Modeling and Enhancement for Voice over IP-Networks

Modeling and Enhancement for Voice over IP-Networks

by

Ahmed Mahmoud Mahmoud Mohamed Shawish

submitted to
Tohoku University
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Supervisor: Professor Susumu Horiguchi

Graduate School of Information Sciences
Tohoku University
Japan

March, 2009

Abstract

There are about 1 billion fixed telephone lines and 2 billion cell phones in the world that use the traditional public switched telephone network (PSTN) systems. Soon, they will move to networks based on open protocols- known as Voice over Internet Protocols (VoIP). This migration is fueled by many factors, like the tremendous growth of the Internet and the World Wide Web, the rapid and low cost coverage capability through wired and wireless networks, the availability of a variety of fixed and mobile Internet accessing tools (e.g., desktop, laptop, pocket pc, dual-band cellular, etc), and the feasibility of integrating voice and data into a single infrastructure. The current IP networks are designed based on an open architecture to mainly support best effort applications, like file transfer, web browsing, and email, which are not delay or delay-jitter sensitive. Due to the current very low voice traffic volume, the available IP networks are still able to report an acceptable quality of service (QoS) for the VoIP applications. Nevertheless, both wired and wireless IP-networks are still not ready to absorb the future expected huge volume of immigrated voice traffic because of many problems. In this dissertation we addressed a group of these problems as follows:

First, the current packet schedulers in wired IP networks meet the delay constraint of VoIP traffic by simply assigning its packets with the highest priority. This treatment is acceptable as long as the amount of VoIP traffic is relatively very small compared to other non-voice traffic. With the notable expansion of VoIP applications, however, the current packet schedulers will significantly sacrifice the fairness deserved by the non-voice traffic. In this thesis, we extend the conventional Deficit Round-Robin (DRR) scheduler by including a packet classifier, a Token Bucket and a resource reservation scheme and propose an integrated packet scheduler architecture for the growing VoIP traffic. We demonstrate through both theoretical analysis and extensive experimental simulation that the new architecture makes it possible for us to significantly improve the fairness deserved by the non-voice traffic while still meeting the tight delay requirement of VoIP applications.

Second, efficient VoIP support at the wireless access point of a Wireless LAN (WLAN) remains a challenge for the last-mile wireless coverage of IP networks with mobility support. Due to the limited bandwidth available in WLANs, an accurate analysis of voice capacity in such networks is crucial for the efficient utilization of their resources. The available analytical models only provide the upper and lower bounds on voice capacity, which may significantly overestimate or underestimate the WLAN's capability of supporting VoIP and thus are not suitable for above purpose. In this thesis, we focus on the voice capacity analysis of a wireless 802.11(a/b) access point running the distributed

coordination function (DCF). In particular, we show that by incorporating the clients' spatial distribution into analysis, we are able to develop a new analytical model for a much more accurate estimation of average voice capacity. By properly exploring this spatial information, we further propose a new scheme for access point placement such that the overall voice capacity can be enhanced. The efficiency of the new voice capacity model and new access point placement scheme is validated through both analytical and simulation studies.

Third, the available bidirectional transmission (BDT) protocol promises to eliminate the VoIP traffic downlink bottleneck at the WLAN's AP by allowing the each transmission session to afford two packets delivery, from sender to receiver and then from receiver to sender if it has a packet to send at such instance. Nevertheless, the current first-in-first-out scheduler adopted in the WLANs' AP does not support the BDT protocol to fulfill its promises. This is because, in the AP, the scheduling of the packet does not consider whether the receivers have also a packet to send-back at such instance or not. In this thesis, we enhance the BDT protocol through a novel packet scheduler and a new contention mechanism. In the proposed scheduler, we used a probabilistic-based scheme to help the AP to schedule its packets according to the higher BDT chance first out rule, while the new contention mechanism is proposed to provide the highly expected BDT sessions with the highest channel access priority. We demonstrate through both analytical analysis and computer simulations that the proposed modifications can significantly enhance the performance of the BDT protocol in terms of throughput and voice capacity. We also show that the enhanced BDT protocol become able to provide a better support to the VoIP applications in over WLANs even under the coexistence of Best-effort traffic.

