Both 11 Mb/s and 54 Mb/s data rates are tested in these experiments. The round trip latency of each packet is recorded for the first 25 seconds in each round. Each experiment configuration was performed ten times. The figures taken are averaged. In each repeated run, the number of VoIP calls (2 pseudo streams bidirectional) is gradually increased by 2 until all 14 nodes are assigned streams to transmit (7 calls). Then any more VoIP calls are assigned to pairs of nodes which are not already having a conversation, until each node has two operating VoIP calls (14 calls in total).

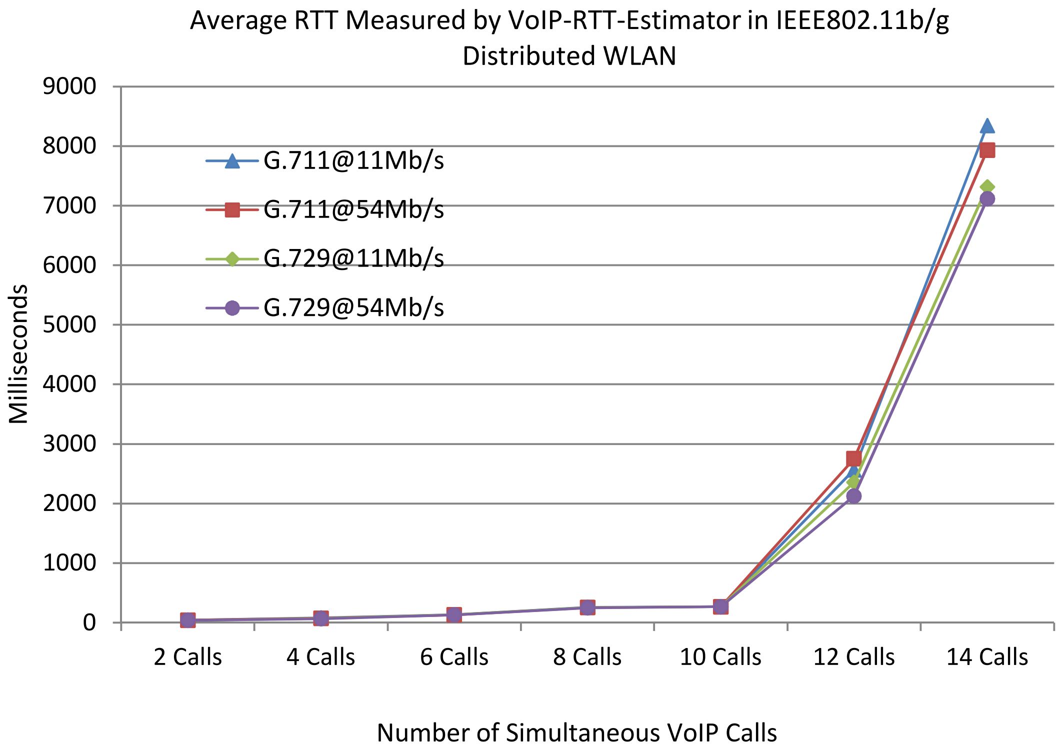

Figure 3.10 Average Round-Trip Time Measured in the Experiments

Figure 3.10 presents the average round trip latency measured in these experiments. This figure indicates that different voice codecs and data rates produce similar behaviours. As soon as the number of VoIP calls exceeds 12, the round trip delay increases dramatically. It is impossible to operate any VoIP calls under such conditions. Because the resulting VoIP service quality (the R value) will be reduced a lot by this high transmission latency and the resulting high I_{e-eff} value. This dramatic increase in latency is due to the delay accumulation phenomenon in the congested wireless channel which is analysed in Section 3.3.3.

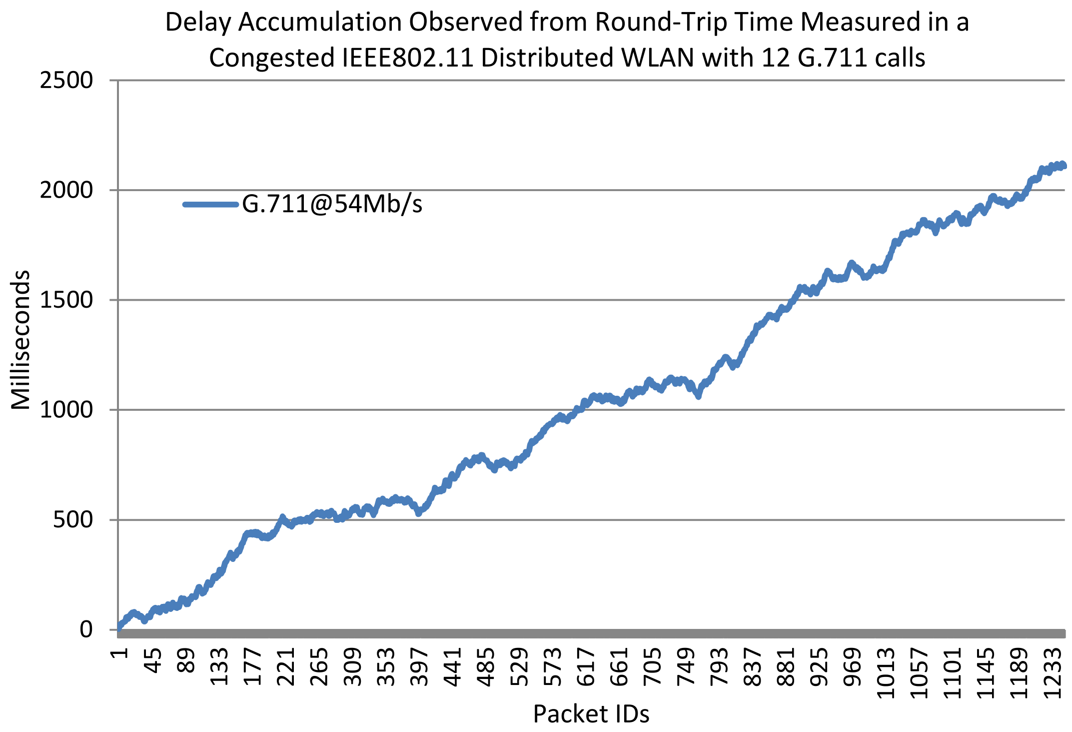

Figure 3.11 Delay Accumulation Effect on Round-Trip Time for each Packet (12 Calls)

Figure 3.11 represents the round trip latency measured for each G.711 packet in a single VoIP stream transmitted using 54 Mb/s in one round with 12 VoIP calls in total operating at the same time. This figure observes the delay accumulation in the outgoing buffers in congested conditions. The round-trip time for each VoIP packet is gradually increasing. After approximately packet ID 150 it is very difficult for users to use the voice service because of the large round trip latency. In a real case, the users may decide to terminate or reset the voice service after a few seconds of this happening. In this experiment, all the streams continue in order to clearly demonstrate the trend of delay accumulation.

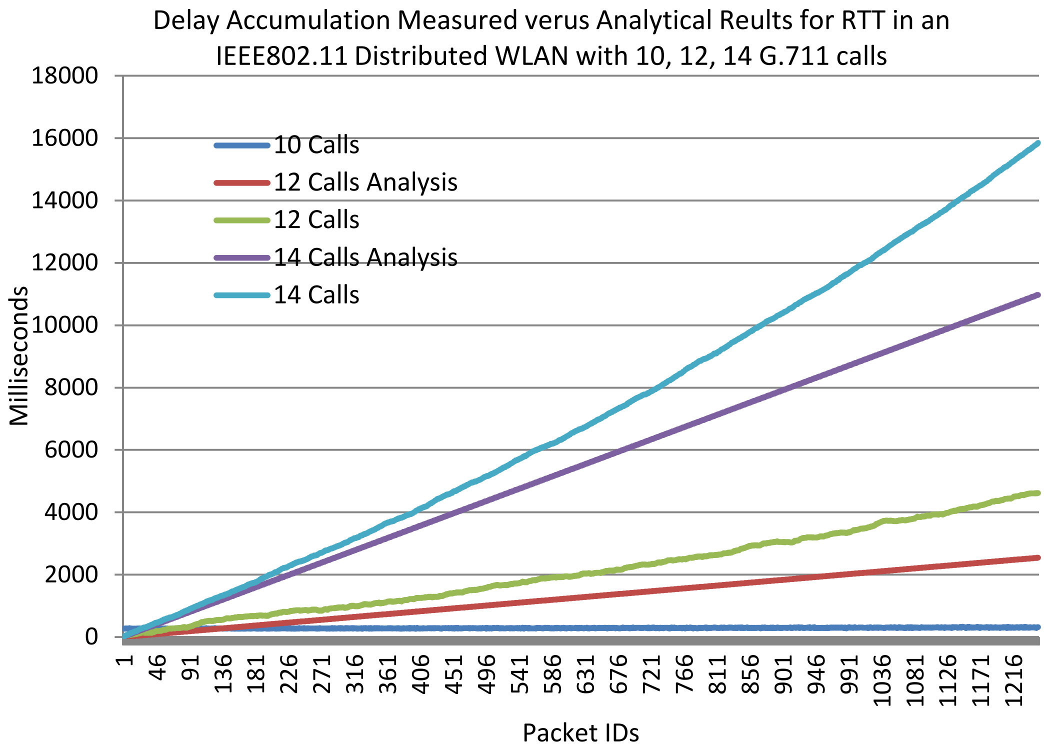

Figure 3.12 Round trip latency measured for G.711@54Mb/s

Figure 3.12 presents the result measured from ten runs of experiments on G.711 VoIP calls transmitted using 54Mb/s data rate. It shows that the round-trip time for 10 VoIP calls or less is not increasing during the experiment, as the wireless channel is not in congested conditions. The calculation results using Equation 3.5 in Section 3.3.3 for 12 and 14 VoIP calls are also plotted in Figure 3.12. The actual results for both are gradually deviating from the calculation results due to contention collisions mentioned at the end of Section 3.4.1 which were not considered in Equation 3.5.

3.5. Simulation Experiments

According to the description of the research scenario in Section 2.3.4, a simulation scenario was built with a 50m×50m space into which 50 wireless nodes are placed randomly with a uniform geometric distribution. This scenario is used to model an environment that is very similar to the IEEE802.11 distributed WLAN previously built in the real device experiments but with more wireless nodes and at varying random locations.

We assume that the shadowing model provided by NS-2 for modelling the signal propagation in an indoor environment gives a packet error rate similar to the real device experiment carried out previously. In this environment, most of the wireless nodes are in range of one another (1 hop). Because previous experiments concluded that there is no significant performance difference for VoIP traffic using different data rates, for simplicity, the simulations focused on 54Mb/s with a frame header and base rate of 1Mb/s. The RTS/CTS function was also disabled. The VoIP streams using the G.711 and G.729 codecs were simulated using the NS-2 constant bit rate (CBR) model. Each VoIP call was set-up on a random pair of wireless nodes. For each VoIP call CBR streams are set-up bi-directionally between these two nodes. Each stream produces 50 packets per second (20 ms interval). The packet size was 172 byte for the G.711 codec or 32 byte for the G.729 codec including the RTP header. Each simulation was run as ten independent iterations. The results are the mean of across these iterations.

3.5.1. Real Device Experiment versus Simulation

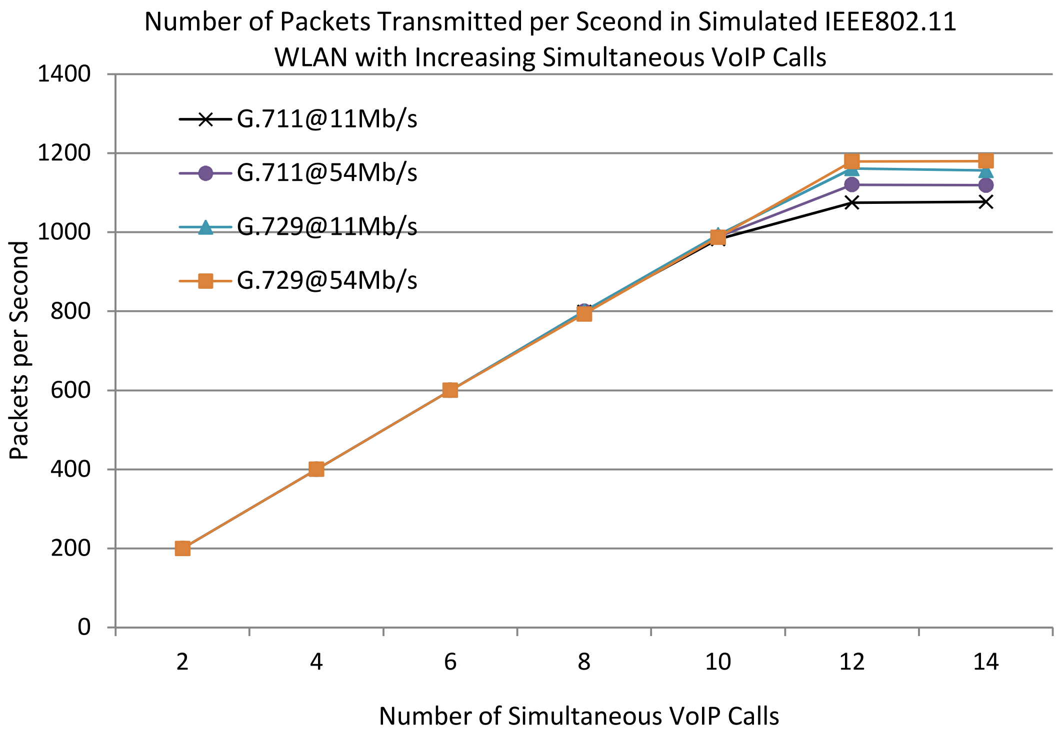

Figure 3.13 Number of Packets Transmitted per Second in Simulations

Figure 3.13 presents the number of packets transmitted per second measured in the simulation. Congestion in the wireless channel happens when the number of simultaneous VoIP calls reaches 12. The wireless channel capacity is estimated to transmit a maximum of about 1150 packets per second in. The network performance in simulations is slightly better than in the real device experiments. This may be due to the difference in packet error rate between the simulation model and the real world environment.

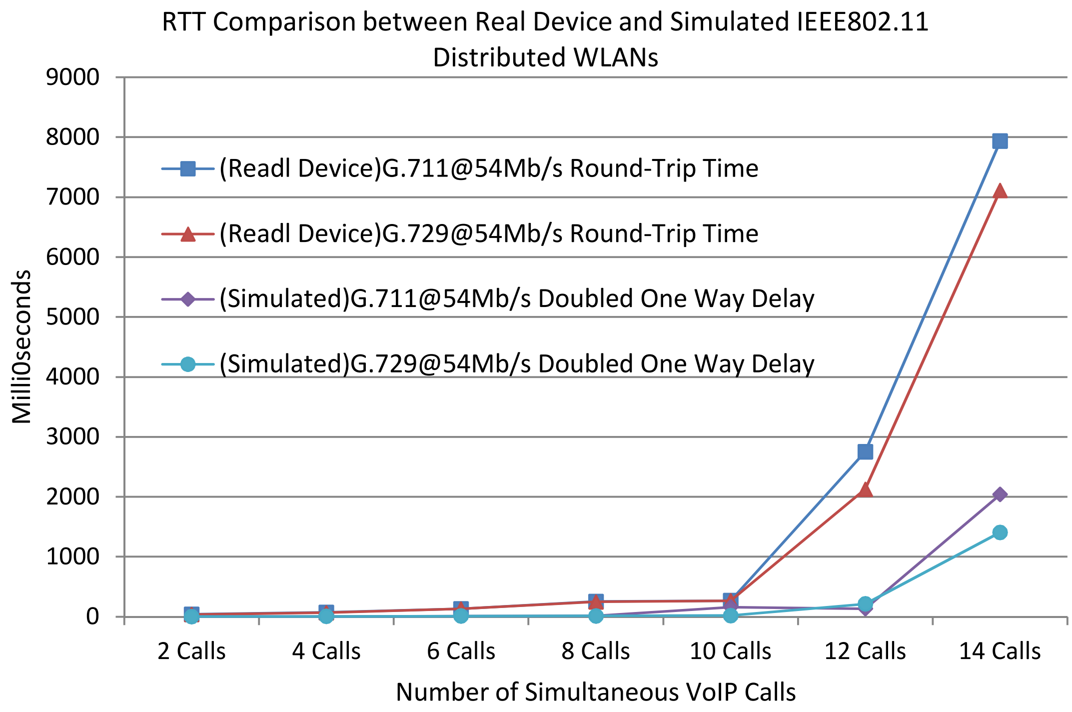

Figure 3.14 Round-Trip Time in Real Device Experiments versus One Way Delay in Simulation

Figure 3.14 compares the doubled one way delay measured in the simulation with the round trip latency measured in the real device experiments for G.711 VoIP streams using 54 Mb/s data rate. Similar behaviours occur after the point where the wireless channel becomes congested in both the simulation and the real device experiments. The delay in transmission increases sharply due to the delay accumulation phenomenon discussed in Section 3.3.3 and 3.4.2.

3.5.2. Network Congestion Simulation

In order to simulate the predicted result of the numerical analysis in a non-congested and then a congested WLAN, a simulation is designed to investigate this dynamic of network performance. In this simulation, 10 VoIP calls will operate in the network initially for 30 seconds to reach a stable network performance. After the 30th second in the simulation, additional VoIP calls 2, 4 or 6 will be added in different runs of simulations in order to congest the network. Throughout, network performance was monitored at 5 seconds intervals.

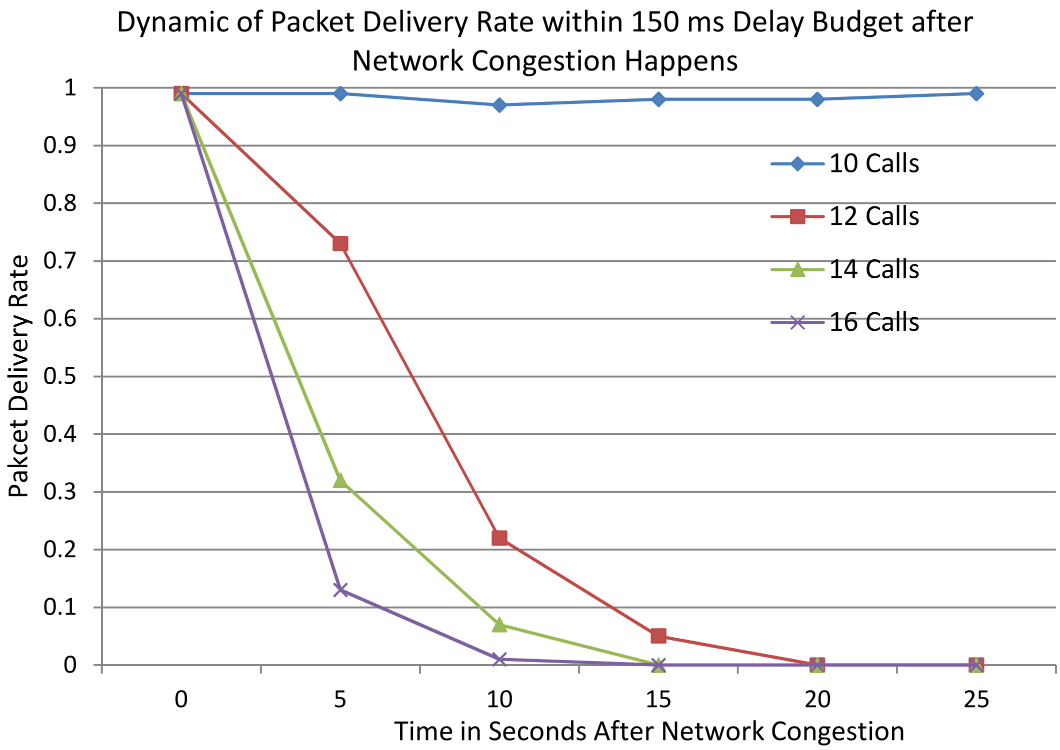

Figure 3.15 Dynamic of Packet Delivery Rate within 150 ms after Network Congestion

Figure 3.15 shows the proportion of voice packets delivered within 150 ms delay budget. This figure indicates that after network congestion happens, it is independent of the number of simultaneous VoIP calls that there will eventually be no packet can be delivered with 150 milliseconds delay budget. The rate of reduction in the network performance is proportional to the number of simultaneous VoIP calls operating and is related to the buffer growth rate and buffer delay mentioned in

Section 3.3.3. If the network is not congested e.g. just 10 VoIP calls are operating, the network performance remains stable with almost all packets delivered on time.

3.6. Summary

This chapter provided a detailed description of the research problems using numerical analysis, real device experiments results and then simulation results. The conclusions drawn from this chapter are below:

1) The numerical analysis presented in Section 3.3.1 and 3.3.2 and the real device experiment results presented in Section 3.4.1 show that the issue of transmission efficiency discussed by the related work [2, 4, 6, 7, 10, 12] based on a single infrastructure WLAN also apply to the scenario of distributed WLANs. The transmission time consumed by the overhead is 818 micro-seconds and by the voice packet payloads can be as little as 3.7 micro-seconds. The maximum VoIP capacity in distributed WLANs is estimated as 12 simultaneous calls which are similar to the estimates from existing research publications based on a single infrastructure WLAN.
2) In the real device experiments presented in Section 3.4.2 and the simulation experiments in Section 3.5, the sharp drop in network performance in congested conditions which was originally found for infrastructure WLANs was also observed for distributed WLANs. When the WLAN is congested, even though the number of simultaneous VoIP calls is just over the capacity e.g. 14 calls the transmission latency will keep increasing. After few seconds the WLAN cannot support VoIP services with acceptable quality as there are no voice packets that are delivered within the delay budget.

3) The numerical analysis in Section 3.3.3, the real device experiment results in Section 3.4.2 and the simulation results in Section 3.5 indicated that the sharp drop in network performance for WLANs in congested conditions is caused by ‘delay accumulation’ which happens as soon as the wireless channel is congested. In the real device experiment, the measured round-trip delay of voice packets can be accumulated over ten seconds just soon after the congestion started. In a real world situation, it is impossible for users to maintain an interactive conversation under such network performance. They might decide to terminate the current call and retry.

So far, the Hypothesis 1 defined in Section 1.1.4 has been proved true. The research Contribution 1 and 2 were made by the conclusions of this chapter. In order to deliver better VoIP service quality over distributed WLANs especially to prevent the collapsing of network performance in congested conditions, the next two chapters of this thesis will propose two different solutions to resolve the ‘delay accumulation’ in such conditions. Chapter 4 will propose an outgoing buffer management strategy which focuses on preventing the ‘delay accumulation’ of voice packets. Chapter 5 will propose a packet aggregation algorithm which focuses on improving the transmission efficiency but also helps to prevent the ‘delay accumulation’ in the outgoing buffer as a side effect.

4. A Queue Management Strategy for VoWLAN

Chapter 3 has defined the problems for VoWLAN that are the focus of this research in detail. It was indicated that when the WLAN is congested by VoIP traffic the outgoing buffer configuration becomes a decisive factor for the service quality. Delay accumulation of packets in the transmitter's LLC sub-layer outgoing buffer is expected in congested conditions. The buffer is required for storing the incoming packets from the network layer whenever the transmitter is not able to send immediately.

Being stored in the outgoing buffer increases the latency of the packet being delivered. The transmission latency may not be critical for non-real-time network applications such as file transfers, emails or web browsing. However, for a real time interactive application such as VoIP, if the packet arrives at the receiver beyond the delay budget, it is equivalent to the packet being lost. The voice codec at the receiver's end will use PLC techniques to generate synthetic voice samples to replace those missing voice samples because of packet loss. Hence, there is no point in transmitting those voice packets that are already too late to deliver and whose samples are already regenerated by LPC techniques at the receiver's end.

In congested conditions the transmission opportunities in the wireless channel are considered as important resources for the network traffic. The delay accumulation in congested conditions happens when the transmitter gives the transmission opportunities to those 'stale' voice packets that wait at or towards the front of the outgoing buffer. When the previous 'fresh' voice packets move from the back of the buffer to the front they probably become 'stale' as well due to the buffer delays. Therefore, a queue management strategy is required to prevent delay accumulation in the transmitter's LLC sub-layer outgoing buffer when the network is congested.

This chapter presents a simple queue management strategy for the LLC sub-layer outgoing buffer that checks the 'freshness' of the voice packets and drops those that are considered as 'stale'. Section 4.1 explains the idea of this queue management strategy in numerical terms. Section 4.2 reviews the existing studies on the relationship between buffer size and transmission latency. Section 4.3 introduces the current strategies that actively drop the packets from the network buffers. Section 4.4 presents this simple queue management scheme, Active Clean Queue (ACQ). Section 4.5 evaluates ACQ.

4.1. Numerical Analysis

The analysis in Section 3.3.2 indicated that the IEEE802.11 WLANs can accommodate up to about 12 VoIP calls. If the number of simultaneous VoIP calls exceeds the upper bound of capacity, the network will be congested. The analysis in Section 3.1.3 indicated the network congestion is going to cause the number of packets in the outgoing buffer of the transmitters to increase. The increase rate ($G_{/s}$) can be estimated using Equation 3.4. Then the delay (D_{buf}) caused by buffering can be estimated using Equation 3.5. This problem is called delay accumulation in this thesis. In order to mitigate this issue, the buffer growth rate ($G_{/s}$) has to be zero or negative.

Most of the existing research on Active Queue Management (AQM) to prevent network congestion was based on the assumptions that the congestion will be resolved very soon, especially when TCP is used by most of the applications. As soon as TCP notices the network becomes congested due to time outs of expected TCP ACKs, most TCP implementations reduce their rates of sending packets. The aim of such behaviour is to try to reduce the occupancy of buffers by reducing the rate of packets arriving to less than the outgoing rate. In other words, TCP increases the average interval of sending packets to the network layer. Using Equation 3.4, if we consider $\frac{1}{T_P}$ is the average rate that TCP sends packets down to the network layer, in order to reduce this rate the average interval of TCP sending packets should increase by R_{TCP} based on the original packet generation interval (T_P). This new expression is:

$$\boldsymbol{G_{/S} = \frac{1}{T_P + R_{TCP}} - \frac{P_{MAX/s}}{N_{call} \times 2}} \qquad \textbf{\textit{Equation 4.1}}$$

Where: $\frac{1}{T_P+R_{TCP}}$ *calculates the number of packets arriving at the outgoing buffer per second;* N_{call} *is the number of simultaneous VoIP calls; there are two VoIP sources (*$N_{call} \times 2$*) for each VoIP call;* $\frac{P_{MAX/s}}{N_{call}\times 2}$ *calculates the average amount of wireless channel capacity shared by each VoIP source.*

In this process, TCP at the transport layer takes responsibility for congestion control. For conventional network applications such as file transfers, emails or web browsing, the traffic demand is generated in bursts. Because such applications are not very sensitive to transmission latency, TCP is helping to distribute the traffic demand over a longer period when the network is congested. The use of buffers allows some of the packets that have already been sent by the TCP to still be delivered sometime after the congestion is resolved. Any packet arriving at an already full buffer will usually be dropped. The network congestion should be resolved by the cooperation between TCP reducing incoming traffic flow and the buffer temporally absorbing some of the excess traffic.

However, UDP used by most of the VoIP applications is not aware of network congestion. On the other hand, real-time applications like VoIP generate a constant rate of packets over the period of usage. Therefore VoIP causes network congestion to persist as long as the number of simultaneous VoIP calls does not reduce. Considering the expression in Equation 3.4, a factor R has to be introduced to reduce the buffer growth rate $\boldsymbol{G_{/S}}$ as follows:

$$\boldsymbol{G_{/S} = (\frac{1}{T_P} - \frac{\frac{P_{MAX}}{s}}{N_{call} \times 2}) - R \leq 0}$$ ***Equation 4.2***

*Where: **R** represents the number of packets dropped from the outgoing buffer.*

When R satisfies the conditions in Equation 4.2, the accumulation of delay will stop. Dropping packets from the outgoing buffer is necessary when either the buffer is full or an Active Queue Managements (AQM) algorithm tells the buffer to actively drop some selected packets from the outgoing buffer.

4.2. Adaptive Buffer Size

The relationship between buffer size and queuing delay especially for wired network routing has been actively studied [66-68]. Papers [66, 67] studied the transmission latency as an important network performance parameter and suggested that a smaller buffer size can reduce the latency. However, paper [68] focused more on the packet delivery rate and argued that smaller buffer size may reduce the network throughput especially for TCP traffic. In the case of the VoIP applications, it requires the network to provide both low latency and high delivery rate transmission. There is also some existing research dedicated to the relationship between buffer size, transmission latency and the VoIP capacity for WLANs such as, [10, 11, 18, 19, 51].

Malone et al. also claimed that their paper [18] was the pioneer of studies on the buffer sizing for VoWLANs. The simulation result in paper [18] concluded that before the WLAN is congested (the number of simultaneous VoIP calls is under 12) the transmission latency is not effected by the buffer size. However, if the WLAN is congested (the number of simultaneous VoIP calls exceeds 12) having a bigger buffer results in larger latency. Shin and Schulzrinne had similar findings in paper [10] that focused more on the AP. They used experiments on WLAN test bed and concluded that the network buffer size configured at the AP is proportional to the queuing delay and inversely proportional to the packet loss rate.

Dangerfeild et al. provided some experimental measurement focusing on the AP's buffer size based on their infrastructure IEEE802.11b WLAN test bed [19]. Their experimental results show that before the WLAN becomes congested by VoIP calls (e.g. 5 or 10 simultaneous calls), the AP's buffer size does not significantly affect the transmission latency when the WLAN is congested (e.g. 15 simultaneous calls) an AP with larger buffer size has the larger transmission latency. Paper [19] also indicated that applying the IEEE802.11e TXOP scheme on the AP can improve the infrastructure VoWLAN capacity up to 16 simultaneous VoIP calls. However it requires the buffer size at the AP should be configured approximately proportional to the number of simultaneous VoIP calls in order to achieve an optimal capacity.

Steockigt and Vu provided similar but more detailed measurement based on numerical analysis, simulation and a test bed in their papers [11, 51]. Their results also show that the AP's buffer size results in different transmission latencies in congested conditions. They also found that when the WLAN is not congested using smaller buffer size (e.g. 10 packets in their scenarios) at the AP results slightly higher packet loss rate than using larger buffers. However, in congested conditions when the AP's buffer is larger than a certain size (30 packets in their scenarios) there will be no significant difference in packet loss rate by increasing the buffer size further. They concluded that in their scenario, the optimum buffer size at the AP is 30 packets. They also investigated the impact of the IEEE802.11e EDCA TXOP scheme in infrastructure WLANs. The maximum VoIP capacity can be extended to about 20 simultaneous calls when the IEEE802.11e TXOP scheme is applied on the AP.

If an adaptive buffer size scheme is applied in order to prevent the delay accumulation when in congested conditions, according to Equation 3.5 the buffer size (S_{buf}) must be reduced to a very small value. In the case of supporting VoIP application assuming the buffer is full in congested conditions, the buffer size must be as small as when a voice packet moves from the end to the front of the buffer so it is still in time to be used after being transmitted to the receiver's end. Hence, the buffer delay (D_{buf}) may be reduced to be within a VoIP application's delay budget.

However, the trade-off for shrinking the outgoing buffer is increased in packet loss when in congested conditions.

4.3. Active Queue Management

Active Queue Management (AQM) is a set of algorithms working on the outgoing buffers that are between the transport layer and network layer to prevent network congestion. The general idea is that when the outgoing buffer occupation starts growing but the buffer has not yet reached the pre-configured maximum occupation, dropping a few packets in the buffer may be noticed by the transport layer protocols, which will assume congestion and reduce the traffic load it is sending to the network layer. The AQM algorithms are mainly designed for TCP because firstly it has the acknowledgment mechanism to detect whether the packets are delivered to the destination, secondly it can react (to reduce the traffic load) to packets being dropped, and thirdly it guarantees the dropped packets will be recovered. This section briefly reviews the development of the AQM.

4.3.1. The Original Random Early Detection

The Random Early Detection (RED) scheme is sometimes also called 'Random Early Drop'. It is the most well-known series of AQM algorithms. The original RED scheme [69] was proposed as early as 1993 for wired packet switched network gateways. It monitors the occupancy of the buffer. There are two thresholds of the buffer size. One is called the minimum threshold (Min_{th}) and the other is called the maximum threshold (Max_{th}). If the buffer occupancy is less than Min_{th} , then no packet will be dropped. If the buffer occupancy is between $Min_{th}$and Max_{th}, the incoming packet will be dropped with a probability (p_d) that is between 0 and a maximum probability

(Max_p) and proportional to how full the buffer is. If the buffer occupancy is greater than Max_{th}, then all the incoming packets are dropped. The simulation results in paper [69] indicate that applying RED at the gateway router can effectively prevent TCP at satellite nodes from over increasing their sliding window, which will result in bursty traffic load and congest the router.

4.3.2. Variants of RED

The Adaptive RED (ARED) [70], the Proportional-Differential RED (PD-RED) [71] and the Dynamic RED (DRED) [72] tried to improve the performance of the original RED by modifying the packet dropping probability. ARED makes the maximum dropping probability (Max_p) vary according to the traffic load. It periodically monitors the buffer occupancy. If the current buffer occupancy is greater than a predefined target (T and $Min_{th} < T < Max_{th}$) the Max_p increases otherwise Max_p decreases [70]. PD-RED does not involve Min_{th} and Max_{th} any more. Similar to ARED, PD-RED also periodically monitors the buffer occupancy and increase or decrease Max_p proportionally to the difference between current buffer occupancy and a predefined target (T) [71]. DRED does not vary Max_p. Instead, it directly calculates the current p_d ($p_d(i)$) based on its previous value ($p_d(i-1)$) and the difference between current buffer occupancy and a predefined target (T) that is similar to PD-RED[72].

RED with in/out bit (RIO) [73] and Weighted RED (WRED) [74] can differentiate network traffic into categories based on for example traffic directions (in bound/out bound), traffic pattern (bursty/CBR) or packet sizes. RIO and WRED assign different Max_p, Min_{th} and Max_{th} to each category. As a result, each traffic category will have a different drop probability p_d. The difference between RIO and WRED is RIO considers packets belonging to different categories as separated queues; and WRED considers the packets from all categories as a single queue [75].

4.3.3 Other AQM Approaches

There are other AQM algorithms that notify the senders about network congestion before the router's buffer becomes fully filled. Adaptive Virtual Queue (AVQ) uses a vertical queue in parallel to the real network buffer [76]. The size of the virtual queue ($\tilde{C}$) is smaller or equal to the size of the real buffer (C). AVQ periodically monitors the occupancy of the real buffer. If greater than $\tilde{C}$ the further incoming packets will be dropped or marked using Explicit Congestion Notification (ECN) bit [77]. The use of ECN bit is an added feature to TCP and IP for notifying both ends of a transmission link about the network congestion without dropping packets. The virtual queue size ($\tilde{C}$) can vary according to the capacity of the link and the desired link utilisation.

GREEN algorithm was proposed in [78]. It estimates the probability (p_d) of randomly dropping packets from the buffer by periodically monitoring the incoming traffic load. The incoming traffic load is estimated by the difference of the buffer occupancy between the current measurement and the previous measurement. If the incoming traffic load is greater than the link bandwidth p_d increases otherwise p_d reduces.

4.4. Active Cleaning Queue for VoWLAN

4.4.1. Motivation

Most of the existing research on buffer sizing for VoWLANs such as, [10, 11, 18, 19, 51] mentioned in Section 4.2 were based on infrastructure topology in which the traffic demand and the distribution of transmission opportunities is unbalanced between the AP and other STAs. The AP is a bottleneck in a condensed VoIP loaded WLAN. The IEEE802.11e TXOP scheme can be applied to prioritise the transmission opportunities for the AP. When this scheme cooperates with the appropriate buffer size at the AP, the optimal theoretical VoIP capacity can be achieved.

A prioritising approach like the IEEE802.11e standard is suitable for the scenarios such as VoIP traffic mixed with other network traffics or the traffic loads are unevenly distributed amongst wireless nodes. However, in the scenarios of this research, the WLAN topology is distributed. The VoIP traffic demand and the transmission opportunities are random uniformly distributed amongst the active nodes in the WLAN. According to the analysis in Section 3.3.2 and 3.3.3, if the number of simultaneous VoIP calls is greater than the maximum capacity of the WLAN, the wireless network will be congested. VoIP applications generate a constant rate of network traffic that needs to be delivered with very small delay. In the scenario of this research it is very difficult to resolve the congestion by prioritising the transmission (e.g. applying TXOP scheme). Prioritising the transmission opportunities for some VoIP streams will make the service quality of other VoIP streams even worse.

For the scenario above, reducing the size of the network buffer for the nodes in the WLAN will result in little gain because the packets dropped will normally be the 'fresher' packets, which have more possibility to be delivered within time than the older queued packets waiting in a larger outgoing buffer.

As stated in Section 2.2.3, VoIP applications can tolerate some packet loss using PLC techniques. If the number of simultaneous VoIP calls is not significantly over the wireless medium capacity, a queue management strategy may be used to drop a few packets from the buffer in order to reduce the buffer growth rate, and help to resolve the congestion.