Fourth, the current widely used IEEE 802.11 distributed coordination function (DCF) protocol is not suitable for efficient support of VoIP, because of both its downlink bottleneck and the large packet overhead caused by the ACK mechanism. Based on the observations that voice packets are of small size, bi-directional and can tolerant certain level of packet loss, we propose a novel MAC protocol to provide an efficient support to voice applications over WLANs. The main idea of the new MAC is to remove the downlink bottleneck by adopting a two-way instead of current one-way transmission mechanism, and then enhancing the channel utilization efficiency by compacting the packets exchanging processes into a fewer number of steps. We first design a protocol that efficiently work in a WLAN that does not suffer from non-ideal channel conditions then we further extend it to suit any condition. We demonstrate through both analytical analysis and computer simulations that the new MAC protocol can dramatically increase the voice capacity and throughput even under the coexistence of best effort traffic.

To My Family

Acknowledgments

First, I would like to sincerely thank ALLAH the almighty who blessed me with all goods along my life and however I thank him, my thanking will not be enough.

I wish to thank my beloved country "Egypt" for taking over the expenses of my study abroad in this great country "Japan" in order to prepare a new generation of skillful researchers in all the branches of science.

I am deeply indebted, grateful to my Supervisor, Prof. Susumu Horiguchi for his continuous guidance, supervision, and warm support which helped me to stay on the right track from the first day of my PhD. His kindness, prudence and work of ethics have made my PhD period one of the best periods of my life.

This dissertation would not be possible without the help of Assoc. Prof. Dr. Xiaohong Jiang. I own him infinite thanks as he tolerated me for the whole PhD period with his careful guiding and teaching, hand by hand and paper by paper. I will never forget his valuable suggestions, discussions and encouragement for the rest of my life.

I would like to give a special thanks to Dr. Masaru Fukushi for his continuous and friendly support. He was always there to help not only me but also all of my lab mates.

I am also thankful to my tutor Fukushima who helped me to stabilize my life at my arrival to Japan. I would like to also thank my colleague Chi for his valuable discussions with me during my research, and my entire lab mates, friends and fellows here in Japan who always helped me to their best and made my life easier.

Acknowledges are also given to my professors in my mother faculty in Egypt "Faculty of Computer and Information Sciences, Ain Shams University", for their strong support and continuous encouragement.

Finally, I would like to express my sincere appreciation to my beloved wife and kids for their complete support and patient with me to continue this research. Without their encouragement and support, my life would have been very hard. In addition to that, I wish to thank my dear mother who was always praying for me and my deceased dear father who was always encouraging me to study and work hard.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

There are about 1 billion fixed telephone lines and 2 billion cell phones in the world that use the traditional public switched telephone network (PSTN) systems. Soon, they will move to networks based on open protocols- known as Voice over Internet Protocols (VoIP) [2, 3]. This migration is fueled by many factors, like the tremendous growth of the Internet and the World Wide Web, the rapid and low cost coverage capability through wired and wireless networks, the availability of a variety of fixed and mobile Internet accessing tools (e.g., desktop, laptop, pocket pc, dual-band cellular, etc), and the feasibility of integrating voice and data into a single infrastructure [4, 5, 6].

The current IP networks are designed based on an open architecture to mainly support best effort applications [7], like file transfer, web browsing, and email, which are not delay or delay-jitter sensitive. Due to the very low relative volume of voice traffic, current IP networks are still able to report an acceptable quality of service (QoS) for the VoIP applications [8, 9]. Nevertheless, both wired and wireless IP-networks are still not ready to absorb the future expected enormous volume of immigrated voice traffic because of many reasons like: the unfair packet scheduling currently employed in wired network, the inaccurate voice capacity estimation model and ineffectual AP placement scheme currently utilized in the wireless networks, in addition to the inefficient (excessively overloaded) medium access control protocol presently operated over the later type of networks.

1.2 Problems

Although there are extensive researches in the literature on the voice over the IP-networks, the available models and approaches are mainly based on the assumption that the voice applications and traffic volume are only a fraction of the current overall IP-networks' traffic, which is nearly going to be unrealistic. Therefore a lot of work should be addressed to provide accurate mathematical models and more efficient approaches so solve the problems addressed in this dissertation.