The existing AQM algorithms are not very suitable for VoIP applications. First of all, they prevent congestion relying on the response of the transport layer protocols, for example, TCP reducing the size of the sliding window. However, in the case of congestion, TCP will reduce the data rate by reducing the sliding window. It may result in the transmission being delayed. The voice data delivered after the congestion has been resolved may be too late to play at the receiver end. Most of the VoIP applications use the RTP/UDP/IP stack. The rate of sending packets is determined by the application. UDP does not provide any functions to adjust the packet sending rate though RTCP can inform VoIP applications using it about the change of network performance after some delay. Because the VoIP application is very sensitive to the transmission delay, the transmission needs to continue while trying to resolve the congestion.

Secondly, the integrity of the data is not considered by the existing AQM algorithms as the choice of the packet to drop from the buffer is random. The transport layer protocols are responsible for the data integrity after packets have been dropped. VoIP applications have different data integrity requirements than normal applications. A small amount of data lost can be recovered by PLC techniques. In order to maintain the delay budget within a small range, the packets that are already very late (e.g. delay > 150 ms) may be given up and not transmitted forward. As a result, a more precise selection of packets to be dropped can help to resolve network congestion without a big reduction in the VoIP application's service quality.

According to Equation 4.2 presented in the numerical analysis in Section 4.1, a factor R as the number of packets being dropped from the outgoing buffer is required in order to resolve the delay accumulation. A new algorithm for queue management strategy is proposed in this chapter to perform packet dropping selection based on transmission latency. This strategy can actively select stale voice packets and drop them from the outgoing network buffer, which is called Active Cleaning Queue (ACQ).

4.4.2. The ACQ Algorithm

```
clean()
{
  for each packet P
    if current_time - P.arrival_time > T_MAX then
      drop (P)
    end if
  end for
}
```

Algorithm 1 Queue Management Strategy ACQ

As shown in Algorithm 1, the packet dropping selection algorithm is very simple. Each packet is assigned a parameter (*arrival_time*) to record the time coming in to the buffer. The parameter *arrival_time* is not implemented within the packet header. Therefore, no modification to the current packet structure is required. A function (*clean()*) is implemented to 'clean out' all the 'stale' packets in the buffer whose delay time is greater than T_{MAX} (see Algorithm 1). The performance of this strategy can be affected by two elements. One is when to call the *clean()* function. In this thesis, every time a packet arrives into or leaves from the buffer the *clean()* function is called. This ensures that all the packets arriving in or leaving are properly assessed by this management function. The *clean()* function only drops voice packets if it needs to. They can be identified by the port numbers in the packet header. Another element is the choice of the maximum packet delay time (T_{MAX}) in the outgoing buffer. It will be discussed in more detail in Section 4.5.

4.5. Evaluation Results

In order to test and evaluate the performance of the proposed queue management strategy ACQ, the *clean()* function was implemented for the LLC sub-layer buffer on top of the IEEE 802.11DCF MAC protocols and simulated using NS-2. In order to compare the network performance for the VoIP traffics with and without ACQ, the simulation scenario and configuration were identical to that defined in Section 3.5.

As indicated in Section 2.2.3, the VoIP traffic delay budget applied throughout this thesis is 150 milliseconds. As a result, if the 150 ms one way transmission delay is considered as the maximum delay that the VoIP users can tolerate, the T_{MAX} was initially set to 150 ms in this simulation. Other T_{MAX} values (100 ms and 200 ms) were also tested in the simulation for comparison. Each simulation was run as ten independent iterations and the results presented are averaged across these iterations.

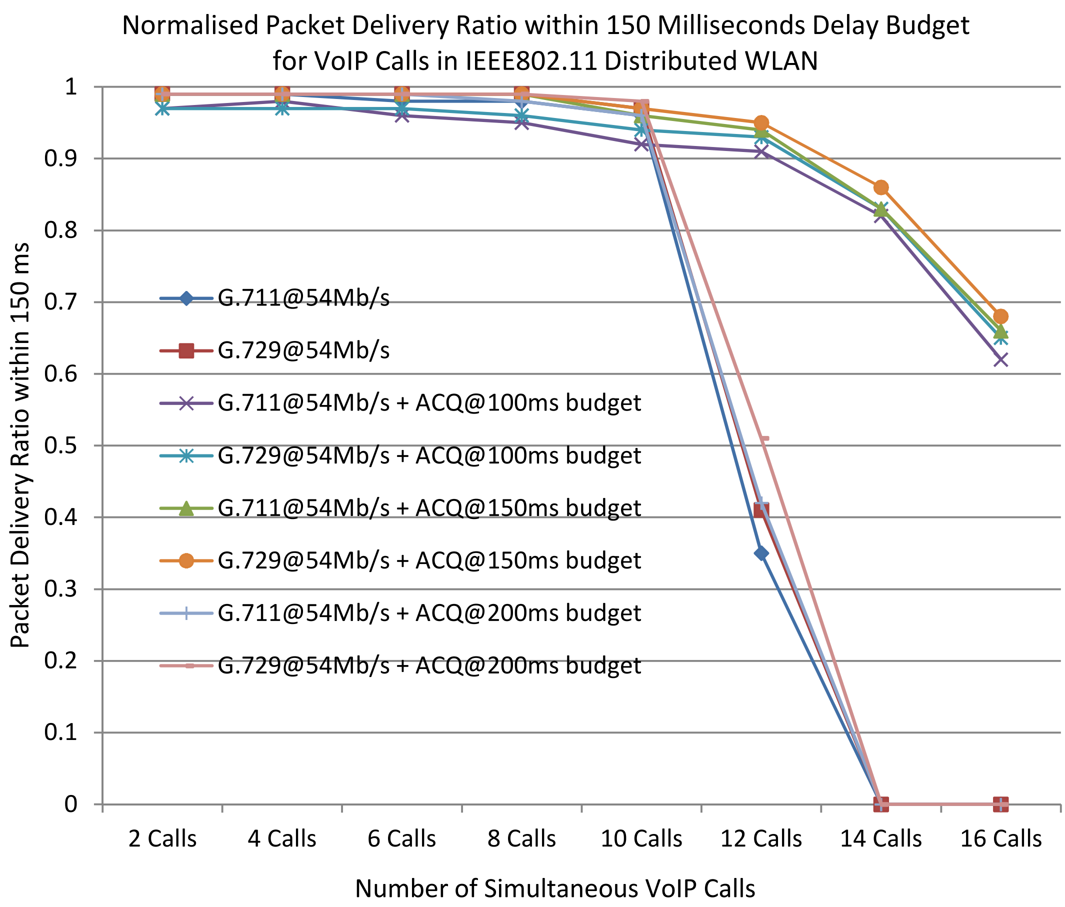

Figure 4.1 Delivery Rate within 150 ms

Figure 4.1 shows the ratio of voice packets delivered within 150 ms without ACQ management and then with ACQ management with T_{MAX} (budget) set to 100, 150 and 200 ms. The simulation results clearly shows that 12 VoIP calls is the normal maximum capacity of the wireless medium in this scenario. As soon as the number of simultaneous VoIP conversations is over 12 the wireless network will be congested by the voice packets. Without any queue management, the IEEE802.11 WLAN finds it nearly impossible to deliver any voice packet in time in this congested condition.

However, by deploying the ACQ management strategy, with 12 simultaneous VoIP calls, most voice packets are still being delivered within the delay budget. The packet delivery rate within the time limit is gradually decreasing when the number of simultaneous VoIP conversations is increasing, but the network performance drop is much less steep.

Figure 4.1 also shows that if the value of $\mathrm{T_{MAX}}$ is smaller than the maximum transmission delay (delay budget) that can be tolerated (e.g. $\mathrm{T_{MAX}}$ = 100 ms), the on-time delivery ratio of voice packets is less than the configuration of $\mathrm{T_{MAX}}$ matching the delay budget. This is because with $\mathrm{T_{MAX}}$ < 150 ms there are some packets being dropped from the outgoing buffer while they still have the possibility to be delivered on time. The results also show that if the value of $\mathrm{T_{MAX}}$ is greater than the delay budget (e.g. $\mathrm{T_{MAX}}$ = 200 ms), there will hardly be any voice packets being delivered within the delay budget in congested conditions.

Figure 4.2 presents the average outgoing buffer occupation measured from the simulation. Without the application of ACQ algorithm, there are many voice packets stuck in the outgoing buffer under congested conditions. The buffer occupancy increases dramatically after 12 simultaneous VoIP calls. By applying the ACQ algorithm, the occupancy of the outgoing buffer (S_{buf}) rises more slowly and stays stable with a few more active calls.

Figure 4.2 also shows the correlation for the choice of the $\mathrm{T_{MAX}}$ value and the occupation of the outgoing buffer (S_{buf}). In congested conditions (number of simultaneous VoIP calls more than 12), the lower the $\mathrm{T_{MAX}}$ value is, the lower the occupancy of the outgoing buffer (S_{buf}) will occur. Therefore, less buffer delay (D_{buf}) will be accumulated to the transmission latency.

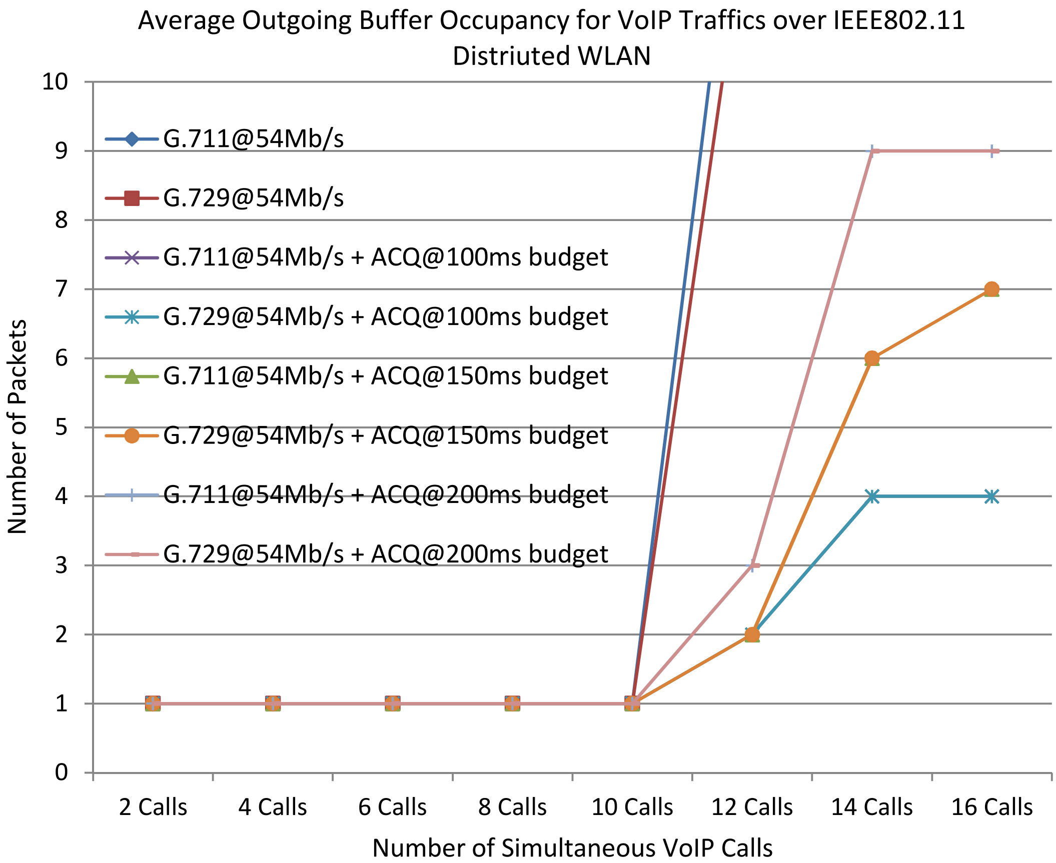

Figure 4.2 Average Outgoing Buffer Occupation

Figure 4.3 shows the average transmission latency of the voice packets. The results show that without any queue management, the transmission latency suddenly increases just after the wireless medium becomes congested. However, by deploying the proposed queue management strategy, the service quality (transmission latency) now gradually decreases due to dropped packets as the number of simultaneous VoIP conversations is increased.

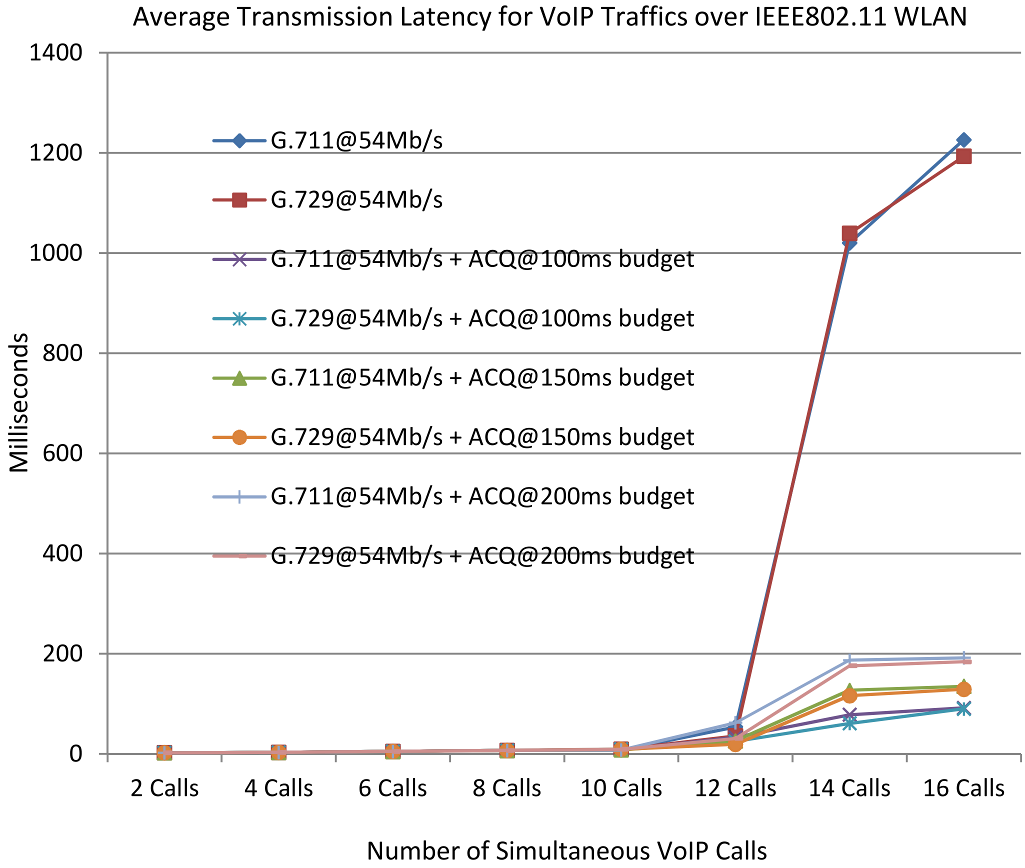

Figure 4.3 Average Transmission Latency

Figure 4.3 also shows that the average transmission latency after applying the ACQ algorithm is correlated to the configured $\mathrm{T_{MAX}}$ values. When $\mathrm{T_{MAX}}$ is set to 100 ms it has the least average transmission latency about 100 ms. When $\mathrm{T_{MAX}}$ is set to 200 it has the highest average transmission latency for using ACQ in the simulation just under 200 ms. According to the previous observations from Figure 4.2, in the network congested conditions, the lower $\mathrm{T_{MAX}}$ value configured, the less buffer delay (D_{buf}) will be accumulated into the resulted transmission latency.

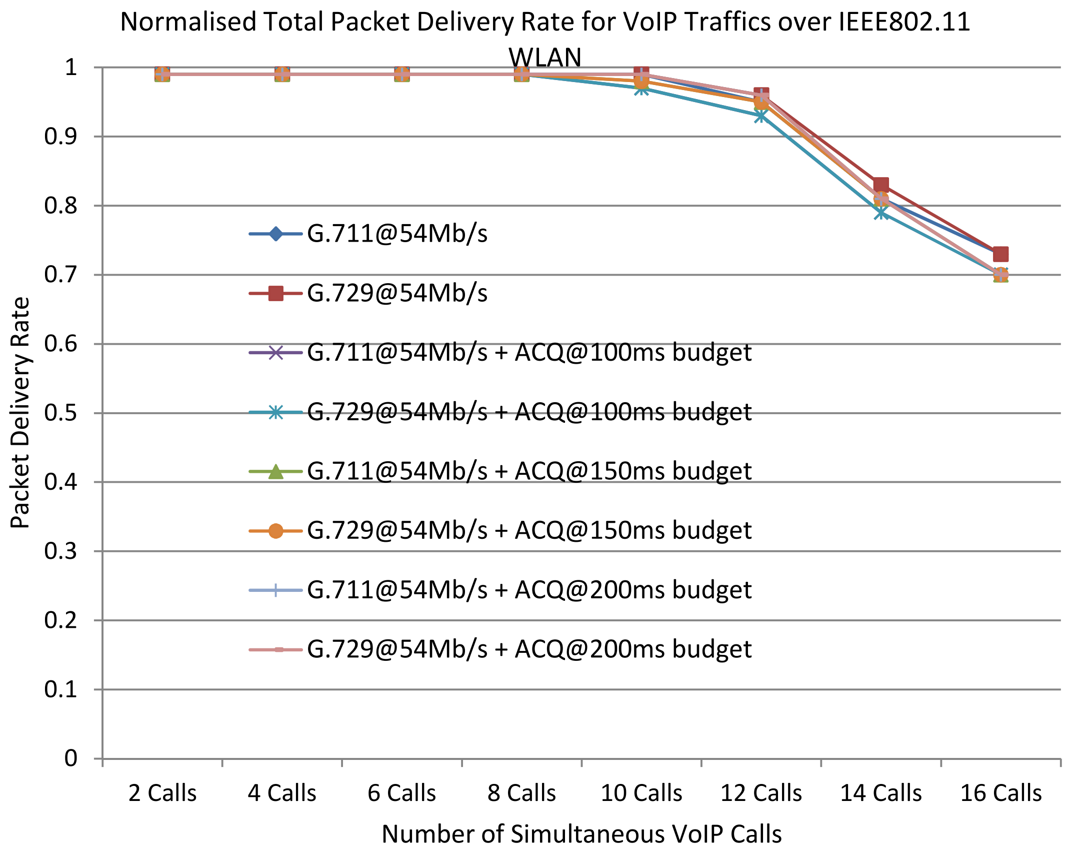

Figure 4.4 Packet Delivery Rate

Figure 4.4 presents the total packet delivery rate (without the consideration of 150 ms delay budget limit) measured from the simulations. It shows that there is no significant difference in packet delivery rate with and without applying the ACQ algorithm in congested conditions. In congested conditions, the rate of packets being dropped due to a full buffer is proportional to the amount by which traffic demand exceeds the wireless channel capacity. The effect of the ACQ algorithms on buffer occupation is similar to what happens by reducing the buffer size. As a result, using either ACQ or a smaller buffer is expected to produce similar performance on packet delivery rate if no transmission latency and delay budget are considered. In other words, dropping packets using ACQ will not cause a noticeable reduction in total packet delivery rate in congested conditions as the WLAN's maximum packets carrying capacity has not changed.