First, in the wired IP networks, the available packet scheduling schemes (like Deficit Round Robin [10] and Weighted Fair Queueing [11]) are mainly designed to provide a fair bandwidth sharing among network traffic without a deliberate consideration about their delay performance. To guarantee the tight delay and delay jitter requirements of VoIP packets, the current packet schedulers simply assign them with the highest priority [12]. This simple priority policy is acceptable as long as the amount of voice traffic is relatively very small in comparison with other non-voice traffic. With the notable expansion of VoIP applications and thus rapid growth of VoIP traffic, however, the above simple priority policy for VoIP will significantly sacrifice the fairness deserved by the non-voice traffic while adopting the fair schedulers alone, as an alternative solution, will severely degrade the QoS of the VoIP applications. Therefore, new schedulers should be developed for future IP networks to efficiently handle the impacts that will arise with the expected growth of VoIP traffic and other delay-sensitive traffic in general.

Second, in the wireless local area networks (WLANs) case, when designing a VoIP system, the most important parameter of concern is the voice capacity of the wireless access point (AP) [13, 14], which is defined as the number of voice connections that can be simultaneously supported through AP. Since the current WLANs have very limited bandwidth and the voice admission control there mainly depends on this parameter to accept or reject new voice calls [15], so a careful voice capacity analysis is crucial for the efficient utilization of WLANs resources. The available models for voice capacity analysis only provide the upper and lower bounds on voice capacity [16, 17, 18], because they were developed based on the assumption that the transmission rate (R) between any client and AP is always either the maximum or the minimum achievable rate. In practice, however, R varies depending on the access distance, shadowing effect and channel fading along the signal path [19, 20]. Thus, the maximum R is only available for those clients who are very close to the AP, while it is sharply stepped down (non-linearly) as the access distance increases. Therefore, the available simple models may significantly overestimate or underestimate the voice capacity and hence severely degrade the QoS of such application.

Third, at the time of building a new WLAN, the network designers usually select the geometric center of the considered area to place the WLAN's access point. Notice that in practice the clients in an area may be non-uniformly distributed, so its geometric center may be far from the high clients' density area. Therefore, placing the AP at the geometric center of such area may significantly degrade the average transmission rate between the mobile clients and the fixed AP, which results in sever degradation of the WLAN's voice capacity and hence waste of its limited resources. Based on the above observations, a new AP placement scheme, based on a careful consideration of the clients' spatial distribution, should be introduced.

Fourth, in addition to the problems of voice capacity estimation and AP placement, the WLAN also suffer from its medium access control (MAC) protocol which was not mainly designed to support the delay sensitive applications like VoIP. The basic access method in the IEEE 802.11 MAC is the distributed co-ordination function (DCF) [21]. In such protocol, the channel access involves a long unbiased competition among all the active stations[1] including the WLAN's AP that may be even ended with a collision. However, once a successful transmission session is initiated, although it is established between two nodes, the sender node is the only one who is allowed to transmit its packet while the receiver node can only acknowledge the reception of such packet. Based on the fact that the VoIP traffic is divided half-by-half upon the downlink (AP-stations) and uplink (stations-AP), and because of such unbiased channel access opportunity, the current DCF protocol usually cause a VoIP traffic downlink bottleneck at the AP and hence restrain the VoIP applications over the WLANs. In general, this protocol can be efficient for best effort applications with big size packets and low sampling rate[2], but for a delay sensitive traffic like VoIP with very small packet size and very fast sampling rate (e.g., 50 packets per second for a payload of 20ms per packet) , is ultimately a bottle-neck.

1.3 Objectives

The overall aim of the thesis is to provide an efficient support to the voice applications over the wired and wireless IP-networks. Our research mainly focuses on the above four main problems which significantly affect the performance of the VoIP. We will provide more accurate models and mathematical analysis for those problems. We will also propose new approaches to efficient enhance the VoIP performance. We will investigate and compare the performance of the new approaches with that of other conventional algorithms.