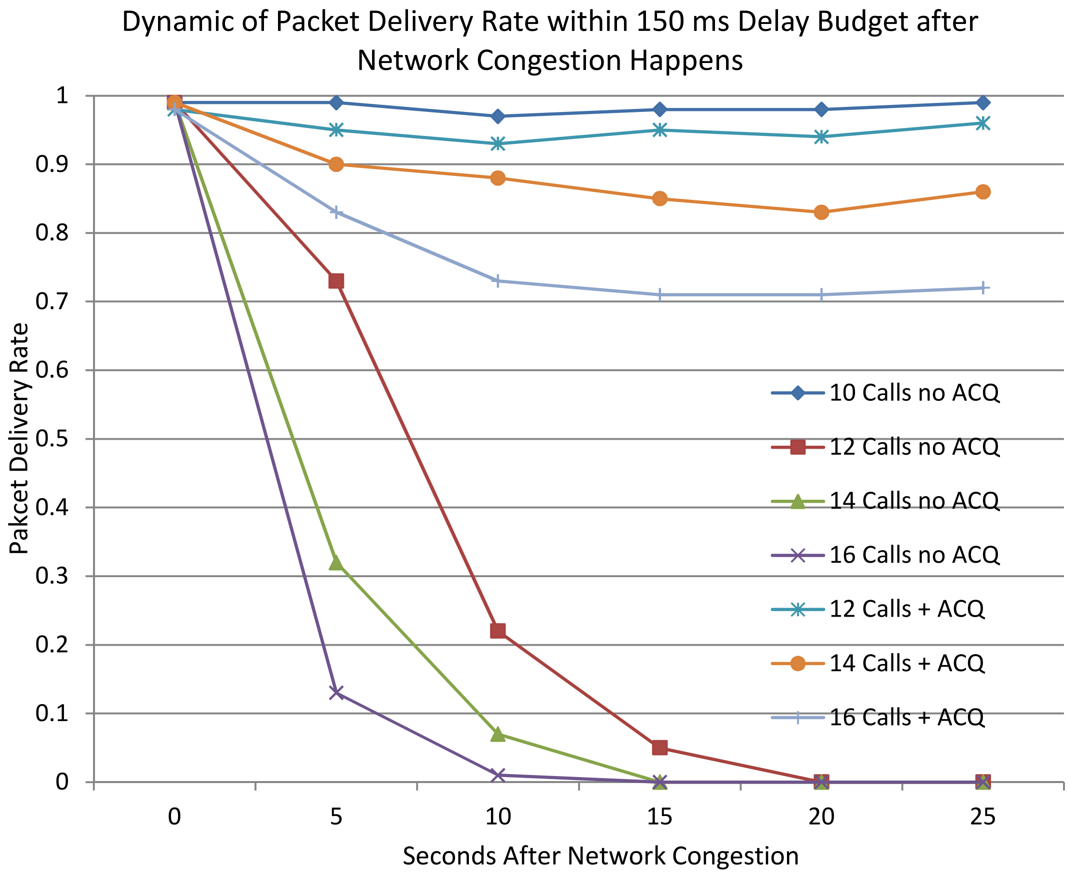

Figure 4.5 Dynamic of Packet Delivery Rate within 150 ms after Network Congestion

The tests for congestion described previously in Section 3.5.2 are also performed for ACQ to compare the performance measured from Section 3.3.2. Figure 4.5 demonstrates that after the network becomes congested the ACQ algorithm maintains the VoIP performance stable, although the best possible performance under congested conditions is reduced proportionally to the number of simultaneous VoIP calls.

4.6. Summary

This chapter proposed a novel simple queue management strategy ACQ for UDP VoIP type traffic for WLANs to solve the delay accumulation in congested conditions identified in Chapter 3. It is one of the research contributions (Contribution 3) presented in this thesis. The simulation results show that without applying ACQ, when the number of simultaneous VoIP calls exceeds 12 the users will suffer in the service quality collapsing caused by delay accumulation. After deploying the ACQ management strategy, the delay accumulation in the outgoing buffer is resolved, the service quality reduction becomes much gentler after the wireless medium becomes congested. Although the WLANs are still congested conditions, the VoIP service quality decreases gently proportional to the number of additional VoIP calls that exceeds the wireless channel capacity.

This strategy is compatible with the original IEEE802.11 or the newer IEEE802.11e EDCA. For the implementation in the original IEEE802.11 WLAN the ACQ can apply to voice packets only by identifying the VoIP packets from examining the UDP and RTP headers. For the implementation in IEEE802.11e EDCA, the ACQ can directly apply to the dedicated voice packet buffer (VC_VO).

There are several possible further improvements to the ACQ management strategy:

1) More research can be done to find the most appropriate strategies for deciding when to drop packets from the outgoing buffer by the ACQ. For example, the queue size can be monitored using similar algorithms to RED and decide whether to call the *clean()* function.

2) RTCP may be used for monitoring the run time service quality. The T_{MAX} parameter might be adapted at run time according to the dynamic of service quality. The ACQ may produce a quick reaction according to the network performance dynamic.

3) More advanced selection of which packet to drop can be implemented, for example, to try and prevent a large amount of consecutive voice packet.

5. Packet Aggregation for VoWLAN

On the one hand, the analysis in Section 3.3.1 clearly shows that sending small packets such as individual voice packets in IEEE802.11 WLANs utilises the wireless channel very inefficiently because the transmission overheads are proportionately much larger than sending large packets. On the other hand, there will be many small voice packets sent by the VoIP applications during a voice call between two users. The inefficiency of wireless channel utilisation persists as long as the voice call goes on. It is sensible to aggregate small packets into a bigger block, provided that this does not result in too large a probability of bit errors per aggregated block that is sent. Moreover, sending multiple packets in one transmission opportunity may also reduce the number of packets being accumulated in the outgoing buffer under congested conditions. However, for supporting interactive voice calls, aggregation must not increase the transmission delay, which can reduce the QoS or even make the voice conversation unusable.

This chapter presents a packet aggregation method, Small Packet Aggregation in Wireless Networks (SPAWN), which works at the data link layer. This method is designed to:

1) Improve the transmission efficiency for IEEE802.11 WLANs by aggregating small packets into a larger block;

2) Reduce the delay accumulation in congested conditions;

3) Be aware of the latency requirement for real-time voice and media applications and minimise the delay for small packets waiting for aggregation.

Section 5.1 provides the numerical explanation of how the delay accumulation can be resolved by packet aggregation in congested conditions. Section 5.2 reviews the existing uses of packet aggregation over different network layers. Section 5.3 introduces the SPAWN scheme, which can be implemented at the data link layer. Section 5.4 presents the results of the testing and evaluation of the SPAWN scheme.

5.1. Numerical Analysis

Equation 3.4 in Section 3.3.3 predicts the growth rate ($\boldsymbol{G_{/S}}$) of outgoing buffer occupation in congested conditions. If the aggregation algorithm is applied $\boldsymbol{G_{/S}}$ can be expressed using the equation below:

$$\boldsymbol{G_{/S} = \frac{1}{T_P} - a \cdot \frac{P_{MAX/s}}{N_{call} \times 2}}$$ ***Equation 5.1***

Where: the component of $\boldsymbol{\frac{1}{T_P}}$ *represents the number of packets arriving to the outgoing buffer per second; the component* $\boldsymbol{a \cdot \frac{P_{MAX/s}}{N_{call} \times 2}}$ *represents the number of packets being sent out from the outgoing buffer per second.*

In Equation 5.1, the term $\boldsymbol{a}$ represents the average number of packets being aggregated into each larger block and transmitted at once. This equation represents that by packet aggregation in congested conditions, the transmitter can send out multiple packets in one transmit opportunity. If the factor $\boldsymbol{a}$ is large enough it is possible for the resulting buffer growth rate $\boldsymbol{G}_{/S}$ to become zero or negative. If $\boldsymbol{G}_{/S}$ equals zero the occupancy of the outgoing buffer stops growing. If $\boldsymbol{G}_{/S}$ becomes negative the occupancy of the outgoing buffer reduces. As a result, the delay accumulation in congested conditions may be resolved by packet aggregation.

5.2. Existing Uses of Aggregation

This section briefly reviews existing packet aggregation strategies for wireless networks categorized by the Open Systems Interconnection (OSI) network layer each method operates at.

5.2.1. Transport Layer Aggregation

VoIP applications almost always use UDP rather than TCP for transport. Because UDP is fire and forget, aggregation of UDP datagrams in an end-to-end manner is not often applied. TCP is not used in most cases because it will re-transmit datagrams repetitively at the source until an acknowledgment (ACK) is received. At the destination, TCP will not forward any traffic to the application layer until a missing datagram has arrived. The problem of late arriving and lost datagrams at the destination and re-transmission of these by the source adds delay or latency, which varies because TCP uses an adaptive timing algorithm but the extra latency is often too long for even a single re-transmission, given that voice conversations are known to break down when the round trip time (RTT) exceeds for example, 300 ms.

TCP uses the Nagle algorithm [79] to aggregate byte streams from applications into datagrams for transmission. This algorithm waits before sending the datagram in order to try and send larger datagrams, thus aggregating the data. TCP limits its aggregation to data from a single port to which an application is attached. Data for different applications being sent to the same destination are always processed separately. Because TCP works end-to-end, data being forwarded on behalf of other network devices to the same port and application on the same destination device cannot be aggregated with locally generated data in the transport layer.

UDP data aggregation does not happen at the transport layer. UDP datagrams may be aggregated in the network, data-link or the physical layer. At the senders transport layer, UDP simply takes a vector of bytes and sends it. Applications may aggregate data before sending it to the transport layer.

SCTP [80, 81] is an alternative to TCP support aggregation in the form of chunk bundling. SCTP connections are called associations. Within an association a wide range of data can be sent in packets using concatenated chunks. To support real-time traffic better than TCP does, SCTP can be told to forward packets arriving at the destination end of the association straight away (in or out of order) or to wait so chucks are forwarded in sequence number order (in-order). Like TCP, re-transmissions are supported based on timeouts or arrival of several selective acknowledgments (SACK) showing one or more gaps in the sequence of received data.

Kim et al. [82] introduced three different packet aggregation schemes for increasing the number of VoIP calls in a wireless ad-hoc network. These schemes are distinguished by where the aggregation takes place: End-to-End (transport layer or below), Hop-by-Hop (data-link layer or below) and Accretion or further aggregation in the path toward the destination node (network layer or below). These three methods can be implemented at various layers, as shown by the bracketed comments above. All of these schemes will force the transmissions to wait in order to aggregate as many packets as possible, then transmit the large packet when either the maximum waiting time or aggregation size threshold has been exceeded.

5.2.2. Network Layer Aggregation

One of the major problems for packet aggregation at the transport layer is the difficulty of performing inter-application stream aggregation. This is because the main transport layer protocols TCP and UDP are each concerned with individual application streams.

Facilities dealing with multiple application streams are provided at the network layer. Choice between end-to-end and point-to-point aggregations is another advantage of aggregation algorithms being implemented at the network layer. Furthermore, the network layer protocols are, like all the higher network stack layers, usually implemented as software in the routers. Aggregation algorithms at this layer could easily be deployed in most IP capable devices.

Raghavendra et al. [83] proposed an end-to-end packet aggregation algorithm at the network layer called IPAC. This algorithm collects packets from the transport layer for the same destination and aggregates them into a larger sized IP packet. The Maximum Concatenation Size (MCS) and the Maximum Concatenation Interval (MCI) are defined in IPAC. If the aggregated IP packet is smaller than the MCS, it will wait for more packets for the same destination until the oldest aggregated packet in the block has been waiting for the MCI period. Then this block will be sent down to the data-link layer. Their evaluation results show that IPAC can increase the system throughput of a WLAN working especially well for CBR traffic. The trade-off is the increase in the delivery latency due to waiting.

PAC-IP is another network layer packet aggregation scheme proposed by Kliazovich and Granelli in paper [84]. The difference from IPAC is that it aggregates the packets point-to-point. It checks routed packets for a shared next hop IP address and aggregates the IP packets being sent to the same next hop into shared blocks. Like IPAC, if a block is smaller than MCS, it will wait for a maximum MCI period before trying to collect more packets. PAC-IP also introduces a QOS classifier for QOS consideration. If the traffic stream is time sensitive, then the packets will not go through the aggregator to minimize their latency. Their evaluation results show PAC-IP achieves WLAN throughput close to the maximum achievable value (about 5.5 Mb/s using 802.11b at 11 Mb/s) even if the application generates small packets.

5.2.3. Data-Link Layer Aggregation

There are two sub-layers at the data link layer. The upper sub-layer is Logical Link Control LLC sub-layer. The lower sub-layer is the Medium Access Control (MAC) layer. The LLC sub-layer is responsible for buffering and packet scheduling. The data link layer is closer to the wireless channel, which is one of the most important resources in a WLAN. The LLC sub-layer can detect changes in the wireless channel state earlier than layers above, e.g. detecting and reacting to congestion. Hence the LLC sub-layer is a good place to implement packet aggregation. The MAC sub-layer is concerned with the transmission of individual packets. Some packet aggregation algorithms also involve MAC sub-layer issues such as packet structure. However, they are not core issues for packet aggregation. Furthermore, as MAC protocols are normally implemented in hardware for most IEEE802.11 devices, any modification of the IEEE802.11 MAC protocols makes the implementation much more difficult.

A number of proposals aggregate packets for the same receiver in the LLC sub-layer buffer into a super-frame before transmission [85, 86]. These also modify the 802.11 MAC protocol to check the CRC of each embedded packet after receiving a super-frame. The original CRC in each IEEE802.11 MAC frame was removed. These schemes require modifications to 802.11. The ACK frames are modified to selectively indicate which parts of the super-frame required re-transmission. Kai-Ten and Po-Tai in [85] aggregate whole MAC frames, whilst PAC-IP [84] at the network layer only aggregates the IP packets complete with their calculated CRCs. The evaluation results show that both methods can achieve significant system throughput when using the highest available data rate e.g. 11Mb/s in IEEE802.11b and 54Mb/s in 802.11g and a.

Another group of proposals [87-91] tried to first predict the optimum frame size and data rate, it then aggregate small packets into the optimum size of packet for transmission. These algorithms are designed to achieve the highest possible network throughput depending on the wireless channel condition such as, signal to noise ratio (SNR).

IPAC [83] at the network layer proposed a packet aggregation algorithm similar to PAC-IP, but implemented at the LLC sub-layer. Furthermore, a better QOS classification scheme is proposed. There are four different QOS classes with different MCIs. The MCIs are also dynamic, being controlled by a channel estimator. Their evaluation results show this algorithm improves the WLAN throughput close to the maximum possible values (about 5.5Mb/s using 802.11b at 11Mb/s) when transmitting small application packets.

Karrlsson et al. in 2007 and 2009 [92, 93] proposed similar schemes for use in wireless mesh networks for packet aggregation for VoIP traffic over UDP and TCP respectively. Each packet coming into the LLC sub-layer buffer is assigned two parameters in the outgoing buffer; the incoming time and the maximum delay that it can wait. The schemes try to aggregate as many packets as possible before any of the packets in a block reach their maximum delay. The 2007 paper scheme supported up to 30 VoIP calls using UDP in an 802.11b network. The 2009 scheme supported up to 20 VoIP calls using TCP in an 802.11a network.

Jeong et al. [94] proposes several algorithms for improving VoIP performance over 802.11b infrastructure networks. One of the algorithms aggregates every other packet to the same destination in the LLC sub-layer's outgoing buffer before sending a VoIP packet. This algorithm introduces a so called "piggyback" scheme to balance the load difference between the AP and its clients. The size of the outgoing buffer at the AP is increased to 500 packets. The 802.11 MAC protocol is modified for further improvement. At the client side, if the number of aggregated packets in a block (B_n) is less than the number in the previous block (B_{n-1}), the client will contend for the wireless channel normally. Otherwise the client will transmit straight away after DIFS. A selective ACK is used for the aggregated block transmission. Their evaluation result shows that this scheme can support up to 40 VoIP calls in an IEEE802.11b infrastructure WLAN. This result appears questionable because it is likely to cause collisions between competing voice transmissions without any deferring.

In wireless mesh networks, the FUZPAG [95] system introduced fuzzy control on the maximum aggregation wait time. The maximum time a block waits for more packets is influenced by the ratio of the wireless channel being busy, the change of this ratio and the difference in traffic loads between the transmitter itself and its neighbours. FUZPAG can be implemented in Linux across both user space and kernel space. Their evaluation shows FUZPAG can support up to 20 VoIP calls with average end-to-end latency less than 20 milliseconds using IEEE802.11a at its 6Mb/s base rate.

Frantti and Koivula [96] introduced a fuzzy expert system to take the packet error rate and the change of the packet error rate into account. The appropriate MAC frame size is estimated so as to optimize the network throughput. Their evaluation results show that this scheme can support up to 15 G.729 VoIP connections (calls) “well” in IEEE802.11b networks at 11Mb/s.

Paper [97] introduced a control scheme to vary the aggregated block size based on the battery charge level in order to improve energy efficiency. Their evaluation result shows the energy efficiency using this scheme is improved 20%-30% more than fixing the block size to be the maximum MAC data frame size, although the resulting maximum network throughput is 10%-20% less as a result.

5.2.4. Application Layer Aggregation

The aggregation at the Application and Presentation layers is not performed packet wise, as the data has not yet been encapsulated into packets. Instead, by increasing the interval of voice data being packetized, more data will be encapsulated into each UDP/RTP packet. There are many investigations of the voice packetizing interval, for example [2, 6, 9]. These works concluded that increasing the packetizing interval can significantly increase the number of VoIP conversations that can be supported in a wireless channel.

However, there are several drawbacks for performing aggregation at such high layers. First of all, only end-to-end aggregation is available at such layers. Secondly, the flexibility is very low. Multi flow aggregation is not supported at those layers above the transport layer. There is limited information available to adapt the packetizing interval (block size). The RTCP feedback may not be able to report the most recent channel states in order to select the appropriate packetizing interval. The lower layers cannot adjust the block size as the blocks are encapsulated as large RTP packets. Finally, a packetizing interval larger than 30 ms is not common for VoIP in wired LANs and Internet.

5.2.5. IEEE802.11n Frame Aggregation

The IEEE802.11n [49] standard also specifies the use of frame aggregation to reduce overheads and enhance network throughput. Paper [98] provided an introduction to frame aggregation in IEEE802.11n. There are two aggregation scheme specified by IEEE802.11n, which are A-MSDU and A-MPDM. The A-MSDU scheme waits for the packets (here each packet is called an MSDU[4]) send to the same receiver from upper layers for a certain period and aggregates them into a large MAC frame (here each packet is called an MPDU[5]) before transmission. A-MPDU scheme encapsulates the packets from upper layer into MPDUs first. Once the transmitter has a transmit opportunity, the A-MPDU scheme will aggregate several MPDUs into a larger MAC frame and transmit.

[4] MAC Service Data Unit
[5] MAC Protocol Data Unit

Considering these two aggregation schemes with the latency requirement for VoIP traffic, A-MPDU would be a better option, because there is no forced delay when waiting for aggregation. However, the aggregation for MPDU aggregation requires modifications to the original IEEE802.11 frame structure, and receiving process. For example, multiple MAC headers and CRC paddings have to be processed during the de-aggregation at the receiver. It is not compatible with other IEEE802.11 MAC schemes. In contrast, the A-MSDU scheme can be implemented at the LLC sub-layer buffer, which does not require modification to the original IEEE802.11 MAC. Hence this thesis is going to present a new packet aggregation scheme to combine the advantages of both A-MSDU and A-MPDU.

5.2.6. Summary on Packet Aggregation

Table 5.1 provides a brief summary on the existing uses of packet aggregation schemes. Table 5.1 is categorised by the layers work at, the type of links, the support for multiple streams, whether a forced wait is required and whether modifications to IEEE802.11 MAC protocol is needed.

Table 5.1 Summary on Existing Uses of Packet Aggregation

Solution	Layer	Link Type	Multi Streams	Forced Wait	Modifications to IEEE802.11 MAC
[79] Nagle	TRA[6]	end-to-end	No	No	No
SCTP [80, 81]	TRA	end-to-end	No	No	No
[82] verion1	TRA/NET[7]	end-to-end	No	Yes	No
[82] verion2	NET/LLC	point-to-point	Yes	Yes	No
[82] verion3	NET+LLC	n/a	Yes	Yes	No
IPAC version1 [83]	NET	end-to-end	No	Yes	No
PAC-IP [84]	NET	point-to-point	Yes	Yes	No
[87-91]	LLC+MAC	point-to-point	Yes	Yes	n/a
[85]	LLC+MAC	point-to-point	Yes	No	Multi CRC Selective ACK
[86]	LLC+ MAC	point-to-point	Yes	No	Multi CRC Selective ACK
IPAC version2 [83]	LLC	point-to-point	Yes	Yes	No
[92, 93]	LLC	point-to-point	Yes	Yes	No
[94]	LLC and MAC	point-to-point	Yes	Yes	adaptive ACK Multi CRC Medium Access
[95]	LLC	point-to-point	Yes	Yes	No
[96]	LLC	point-to-point	Yes	Yes	No
[97]	LLC	n/a	Yes	Yes	No
A-MSDU[49]	LLC	point-to-point	Yes	Yes	No
A-MPDU[49]	MAC	point-to-point	Yes	No	New IEEE802.11
Proposed SPAWN	LLC	point-to-point	Yes	No	No

We can see from Table 5.1 that the end-to-end algorithms that are usually designed to work at the higher layers (e.g. the transport layer) can be implemented as software. Furthermore, the aggregation and de-aggregation only happen at both ends of the traffic stream. In a multi-hops transmission path the intermediate nodes, like the routers, do not need to deal with the aggregated packets. However, the end-to-end algorithms do not support aggregation amongst multiple traffic streams to a

[6] Transport Layer
[7] Network Layer

single receiver node, because the aggregation of packets only happens for the network traffic sent from the same destination port.

Another disadvantage of implementing packet aggregation algorithms at higher layers for VoWLAN is the algorithm cannot react quickly to the network congestions. For example, the transport layer packet aggregation algorithms such as Nagle and STCP, they start to aggregate packets when the transport layer buffer grows. In this case, the buffers at the lower layers may already be full. Hence several previous packets may miss the chance to be aggregated.

In contrast, the point-to-point algorithms are usually implemented at the lower layers and only consider the transmission between two physical nodes in a single hop. Therefore point-to-point algorithms can aggregate the packets from different network traffic streams as long as they are sent to the same physical (MAC) address. One of the disadvantages of point-to-point algorithm is the packets are aggregated and de-aggregated in each hop. It increases the computational complexity if the transmission path has multiple hops. Furthermore, some of the point-to-point algorithms' implementations require modifications to the original IEEE802.11 standards. As a result, such algorithms may not be able to be implemented on the off the shelf WiFi devices.

The new SPAWN algorithm mentioned in the introduction part of this chapter fills the research gap. It works at the LLC sub-layer and performs point-to-point packets aggregation. In order to avoid any extra transmission latency for VoWLAN, SPAWN does not force the packets in the LLC sub-layer buffer to wait for aggregation once they have the opportunity to transmit. Furthermore SPAWN does not require any modification to the original IEEE802.11 standards. Hence it is possible for it to be implemented for use with almost any of the WiFi devices in the market.

The evaluation results from the existing studies on packets aggregation are based on a variety of topology scenarios. However, because of the implementation complexity and the lack of code availability, it is not possible to provide a crosswise comparison amongst all these algorithms based on the scenario used in this thesis.

5.3. New LLC sub-layer Aggregation Scheme

A new packet aggregation scheme called Small Packet Aggregation for Wireless Networks (SPAWN) is presented here. In order to minimise the packet collecting process delay SPAWN opportunistically exploits the transmitter's behaviour, which uses CSMA/CA scheme in IEEE 802.11 DCF protocol. Rather than forcing the outgoing packets to wait for the more of the others for aggregation, SPAWN only aggregates packets when the transmitter is waiting for a transmission opportunity in the LLC sub-layer buffer.

If the medium is not busy, each packet may be transmitted as soon as it arrives in the buffer with no extra wait time imposed. In this case, the wireless channel efficiency is not significantly affecting the service quality. Hence, there will be no aggregation process required. When the medium is busy, transmitters have to contend to transmit. Because the packet generation interval is quite small (20 ms in our scenarios) in VoIP applications, if the traffic load is high enough to cause congestion, there could be several voice packets accumulated in the outgoing buffer at the LLC sub-layer before a transmitter wins the contention for transmission (has the transmission opportunity). Apart from the delay for wireless channel contention, there will be no extra delay required to collect packets.

5.3.1. SPAWN Packet Flow

In order to implement the SPAWN scheme, there are mainly three components required, the aggregator, the de-aggregator and the outgoing buffer. Considering that the IEEE802.11 standards cover the specifications at the MAC layer and below. All of these components are implemented at the LLC sub-layer. This plan does not require any modification to the IEEE802.11 standards and therefore SPAWN is feasible to be implemented based on existing off the shelf WiFi enabled devices.

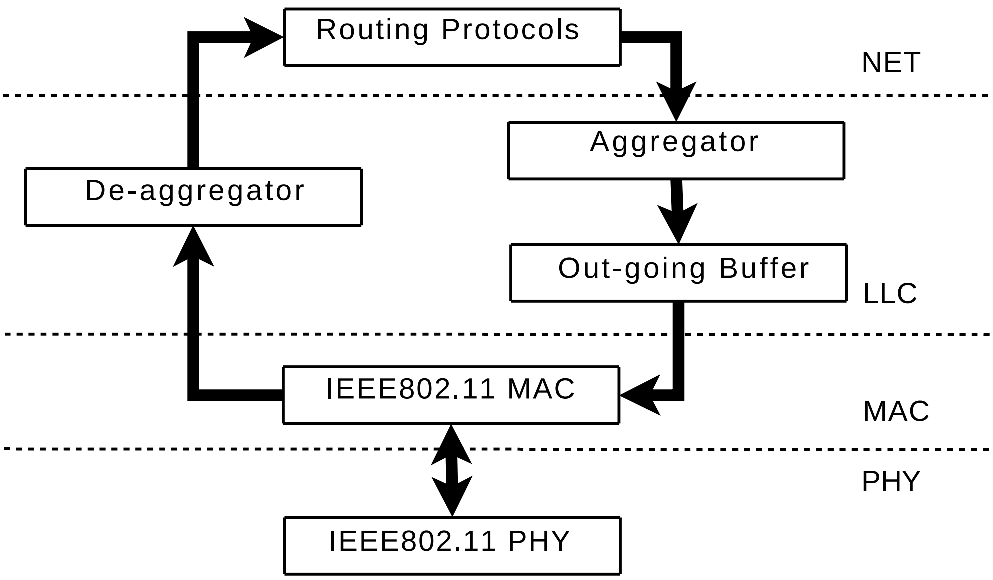

Figure 5.1 SPAWN Packet Flow

Figure 5.1 shows the flows for both outgoing and incoming packets using SPAWN. When the IP packets are sent by the routing protocol at the network layer, they will be processed by the aggregator. The aggregated blocks sent down by the outgoing buffer will be treated as normal IEEE802.11 packets. When the MAC protocol receives packets from the physical layer, the payload will be passed to the LLC sub-layer de-aggregator. The de-aggregator will then separate the IP packets by the indication of block offsets and pass them to the network layer routing protocol.

5.3.2. SPAWN Block and Outgoing Buffer Structures

Small (less than the maximum MAC payload length) packets from upper layer to the same destination MAC address will be aggregated into a block until this block reaches the maximum MAC payload length. The block structure is encapsulated within the IEEE802.11 MAC packet payload. Blocks are queued in the outgoing LLC sub-layer buffer.

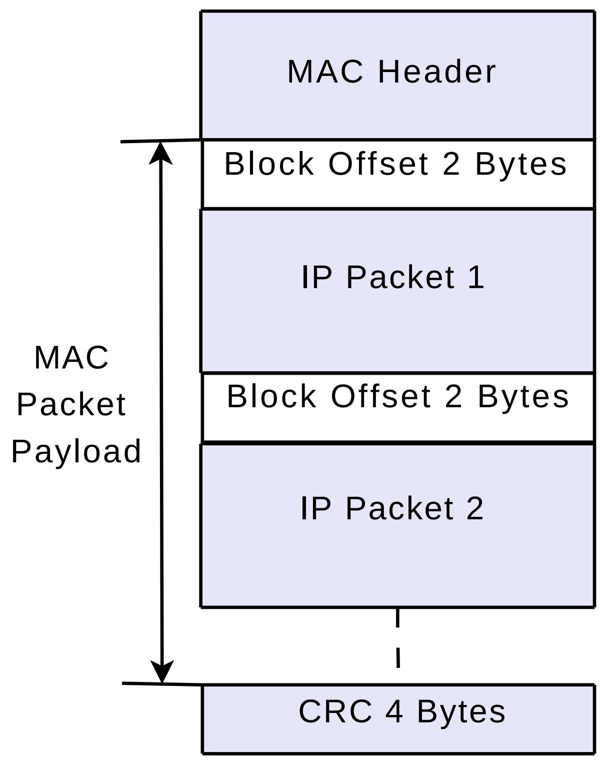

Figure 5.2 SPAWN Block Structure

As shown in Figure 5.2, IP packets are concatenated within a block. The block is offset, forming a linked-list of IP packets in the block. There is a 2 bytes block offset at the beginning of each IP packet. It indicates the length of the following IP packets. Both IPv4 and IPv6 headers use 2 bytes to indicate the length of the IP packets [26, 99]. Therefore, 2 bytes block offset is sufficient to indicate any length of an IP packet. At the end of each IP packet there is another block offset expected unless the end of this block is reached. The number of IP packets within each block can vary. This block implementation is fully compatible with the MAC packet structure defined in IEEE802.11 standards and these are sent as normal packets. Aggregation is therefore transparent to both higher and lower layers in the network stack.

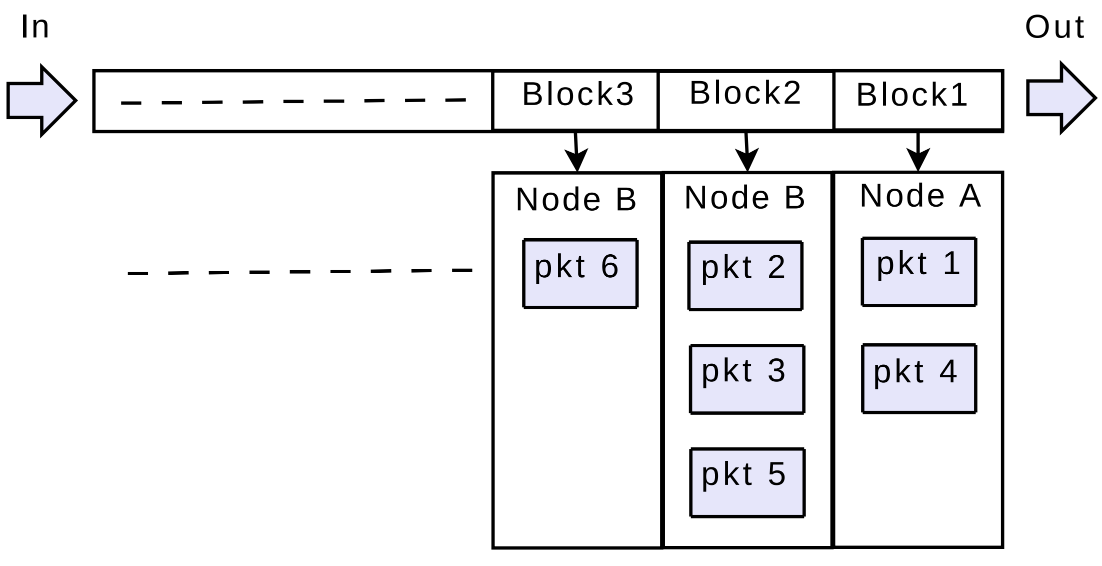

Figure 5.3 SPAWN LLC sub-layer Outgoing Buffer

A special outgoing buffer for blocks is designed to enable the aggregation of packets while the transmitter is waiting for an opportunity to transmit. As shown in Figure 5.3, the blocks in the buffer are in first in first out order. Each block is encapsulated into an 802.11 MAC packet and is ready to be sent out. There can be multiple blocks to the same MAC address in the buffer at a given time. As Figure 5.3 shows, because the first block to Node B containing IP packets 2, 3 and 5 reaches the maximum MAC packet payload length packet 6 is encapsulated into a new block to Node B.

5.3.3. SPAWN Aggregation Process

The aggregator will implement the packet aggregation process shown in Figure 5.4. When a packet is sent down to the LLC sub-layer by the routing protocol, the aggregator examines the receiver's MAC address and checks to see if there is any block in the outgoing buffer for this receiver. If there is a block already created for this receiver and the post aggregation block size is still within the maximum MAC packet payload size, the aggregator will append this packet into the existing block. If there is no block created for this receiver or the block will be over the maximum payload size by adding this additional packet, the aggregator will create a new block for this outgoing packet. When the transmitter has a chance for transmission, the buffer will de-queue the first block to the MAC protocol. As the blocks are in a first in first out order, the fairness to the receivers is maintained.

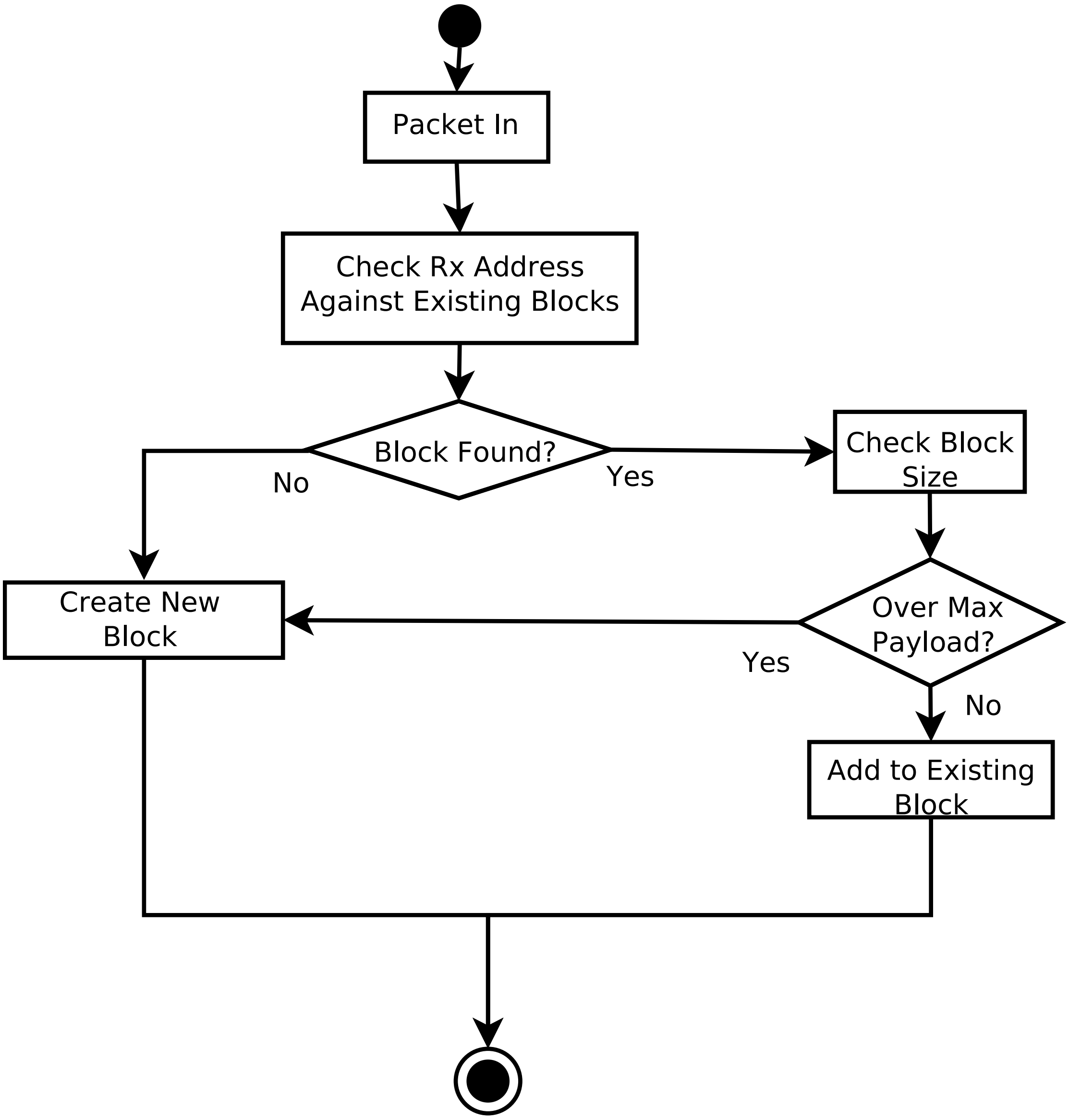

Figure 5.4 SPAWN Aggregation Process

5.3.4. SPAWN Co-operation with IEEE802.11e

IEEEE802.11e specifies multiple priority queues are used for packets with different QoS requirements. SPAWN's aggregator and outgoing buffer are also compatible with IEEE802.11e. There are different plans of SPAWN implementation available in order to co-operate with multiple priority queues. One implementation would use a SPAWN aggregator for each priority queue. However, this solution cannot aggregate packets of mixed traffic types to the same receiver. Another solution would use only one SPAWN aggregator feeding all the priority queues. In this second solution, a block can contain packets of mixed traffic types. Each block is classified according to the highest priority of the packet that it contains.

5.4. Testing and Evaluation

In order to test and evaluate the performance of SPAWN, the aggregator, the de-aggregator and the outgoing buffer are implemented on top of the IEEE802.11 MAC protocols and simulated using NS-2. In order to compare the network performance for the VoIP traffic with and without SPAWN, the simulation configurations were identical to that defined in Section 3.5. The simulations were run in two scenarios, the distributed WLAN and the infrastructure WLAN. The former was identical to that defined in Section 3.5. The later will be described in Section 5.4.2.

5.4.1. Distributed WLAN Scenario Results

Figure 5.5 shows the average transmission latencies measured in the simulations of distributed IEEE802.11 WLAN, using the different codecs with and without applying the packet aggregation scheme. The results show that without applying the packet aggregation scheme, the average transmission latency will increase significantly as soon as the number of simultaneous VoWLAN conversations is greater than 12 as predicted by Equation 3.2 in Section 3.3.2. With 12 conversations and no aggregation, the wireless channel is saturated and the latency is too large for normal conversation. When the packet aggregation scheme is applied, the average transmission latency stays low (less than 100 ms) by achieving higher channel efficiency. The result also shows that no noticeable extra delay is introduced by the packet aggregation scheme. Another observation from this result is the performance using either of the G.711 and G.729 codecs is very similar. The conclusion drawn from this is that as predicted in Chapter 3, extra voice compression does not significantly affect the transmission latency in IEEE802.11 WLANs.

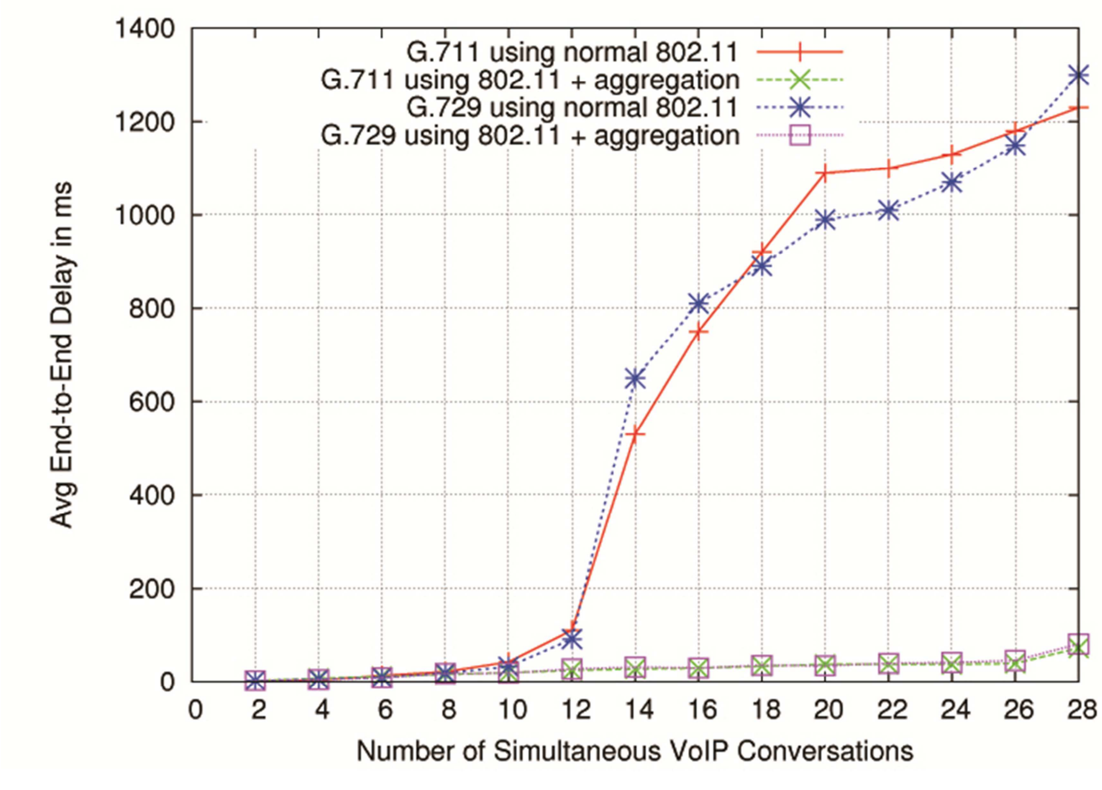

Figure 5.5 Average Transmission Latency in Distributed WLAN

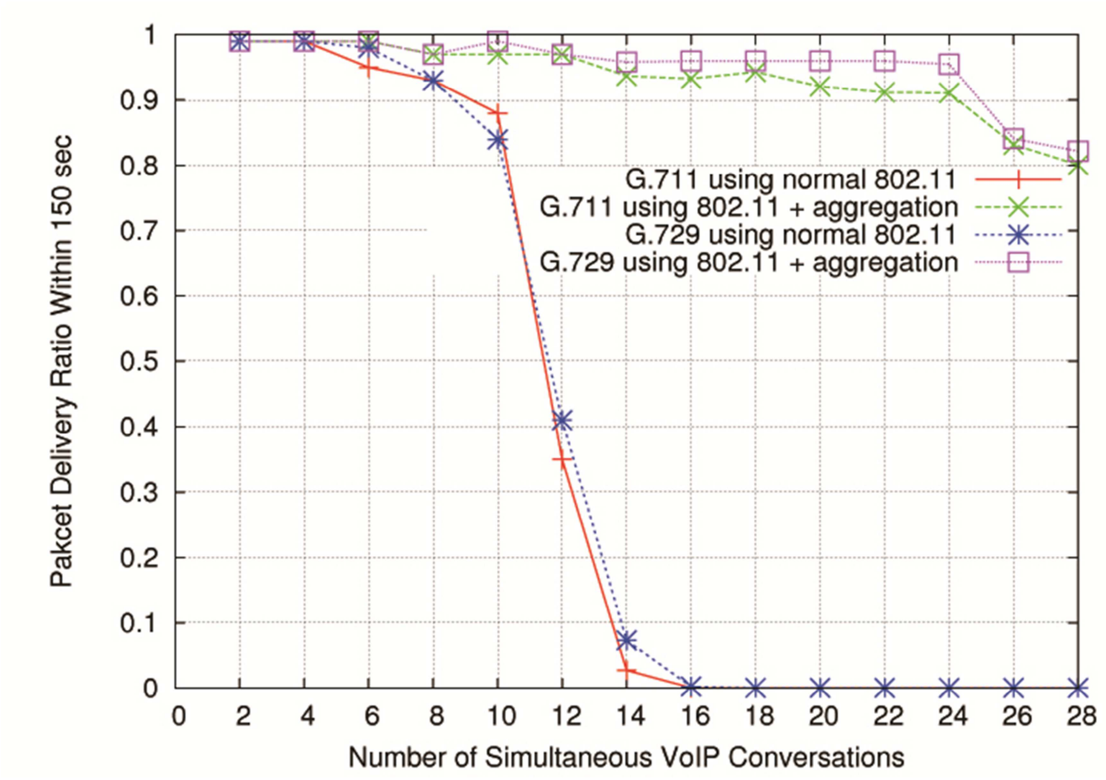

Figure 5.6 Packet Delivery Rate within 150 ms Delay Budget in Distributed WLAN

As indicated in Section 2.2.3, we regard a VoWLAN call as having acceptable quality if 90% of the voice packets are delivered within 150 milliseconds delay budget. The simulation results in Figure 5.6 show that without applying packet aggregation, the packet delivery rate within 150 milliseconds drops significantly shortly before the wireless channel is saturated, at around 10 conversations. The voice packets are either suffering large buffering delays or are dropped by the transmitter. VoWLAN conversations are difficult to continue at this quality. When the packet aggregation scheme is applied, the wireless ad-hoc network can support up to 24 simultaneous VoWLAN conversations with good quality, with G.729 slightly outperforming G.711 now, presumably because G.729 results in more packets per block being aggregated than G.711.

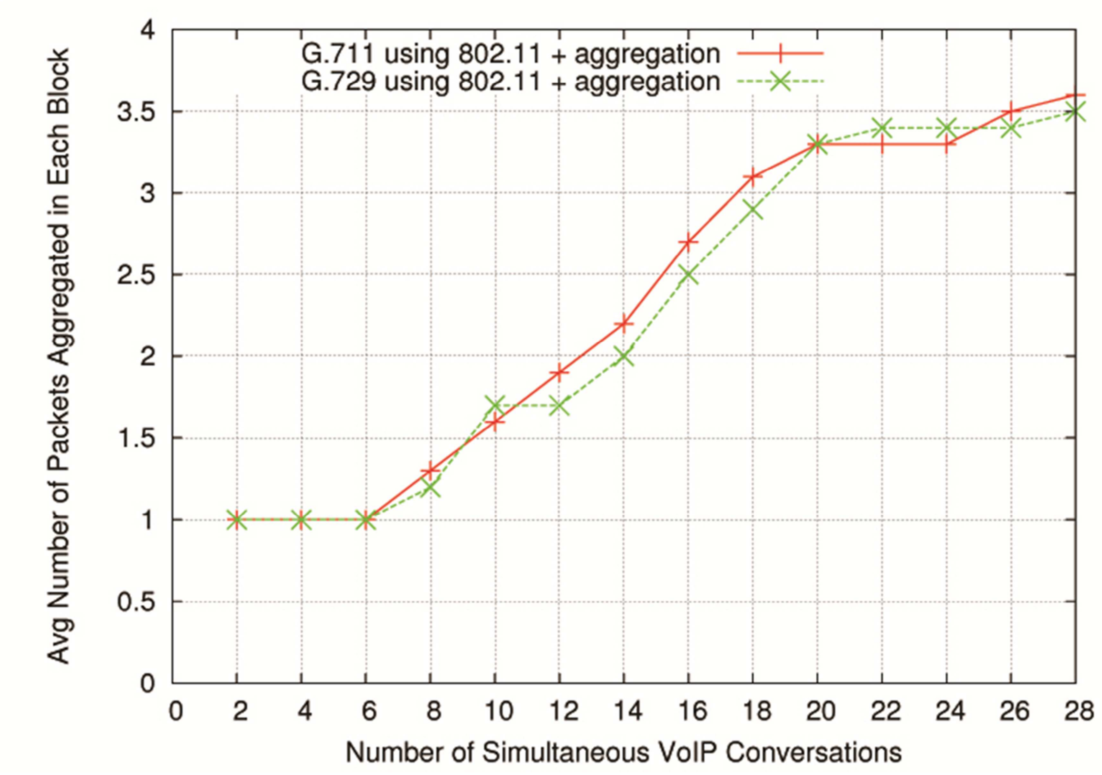

Figure 5.7 Average Number of Voice Packets Aggregated per Block

Figure 5.7 shows the number of voice packets aggregated into each block is increasing when the number of simultaneous VoWLAN conversations increases. This result indicates that no aggregation is required when the wireless channel is not busy. When the number of VoWLAN conversations grows, the packet aggregation scheme starts to be significant.

5.4.2. Infrastructure Scenario Results

The second simulation scenario is modelling IEEE802.11 infrastructure WLAN. Both the AP and the stations are using IEEE802.11 DCF. One wireless node is placed in the middle of the 50mx50m space acting as an AP. The other 49 nodes are randomly distributed as above and become the client stations. A TCP downlink will be generated by the NS2 TCP agent, which is configured to send 1000 byte packets from the AP to one client station. This simulates normal application traffic from the Internet via the AP to the station. This packets size allows a few of these TCP packets to be aggregated by SPAWN along with some voice packets with the block size still less than the typical Maximum Transmission Unit (MTU) for the MAC frame. All the stations including the TCP active one are randomly selected to setup G.711 voice conversations between themselves and the AP.

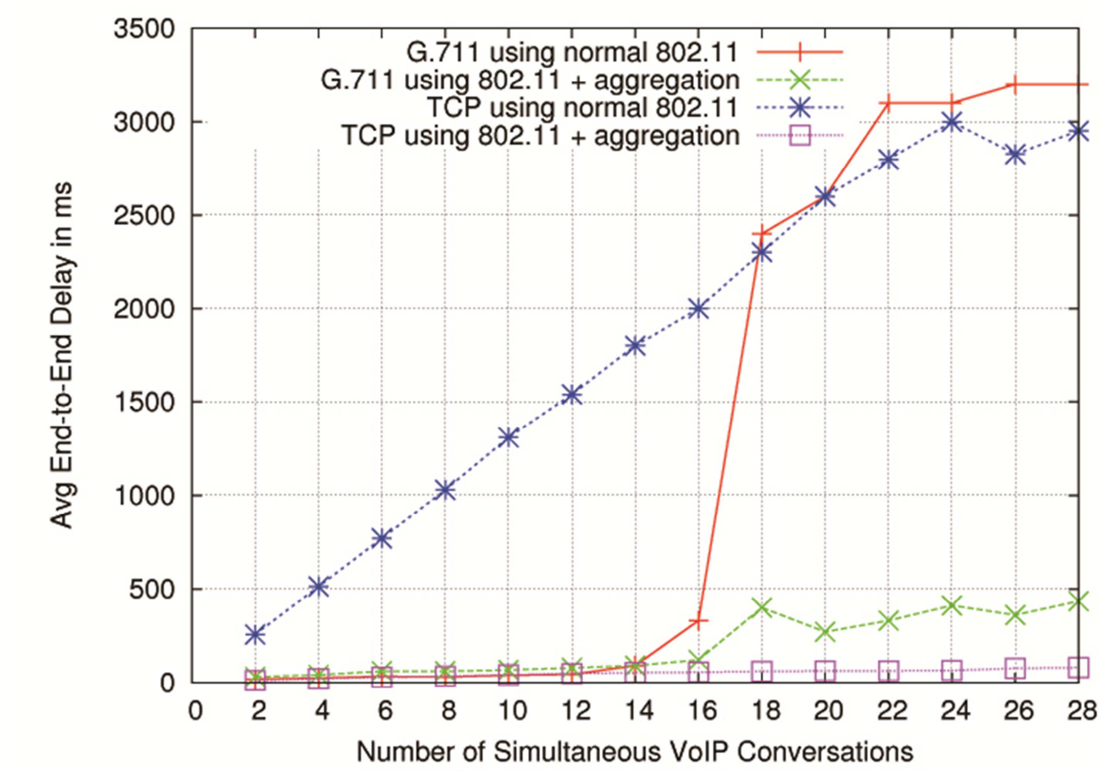

Figure 5.8 Average Transmission Latency in Infrastructure WLAN

Figure 5.8 shows the transmission latency for both the VoIP traffic and the TCP traffic both with and without SPAWN's aggregation. The result shows that without aggregation, as the number of simultaneous conversations increases, network congestion happens later in the infrastructure scenario than in distributed WLAN scenario. This is because overall there are fewer transmitting nodes contending for the wireless channel than in the ad-hoc configuration. In order to support N simultaneous VoIP calls, there are one AP and N satellite nodes in the infrastructure scenario instead of 2N nodes in the fully distributed scenario. For example in order to support 16 simultaneous VoIP calls, there are 17 transmitting nodes including the AP in the infrastructure scenario and 32 transmitting nodes in the fully distributed scenario.

However, the average end-to-end delay of the infrastructure scenario is more than twice that of the ad-hoc scenario once congestion occurs. The cause of the extra end-to-end latency is the unbalance between the uplink traffic loads to the AP and its downlink transmission opportunities. Of course, the simulated AP in this scenario does not have QoS prioritising scheme such as the IEEE802.11 TXOP applied. As expected, the end-to-end delay of the TCP traffic without aggregation is gradually increasing as the number of simultaneous conversations increases. Aggregation delays the increase in end-to-end delay for the TCP traffic supporting either more conversations or more TCP traffic.

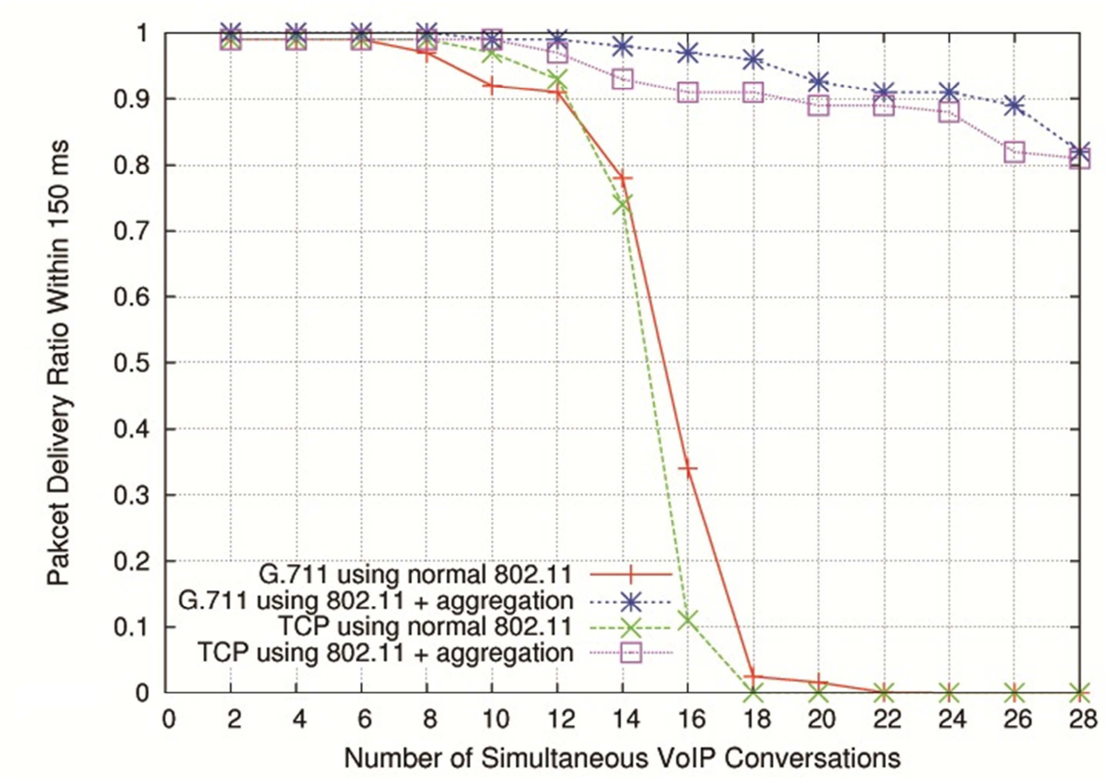

Figure 5.9 Packet Delivery Rate within 150 ms Delay Budget in Infrastructure WLAN

Figure 5.9 shows that the ratio of packets being delivered within 150 ms drops dramatically for both VoIP and TCP traffic without aggregation after congestion occurs. There are nearly no packets delivered within 150 ms when there are more than 20 simultaneous conversations operating in the network. After applying SPAWNs aggregation, the infrastructure network can support up to 24 simultaneous conversations, with the same TCP load, achieving over 90% of the packets being delivered within 150 ms.

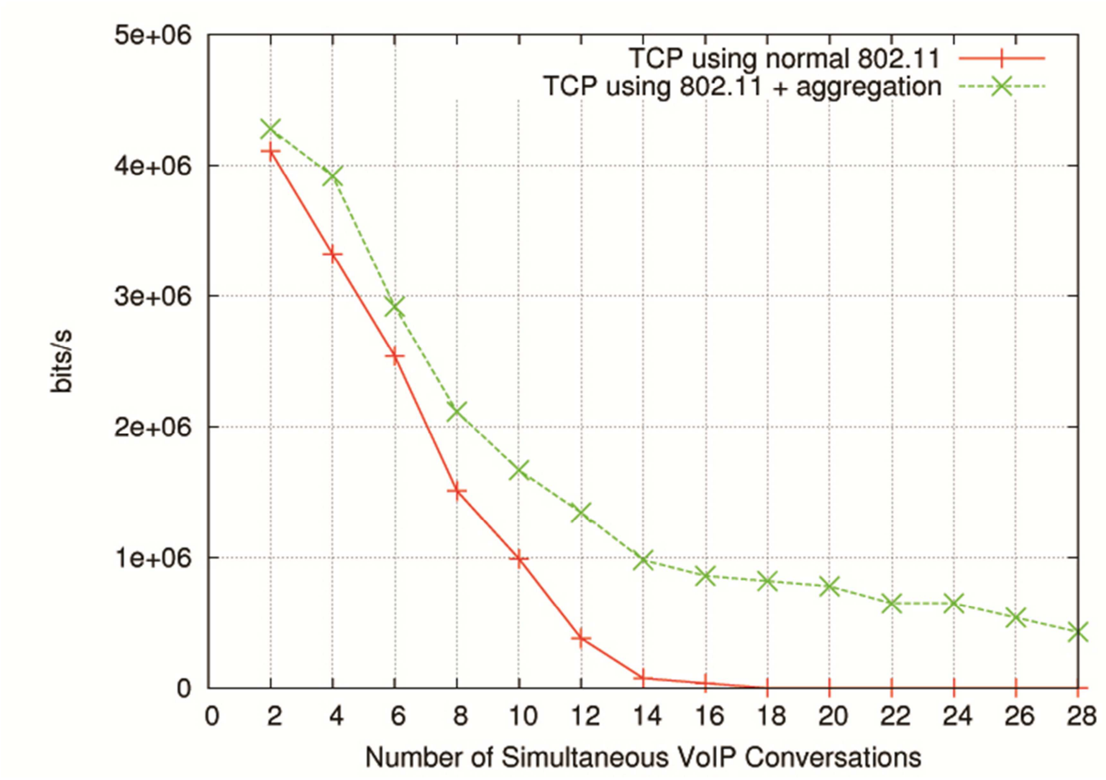

Figure 5.10 Infrastructure WLAN TCP Throughput versus. Number of Conversations

Figure 5.10 shows the TCP link throughput decreases as the number of simultaneous conversations increases. The figure shows that there is hardly any throughput via the TCP link when aggregation is not used when more than 18 simultaneous conversations are taking place. This is due to the TCP link decreasing its sliding window size when the network is badly congested. When SPAWN's aggregation is applied, the TCP throughput can be maintained at about 0.6Mb/s as some TCP packets are aggregated with the voice packets and get through.

5.5. Summary

According to Equation 5.1, packet aggregation can often be used to prevent delay accumulation in congested conditions by sending multiple packets in one transmission opportunity. The existing uses of packet aggregation schemes were analysed. Amongst the data link layer schemes two issues were found that are inappropriate to the scenario requirement in this research. Firstly, some of them [87-97] require a forced delay waiting for more packets to aggregate, which may result in additional delay. Secondly, the implementations of some aggregation schemes [49, 85, 86, 94] require modification to the original IEEE802.11 MAC standard.

The algorithm proposed in paper [92] does not require modification to the original IEEE802.11 standards but forces the packets to wait for aggregation. The evaluation results in [92] report that their aggregation algorithm can support up to 30 simultaneous VoIP calls in IEEE802.11b WLAN. However, the simulation scenario in their evaluation is very different from here. There are only two satellite nodes in their infrastructure WLAN. In that case, there are 15 VoIP traffic streams sharing the same physical point-to-point link in a single direction, which is very favourable to the point-to-point packet aggregation algorithm.

The aggregation algorithms proposed for VoIP in papers [95] and [96] also have no modification required to the existing IEEE802.11 standards. They can support up to 20 and 12 simultaneous VoIP calls respectively according to their evaluation results.

A new packet aggregation scheme, SPAWN was presented in this chapter to overcome these issues as one of the research contributions (Contribution 4). This solution is implementable between the routing protocols and MAC protocols and is compatible with current IEEE802.11 standards.

According to Equation 5.1, SPAWN is able to increase the term $\boldsymbol{\alpha}$ from 1 up to about 3.5 under the congested conditions in our scenario simulations. The results also show that SPAWN is also beneficial to TCP traffic in terms of transmission latency and packet delivery ratio within the sensible delay budget.

The evaluation results show that there is a considerable improvement in the number of simultaneous VoWLAN conversations possible in the IEEE802.11 WLANs. The two main objectives, improving transmission efficiency and minimising extra delay for packet aggregation can be achieved by implementing SPAWN. The evaluation results show that without packet aggregation as soon as the number of simultaneous VoIP calls exceeds 12, the average transmission latency increases dramatically and the number of packets delivered within 150 ms drops significantly in both IEEE802.11 distributed and infrastructure WLANs. However, after SPAWN is applied, both WLAN scenarios are able to support up to 24 simultaneous VoIP calls with the average transmission latency below 100 milliseconds and over 90% of the voice packets being delivered within 150 milliseconds.

So far we proved that the hypothesis defined in Section 1.1.4 is true. Both research contributions in Chapter 4 and this chapter (Contribution 3 and 4) provided solutions to improve the VoIP service quality and capacity of VoIP calls over distributed WLANs without modifications to the existing IEEE802.11 standards.

6. Conclusions and Discussions

As stated in Chapter 1, there is a growing use of both WiFi (IEEE802.11 WLANs) and VoIP service. Therefore, there are many possible scenarios where people use VoIP over IEEE802.11 WLANs. Most of the existing WiFi devices are based on IEEE802.11a, b or g standards. According to the existing research on VoWLANs the standards were found not efficient to carry VoIP service over wireless channels. This research is motivated by the needs of improving VoWLANs efficiency base on existing IEEE802.11 a, b or g devices and hence expand the VoIP service capacity for each wireless channel.

This thesis has investigated the capacity of VoIP calls based on fully distributed IEEE802.11 WLANs scenario. The analysis was carried out mainly by simulations. Simplified numerical analysis and real device experiments were auxiliary. This thesis also extended the investigation to the effect of network buffering in congested distributed WLANs. Delay accumulation was identified to be the cause of VoIP service quality collapsing in such conditions. A queue management strategy ACQ and a packet aggregation algorithm SPAWN were proposed in this thesis to mitigate the delay accumulation and improve the VoWLAN traffic efficiency. This chapter concludes this thesis, evaluates this research, discusses the limitations on this research and proposes some possible future works.

6.1. Conclusions

This section recasts the conclusions made by previous chapters to answer the research questions indicated in Chapter 1.

6.1.1. VoWLAN under Congested conditions

Chapter 3 answered the research question 1) Are the factors imposing the upper bound of VoIP capacity for infrastructure WLANs the same for distributed WLANs?

As indicated in Chapter 3, the VoIP capacity over distributed WLANs has been observed to be similar to a single infrastructure WLAN, which has been studied by many existing research projects. It is because in both WLAN topologies, wireless nodes implement IEEE802.11 DCF, which uses the CSMA/CA MAC scheme to access and share the wireless channel. The transmission efficiency of voice packets, the streaming patterns of voice traffic and the delay sensitivity of real time network applications are the major factors limiting the VoIP capacity over IEEE802.11 WLANs rather than the network topology.

Chapter 3 also answered the research question 2) How will the VoIP quality of service be affected in congested WLANs?

If the number of simultaneous VoIP calls exceeds the capacity of the wireless channel the network is congested. The analysis in Chapter 3 also indicated that in distributed WLANs VoIP performance will collapse from a low latency, high delivery rate to drop sharply to a high latency, low delivery rate beyond the boundary of the network becoming congested. Similar performance behaviour was concluded for infrastructure WLANs by Malone et al. in paper [18]. To extend this research finding, the delay accumulation phenomenon under congested conditions in the data link layers buffer was discovered and studied. When the network is congested, the outgoing buffer occupancy for the sender will increase over time as the VoIP application continuously generates the voice packets to send. As a result, the buffer delay increases.

Furthermore, the study on buffer delay in distributed WLANs here was not identical to the existing research in infrastructure topology. Most of the existing analyses on buffer delay focused on the AP in infrastructure WLAN. The buffer delay that they concerned is mainly caused by the unbalance transmission opportunities between downlinks and uplinks. These unbalance can be resolved by applying IEEE802.11e EDCA TXOP function to prioritise the AP's transmission. However, in distributed WLANs the buffer delay happens at all the STAs participating in VoIP calls. All STAs have the same transmission opportunity in the long term. Prioritising any STA is going to result in the unbalance of transmission opportunity. A new algorithm is needed to resolve the delay accumulation in distributed WLANs.

6.1.2. Active Cleaning Queue Management Strategy

Chapter 4 answered the research question 3) How can the collapsing in VoIP service quality be prevented in congested conditions?

Chapter 4 continued the analysis in Chapter 3 and indicated that in order to prevent the delay accumulation that results in the sharp dropping in VoIP service quality, a certain amount of voice packets need to be dropped from the outgoing buffer to cancel the growth in buffer occupancy in congested conditions. However, a precise selection of voice packets to drop is required in order to minimise the damage to the voice speech.

According to this conclusion, a new queue management strategy ACQ was proposed to actively clean the "stale" voice packets from the outgoing buffer. By applying this strategy ACQ drops the voice packets from the outgoing buffer that have little or no probability of being delivered on time. The transmission opportunities are therefore given to those "fresh" voice packets.

Currently, ACQ uses a hard boundary threshold in time $\mathrm{T_{MAX}}$ to identify whether a voice packet is "stale" for being stored in the outgoing buffer. Different values of $\mathrm{T_{MAX}}$ were tested in the simulations. The results show that the ACQ changes the VoIP service quality performance under congested conditions from collapsing to smooth reduction. The result also shows that there are correlations between the $\mathrm{T_{MAX}}$ values and the packet delivery ratio within time limits and average transmission latency. If the $\mathrm{T_{MAX}}$ is greater than the VoIP delay budget (e.g. 150 ms in this thesis) the ACQ cannot clean out all the 'stale' voice packets. If the $\mathrm{T_{MAX}}$ is less than the VoIP delay budget the ACQ may clean out some voice packets that still have a chance of being delivered on time.

6.1.3. Small Packet Aggregation for Wireless Network

Chapter 5 answered the research question 4) What is the solution at the data link layer for distributed WLANs to improve the VoIP traffic efficiency?

Chapter 5 extended the analysis in Chapter 3 and indicated that in ordinary IEEE802.11 WLAN, each time the sender has an opportunity to transmit, the occupancy of the outgoing buffer reduces by one ($S_{buf} - 1$). In congested conditions, the occupancy of the outgoing buffer is always greater than one. If there are many packets need to be sent to the same receiver, it is possible to send multiples of them in a single transmission opportunity ($S_{buf} - a$). If a small payloads are aggregated into a larger block and transmitted using the overhead of one single transmission, the transmission efficiency increases in approximately a folds. Furthermore, the delay accumulation might also be resolved if the resulting number of packets being sent per second greater than the buffer growth rate ($G_{/s}$).

According to the analysis in Chapter 5, a packet aggregation algorithm SPAWN was proposed. This algorithm instantly aggregates the packets available to send to the same receiver in the outgoing buffer and transmits in one large block. The aggregation and de-aggregation processes at both transmitter and receiver are also specified. Furthermore, the specified packet structure for the aggregated payloads allows the SPAWN algorithm to be implemented without modifications to existing IEEE802.11 standards.

6.2. Research Evaluations

6.2.1. Research Objectives and Hypothesis

Reviewing the objectives of this research, we achieved objective 1) and successfully demonstrated that the problems for limiting the VoIP capacity over a single infrastructure WLAN found in existing research also apply to the distributed WLANs. To extend the analysis on this problem, we also achieved objective 2) and carried out a deeper analysis of the network performance in congested conditions and identified that the voice packets delay accumulation in the outgoing buffer causing the collapse in VoIP service quality. Hypothesis 1 declared in Section 1.1.4 is proven true by the research outcome presented in Chapter 3.

Furthermore, in order to achieve objective 3), based on the research findings in Chapter 3 two novel solutions ACQ and SPAWN were proposed to prevent the delay accumulation and improve the VoIP service quality in congested conditions. The SPAWN algorithm can double the VoIP capacity per wireless channel for distributed WLANs. Both solutions can be implemented on off the shelf WiFi devices as no modification is required to the IEEE802.11 standards. The Hypothesis 2 declared in Section 1.1.4 is proven true by the research outcome presented in Chapters 4 and 5.

6.2.2. Research Contributions

Chapter 3 answered Question 1 and 2 and led to the research contributions below:

> **Contribution 1** We demonstrated that limitations on VoIP service quality in infrastructure WLAN, described by existing research publications, also apply to distributed WLANs.
>
> **Contribution 2** We provided an analysis to extend the existing research findings on the VoWLAN performance collapsing in congested conditions; and identified that the delay accumulation in outgoing buffer causes this collapsing.

Chapter 4 answered Question 3 and led to the contribution to knowledge:

> **Contribution 3** We proposed a new queue management strategy ACQ for WLANs to prevent the sharp drop in VoIP service quality in congested conditions.

Chapter 5 answered Question 3 and led to the contribution to knowledge:

> **Contribution 4** We proposed a new packet aggregation algorithm SPAWN for WLANs to improve VoIP traffic efficiency.

6.3. Research Limitations

6.3.1. Limitation on Research Scenarios

This research was mainly based on fully distributed IEEE802.11 WLANs. Within this scenario, we assume that the traffic demand and transmission opportunities are evenly distributed over all STAs in the long term average. The only interaction between simultaneous VoIP calls is contention for transmission opportunities in the wireless channel. Each pair of STAs of a VoIP call can in fact be considered as a minimised WLAN in the global topology in a wireless channel.

If the fully distributed topology is considered as one extreme case, the single infrastructure topology that most of the related research focusing on is considered as the other. The realistic cases are in between. For instance, especially in urban area, multiple WLANs may exist nearby and share one wireless channel and each WLAN has several STAs.

6.3.2. Elimination of Issues at the Other Layers

This research focused on the data link layer issues. For the reasons of simplification, the possible impacts on VoIP service quality from other layers have not yet been studied in detail in this research.

At the physical layer, the dynamic of channel conditions may affect the packet error rate (PER) and therefore the number of retransmissions. Retransmission of a voice packet results in further transmission latency and an extra transmission opportunity is consumed. Higher data rates than 54Mb/s are available when IEEE802.11n MIMO antenna is equipped on the latest devices. However the VoIP capacity increase is limited because most of the overheads still exist. Furthermore, the delay accumulation will still happen under congestion.

At the network layer, many published research results e.g. [100-102] indicated that the VoIP capacity is reduced in the orders of the number of average hops within a single interference area. Furthermore, the transmission latency is proportionally increased by the number of hops in the path. As the data link layer focuses on the point-to-point transmissions, we place the wireless nodes within a small area and assume that each pair of wireless nodes can reach each other in one hop. Hence, such multi-hop issues have not been considered.

6.3.3. Co-existing with Other Traffics

The main focuses of this thesis are delay accumulation in outgoing buffer and its solutions. We assumed the pure VoIP traffic is carried over the distributed WLANs in most of our simulations. The co-existence between VoIP and other network traffics has not yet been extensively considered.

Furthermore, IEEE802.11e already has the solutions of prioritising the VoIP traffics in WLANs. In congested conditions, we assumed that IEEE802.11e delays most of the transmission of data traffics and ensure the best possible quality of service to VoIP. In IEEE802.11e, the dedicated outgoing buffer for voice packets is provided above the MAC layer. Therefore, if we consider to deploying ACQ or SPAWN in such buffer, there will be no co-existence of voice and data packets.

6.3.4. Accuracy of Simulation Model and Performance Criterion

The validation of this research was mainly based on the extensive simulations using NS-2 simulator as its simulation results are widely accepted by a large amount of research publications. The implementation of this simulator is based on statistical models. However, the statistical models used, the protocols implemented and the parameter configurations may not fully represent the realistic cases. The simulation figures produced are mainly used as a prediction of performance behaviours under certain conditions. Chapter 3 already validated that the WLANs simulated using NS-2 simulator have the similar behaviours to the numerical analysis and real devices.

As mentioned in Section 3.2, due to the resource limitation this research has not carried out any subjective test by users using real VoIP applications on wireless devices. Instead, we transmit randomly generated data in each voice packet and judge whether the VoIP calls can be maintained by whether the WLANs can provide the lowest performance in order to deliver the minimum acceptable service quality to users. In the real world, one of the actual factors affecting the VoIP service quality is round trip latency. The transmission latencies may not be symmetric; however, our research scenarios focus on point-to-point links in small areas. We assume that all the wireless links are under similar conditions in the long term average. Hence, we treat the links in both direction of each VoIP conversation as having symmetric transmission latency. The minimum acceptable VoIP service quality (90% of voice packets are delivered within 150 ms) used in this research is derived according to the ITU recommendations [17, 37].

6.4. Future Work

There are several possible future work areas in order to improve the research quality and extend this research in further depth.

6.4.1. Implementation of ACQ and SPAWN in Real Devices

So far, we have validated and evaluated the ACQ and SPAWN algorithms under our scenarios and assumptions. As indicated in the conclusions in Chapter 4 and 5, both ACQ and SPAWN can be implemented into existing IEEE802.11 a, b, g devices or test beds. Further validations and evaluations of both algorithms can be performed using real devices in realistic scenarios.

There are also several further research avenues possible on real devices:

- Most of the existing evaluations on VoIP service quality are objectives and with static criterions e.g. latency and packet loss budgets. With the real device implementation, the effects of ACQ and SPAWN algorithms on real VoIP application can be observed directly by users.
- With the implementation on real devices, the proposed algorithms are not isolated from the issues at the other layers. More research can be carried out to analyse the cooperation across layers. Furthermore, it is also interesting to test the cooperation between these two algorithms and the IEEE802.11e EDCA.

6.4.2. Further Analysis on T_{MAX} for ACQ

As indicated in Section 4.5, on the one hand, if T_{MAX} for ACQ is configured greater than the maximum one way delay budget for voice service, the delay accumulation in the outgoing buffer can not be prevented in congested conditions. On the other hand, if T_{MAX} for ACQ is configured smaller than the maximum one way delay budget, more packets than necessary may be selected and dropped. It is worth carrying out further research on the correlation between the T_{MAX} value configuration in ACQ and the resulting VoIP service quality:

1) Research can be carried out to identify what parameters, such as the number of hops in a path, affect the choices of T_{MAX} values. An analytical model could be built using the resulting VoIP service quality, the T_{MAX} and other parameters.
2) The configuration of T_{MAX} value could be enabled to be dynamic over the change of network performance in order to achieve the best possible service quality.

6.4.3. Aggregation to Multiplexing for SPAWN

Current proposed SPAWN algorithm is to aggregate the small packets sent to the same receiver. There are already some proposed algorithms such as in paper [19], for AP in the infrastructure WLANs to aggregate the voice packets in the downlinks and broadcast to different receivers in one larger packet. Such algorithms are called multiplexing.

The existing multiplexing algorithms are focusing on the uplinks/downlinks unbalance problem for the APs. However, more potential can be explored for such algorithms, such as further increasing the transmission efficiency. SPAWN can be easily modified to enable the multiplexing function in the case that a sender needs to transmit packets to more than one receiver at the same time.

References

1. *Skype*. 2012 [cited 2012 29 June]; Available from: http://www.skype.com/intl/en-gb/home.
2. Garg, S. and M. Kappes. *Can I add a VoIP call?* in *Communications, 2003. ICC '03. IEEE International Conference on*. 2003.
3. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.* IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999), 2007: p. 1-1076.
4. Garg, S. and M. Kappes. *An experimental study of throughput for UDP and VoIP traffic in IEEE 802.11b networks*. in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*. 2003.
5. Anjum, F., et al. *Voice performance in WLAN networks - an experimental study*. in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*. 2003.
6. Hole, D.P. and F.A. Tobagi. *Capacity of an IEEE 802.11b wireless LAN supporting VoIP*. in *Communications, 2004 IEEE International Conference on*. 2004.
7. Hegde, N., A. Proutiere, and J. Roberts. *Evaluating the voice capacity of 802.11 WLAN under distributed control*. in *Local and Metropolitan Area Networks, 2005. LANMAN 2005. The 14th IEEE Workshop on*. 2005.
8. Elaoud, M., D. Famolari, and A. Ghosh. *Experimental VoIP capacity measurements for 802.11b WLANs*. in *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*. 2005.
9. Cai, L.X., et al., *Voice capacity analysis of WLAN with unbalanced traffic.* Vehicular Technology, IEEE Transactions on, 2006. **55**(3): p. 752-761.
10. Sangho, S. and H. Schulzrinne. *Experimental Measurement of the Capacity for VoIP Traffic in IEEE 802.11 WLANs*. in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. 2007.
11. Stoeckigt, K.O. and H.L. Vu, *VoIP Capacity - Analysis, Improvements, and Limits in IEEE 802.11 Wireless LAN.* Vehicular Technology, IEEE Transactions on, 2010. **59**(9): p. 4553-4563.
12. Wei, W., L. Soung Chang, and V.O.K. Li, *Solutions to performance problems in VoIP over a 802.11 wireless LAN.* Vehicular Technology, IEEE Transactions on, 2005. **54**(1): p. 366-384.
13. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band.* IEEE Std 802.11-1999, 1999.
14. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access*

Control (MAC) and Physical Layer (PHY) Specifications: Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band. IEEE Std 802.11-2003, 2003.
15. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz Band.* IEEE Std 802.11-1999, 1999.
16. Goode, B., *Voice over Internet protocol (VoIP).* Proceedings of the IEEE, 2002. **90**(9): p. 1495-1517.
17. ITU, *G.114 : One-way transmission time.* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS; International telephone connections and circuits – General; Recommendations on the transmission quality for an entire international telephone connection; One-way transmission time, 2003.
18. Malone, D., P. Clifford, and D.J. Leith, *On buffer sizing for voice in 802.11 WLANs.* Communications Letters, IEEE, 2006. **10**(10): p. 701-703.
19. Dangerfield, I., D. Malone, and D.J. Leith. *Understanding 802.11e Voice Behaviour via Testbed Measurements and Modeling*. in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, 2007. WiOpt 2007. 5th International Symposium on*. 2007.
20. *RFC3261: SIP: Session Initiation Protocol*. 2002 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc3261.
21. ITU, *H.323: Packet-based multimedia communications systems.* SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS; Infrastructure of audiovisual services – Systems and terminal equipment for audiovisual services, 2009.
22. *RFC3550: RTP: A Transport Protocol for Real-Time Applications*. 2003 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc3550.
23. *RFC3605: Real Time Control Protocol (RTCP) attribute in Session Description Protocol (SDP)*. 2003 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc3605.
24. *RFC768: User Datagram Protocol*. 1980 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc768.
25. *RFC793: TRANSMISSION CONTROL PROTOCOL*. 1981 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc793.
26. *RFC791: INTERNET PROTOCOL*. 1981 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc791.
27. Valin, J.-M. *The Speex Codec Manual* 2007 [cited 2012 28 June]; Available from: http://www.speex.org/docs/manual/speex-manual/manual.html.
28. Binod, V., et al., *Investigating voice communication over multipath wireless mobile ad hoc network*, in *Proceedings of the 2nd international conference on Ubiquitous information management and communication*2008, ACM: Suwon, Korea.
29. ITU, *G.729 : Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP).* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS; General Aspects of Digital Transmission Systems, 1996.

30. ITU, *G.711: Pulse code modulation (PCM) of voice frequencies.* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS; General Aspects of Digital Transmission Systems: Terminal Equipments, 1988.
31. ITU, *G.726 : 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS; General Aspects of Digital Transmission Systems: Terminal Equipments, 1990.
32. ITU, *G.723.1 : Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s.* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS; General Aspects of Digital Transmission Systems, 1996.
33. Guangwei, B. and C. Williamson. *Simulation evaluation of wireless Web performance in an IEEE 802.11b classroom area network*. in *Local Computer Networks, 2003. LCN '03. Proceedings. 28th Annual IEEE International Conference on*. 2003.
34. Goodman, D., et al., *Waveform substitution techniques for recovering missing speech segments in packet voice communications.* Acoustics, Speech and Signal Processing, IEEE Transactions on, 1986. **34**(6): p. 1440-1448.
35. Elsabrouty, M., M. Bouchard, and T. Aboulnasr, *Receiver-based packet loss concealment for pulse code modulation (PCM G.711) coder.* Signal Processing, 2004. **84**(3): p. 663-667.
36. Sun, L.F., et al., *Impact of Packet Loss Location on Perceived Speech Quality*, in *2nd IP-Telephony Workshop*2001. p. 114--122.
37. ITU, *G.113: Transmission impairments due to speech processing.* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS International telephone connections and circuits – General Recommendations on the transmission quality for an entire international telephone connection, 1999.
38. Perkins, C., O. Hodson, and V. Hardman, *A survey of packet loss recovery techniques for streaming audio.* Network, IEEE, 1998. **12**(5): p. 40-48.
39. ITU, *P.800 : Methods for subjective determination of transmission quality.* SERIES P: TELEPHONE TRANSMISSION QUALITY; Methods for objective and subjective assessment of quality, 1996.
40. ITU, *P.861 : Objective quality measurement of telephoneband (300 - 3400 Hz) speech codecs.* SERIES P: TELEPHONE TRANSMISSION QUALITY; Methods for objective and subjective assessment of quality, 1996.
41. ITU, *P.861 :Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs* SERIES P: TELEPHONE TRANSMISSION QUALITY; Methods for objective and subjective assessment of quality, 2001.
42. Ding, L. and R.A. Goubran. *Assessment of effects of packet loss on speech quality in VoIP*. in *Haptic, Audio and Visual Environments and Their Applications, 2003. HAVE 2003. Proceedings. The 2nd IEEE Internatioal Workshop on*. 2003.
43. Songun, N. and Y. Seungwha, *Allowable Propagation Delay for VoIP Calls of Acceptable Quality*, in *Proceedings of the First International Workshop on Advanced Internet Services and Applications*2002, Springer-Verlag.

44. Hoene Christian, R.B., Wolisz, Adam *On the importance of a VoIP packet*, in *ISCA Tutorial and Research Workshop on Auditory Quality of Systems; AQS-2003*2003. p. 55-63.
45. ITU, *G.107 : The E-Model: a computational model for use in transmission planning.* SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS; International telephone connections and circuits – General; definitions; The E-Model: a computational model for use in transmission planning, 1998.
46. Clark, A.D., *Modeling the effects of burst packet loss and recency on subjective voice quality*, in *IP Telephony Workshop 2001*2001.
47. Cole, R.G. and J.H. Rosenbluth, *Voice over IP performance monitoring.* SIGCOMM Comput. Commun. Rev., 2001. **31**(2): p. 9-24.
48. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements.* IEEE Std 802.11-2005, 2005: p. 1-1076.
49. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 5: Enhancements for Higher Throughput.* IEEE Std 802.11-2009, 2009.
50. *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 10: Mesh Networking.* IEEE Std 802.11-2005, 2011: p. 1-1076.
51. Stoeckigt, K.O. and H.L. Vu. *VoIP capacity analysis in IEEE 802.11 WLAN*. in *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*. 2009.
52. Obeidat, S.A., A.N. Aldaco, and V.R. Syrotiuk, *Cross-layer opportunistic adaptation for voice over ad hoc networks.* Computer Networks, 2012. **56**(2): p. 762-779.
53. Islam, M.S., et al. *Performance evaluation of various vocoders in mobile ad hoc network (MANET)*. in *Electrical and Computer Engineering (ICECE), 2010 International Conference on*. 2010.
54. *Ubuntu 9.04 Release Notes*. 2009 [cited 2012 28 June]; Available from: https://wiki.ubuntu.com/JauntyJackalope/ReleaseNotes.
55. *JPerf*. 2012 [cited 2012 24 June]; Available from: http://sourceforge.net/projects/jperf/.
56. *ndiswrapper*. 2012 [cited 2012 28 June]; Available from: http://sourceforge.net/projects/ndiswrapper/.
57. *Wireless G USB Adapter*. 2012 [cited 2012 28 June]; Available from: http://www.belkin.com/ProductPDF/documents/F5D7050_179211.pdf.
58. *IPerf*. 2012 [cited 2012 28 June]; Available from: http://sourceforge.net/projects/iperf/.

59. Bianchi, G., *IEEE 802.11-saturation throughput analysis.* Communications Letters, IEEE, 1998. **2**(12): p. 318-320.
60. Bianchi, G., *Performance analysis of the IEEE 802.11 distributed coordination function.* Selected Areas in Communications, IEEE Journal on, 2000. **18**(3): p. 535-547.
61. Yubin, X., et al. *A self-adaptive minimum contention window adjusting backoff algorithm in IEEE 802.11 DCF*. in *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on*. 2012.
62. Baoguo, C. and M. Xuening. *Saturation throughput analysis of IEEE 802.11 DCF under capture effect*. in *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on*. 2012.
63. Nguyen, S.H., H.L. Vu, and L.L.H. Andrew, *Performance Analysis of IEEE 802.11 WLANs With Saturated and Unsaturated Sources.* Vehicular Technology, IEEE Transactions on, 2012. **61**(1): p. 333-345.
64. Tadayon, N., et al., *A Novel Analytical Model for Service Delay in IEEE 802.11 Networks.* Systems Journal, IEEE, 2012. **PP**(99): p. 1-1.
65. *The Network Simulator ns-2*. [cited 2012 28 June]; Available from: http://www.isi.edu/nsnam/ns/.
66. Barman, D., G. Smaragdakis, and I. Matta. *The effect of router buffer size on HighSpeed TCP performance*. in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*. 2004.
67. Avrachenkov, K., U. Ayesta, and A. Piunovskiy. *Optimal choice of the buffer size in the Internet routers*. in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*. 2005.
68. Dhamdhere, A. and C. Dovrolis, *Open issues in router buffer sizing.* SIGCOMM Comput. Commun. Rev., 2006. **36**(1): p. 87-92.
69. Floyd, S. and V. Jacobson, *Random early detection gateways for congestion avoidance.* Networking, IEEE/ACM Transactions on, 1993. **1**(4): p. 397-413.
70. Sally, F., G. Ramakrishna, and S. Scott, *Adaptive RED: An Algorithm for Increasing the Robustness of RED*, 2001.
71. Jinsheng, S., et al., *PD-RED: to improve the performance of RED.* Communications Letters, IEEE, 2003. **7**(8): p. 406-408.
72. Aweya, J., M. Ouellette, and D.Y. Montuno, *A control theoretic approach to active queue management.* Computer Networks, 2001. **36**(2–3): p. 203-235.
73. Clark, D.D. and W. Fang, *Explicit allocation of best-effort packet delivery service.* IEEE/ACM Trans. Netw., 1998. **6**(4): p. 362-373.
74. CISCO. *Weighted Random Early Detection (WRED)*. [cited 2012 18 Apr]; Available from: http://www.cisco.com/en/US/tech/tk543/tk760/tk725/tsd_technology_support_sub-protocol_home.html.
75. Makkar, R., et al. *Empirical study of buffer management scheme for Diffserv assured forwarding PHB*. in *Computer Communications and Networks, 2000. Proceedings. Ninth International Conference on*. 2000.
76. Kunniyur, S.S. and R. Srikant, *An adaptive virtual queue (AVQ) algorithm for active queue management.* Networking, IEEE/ACM Transactions on, 2004. **12**(2): p. 286-299.

77. *RFC3168: The Addition of Explicit Congestion Notification (ECN) to IP*. 2001 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc3168.
78. Wydrowski, B. and M. Zukerman. *GREEN: an active queue management algorithm for a self managed Internet*. in *Communications, 2002. ICC 2002. IEEE International Conference on*. 2002.
79. Nagle, J. *RFC896 Congestion control in IP/TCP internetworks* 1984 2012]; 28 June]. Available from: http://www.dspace.cam.ac.uk/handle/1810/13500.
80. Stewart, R. and Q. Xie, *Stream Control Transmission Protocol (SCTP)*2001: {Addison-Wesley Professional}.
81. Ong, L., C. Corporation, and J. Yoakum. *RFC3286 An Introduction to the Stream Control Transmission Protocol (SCTP)*. 2002 [cited 2012 28 June]; Available from: http://grenache.tools.ietf.org/html/rfc3286.
82. Kim, K., et al., *On Packet Aggregation Mechanisms for Improving VoIP Quality in Mesh Networks*, in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*2006. p. 891 -895.
83. Ramya, R., et al., *IPAC: IP-based adaptive packet concatenation for multihop wireless networks*, 2006.
84. Kliazovich, D. and F. Granelli, *Packet concatenation at the IP level for performance enhancement in wireless local area networks.* Wireless Networks, 2008. **14**(4): p. 519-529.
85. Ka, Y., *802.11a Modeling and MAC enhancements for high speed rate adaptive networks*, 2002.
86. Hongqiang, Z. and F. Yuguang. *A distributed packet concatenation scheme for sensor and ad hoc networks*. in *Military Communications Conference, 2005. MILCOM 2005. IEEE*. 2005.
87. Chun-cheng, C., et al. *Rate-Adaptive Framing for Interfered Wireless Networks*. in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. 2007.
88. Ci, S. and H. Sharif. *Adaptive optimal frame length predictor for IEEE 802.11 wireless LAN*. in *6th International Symposium on Digital Signal Processing for Communication Systems*. 2002.
89. Yuxia, L. and V.W.S. Wong. *WSN01-1: Frame Aggregation and Optimal Frame Size Adaptation for IEEE 802.11n WLANs*. in *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*. 2006.
90. Kai-Ten, F. and L. Po-Tai. *Frame-aggregated link adaptation algorithm for IEEE 802.11n networks*. in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*. 2009.
91. Feng, K.-T., P.-T. Lin, and W.-J. Liu, *Frame-aggregated link adaptation protocol for next generation wireless local area networks.* EURASIP J. Wirel. Commun. Netw., 2010. **2010**: p. 1-13.
92. Castro, M.C., et al. *Capacity Increase for Voice over IP Traffic through Packet Aggregation in Wireless Multihop Mesh Networks*. in *Future Generation Communication and Networking (FGCN 2007)*. 2007.
93. Karlsson, J., A. Kassler, and A. Brunstrom. *Impact of packet aggregation on TCP performance in Wireless Mesh Networks*. in *World of Wireless, Mobile and Multimedia Networks & Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*. 2009.

94. Jeong, Y., et al., *VoIP over Wi-Fi Networks: Performance Analysis and Acceleration Algorithms.* Mobile Networks and Applications, 2009. **14**(4): p. 523-538.
95. Dely, P., et al., *FUZPAG: A fuzzy-controlled packet aggregation scheme for wireless mesh networks*, 2010. p. 778-782.
96. Frantti, T. and M. Koivula, *Fuzzy packet size control for delay sensitive traffic in ad hoc networks.* Expert Systems with Applications, 2011. **38**(8): p. 10188-10198.
97. Seungwoo, J. and L. Jihoon, *Adaptive frame aggregation scheme for energy efficiency in WLAN*, 2011. p. 463-464.
98. Skordoulis, D., et al., *IEEE 802.11n MAC frame aggregation mechanisms for next-generation high-throughput WLANs.* Wireless Communications, IEEE, 2008. **15**(1): p. 40-47.
99. *RFC2460: Internet Protocol, Version 6 (IPv6) Specification*. 1998 [cited 2012 28 June]; Available from: http://tools.ietf.org/html/rfc2460.
100. Kekre, H.B., et al., *Performance problems of VoIP in 802.11 wireless mesh networks & their solutions*, in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*2011, ACM: Mumbai, Maharashtra, India. p. 891-898.
101. Niculescu, D., et al. *Performance of VoIP in a 802.11 Wireless Mesh Network*. in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*. 2006.
102. Kashyap, A., et al. *VoIP on Wireless Meshes: Models, Algorithms and Evaluation*. in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. 2007.