First, we will introduce a new packet scheduling architecture that has the capabilities to both meet the tight delay requirement of VoIP applications and also avoid the aggressive resource unfairness to other non-voice traffic. This new scheduling architecture will allow a graceful trade-off to be initialized between priority and fairness in the future VoIP-capable IP network.

Second, we will develop a new analytical model for a much more accurate estimation of the average transmission rate between the access point and the mobile clients. Such new model will be conducted based on a careful consideration of the clients' spatial distribution

[1]Active station means that it has a packet to send at that instance
[2]the number of generated packets per second

inside the WLAN. We will investigate the efficiency of the new model in terms of the voice capacity estimation under different simulation scenarios.

Third, by properly exploring the clients' spatial distribution information, we will further propose a new scheme for access point placement such that the overall voice capacity can be relatively enhanced. We will demonstrate through our scheme the importance of considering the clients' spatial distribution when placing the AP of any WLAN.

Fourth, we will enhance the performance of one of the available unused MAC protocol (so called the bidirectional-transmission protocol). Such protocol promise to solve the downlink unfairness problem of the voice traffic in the WLANs but without any increase in the voice capacity. A novel packet scheduler will be proposed to overcome its current drawbacks. We will further enhance its performance even under the coexistence of best-effort traffic condition using a new proposed contention mechanism.

Fifth, addressing the same MAC protocol problem, we will develop a new protocol to dramatically increase the voice capacity in the current IEEE 802.11 WLANs under ideal wireless channel condition. We will further extend it to suit the WLANs that suffer from non-ideal channel condition. We will demonstrate through mathematical model and simulation studies that the improvement in the voice capacity can be still achievable yet under the coexistence of best-effort traffic.

1.4 Organization of This Thesis

The following of the thesis is organized as follows:

Chapter 2. Fair Scheduler for Voice Traffic over Wired IP-Networks. In this chapter, we focus on the voice traffic scheduling over the wired IP-networks. We introduce a new packet scheduling architecture that has the capabilities to both meet the tight delay requirement of VoIP applications while allowing a fair resources sharing to other non-voice traffic. We also discuss the related works and their disadvantages. The out performance of our proposed scheduling architecture in comparison with the available ones is demonstrated through both mathematical analysis and extensive simulation studies. The end-to-end quality of service provided by the proposed scheduler is even examined in order to prove its applicability.

Chapter 3. Voice Capacity Analysis and Enhancement in WLAN. In this chapter, we focus on the voice capacity estimation and enhancement over the wireless IP-networks. With a careful consideration of clients spatial distribution (CSD) information, we develop a new analytical model for a much more accurate estimation of the average WLANs' voice capacity instead of the current upper and lower bound estimation models. Furthermore, by properly exploring such CSD information, we further propose a new scheme for access point placement such that the overall voice capacity can be enhanced without any extra resources. The mathematical analysis and extensive simulation studies conducted in this chapter demonstrate that our new model and AP placement scheme have accurately estimate a significant enhanced the achievable voice capacity in WLANs, respectively.

Chapter 4. An Enhanced bidirectional-transmission Protocol. In this chapter, we focus on enhancing the available MAC protocols of the wireless IP-networks in order to provide a more efficient support to the VOIP. We enhance the available bidirectional-transmission MAC protocol by introducing a novel packet scheduler and a new contention mechanism. It worth to mention here that such protocol promised to solve the downlink unfairness problem of the voice traffic in the WLANs but without any increase in the WLANs' voice capacity. Therefore, the our modifications were proposed to get the full advantages of such protocol and overcome its drawbacks. The developed analytical model and conducted simulation studies show that expected gains in terms of throughput and voice capacity are quite remarkable even under the coexistence of best-effort traffic thanks to the proposed modifications.

Chapter 5. A Novel MAC Protocol for VoIP support over WLANS. In this chapter, In this chapter, we again focus the wireless IP-networks' MAC protocol. We propose a novel MAC protocol to dramatically increase the voice capacity in the current IEEE 802.11 WLANs with ideal wireless channel condition, so called DCFvs (DCF voice support). Furthermore, we extend the DCFvs protocol to suit the WLANs that suffer from non-ideal channel conditions. We develop analytical models for estimating both of the throughput and voice capacity of the proposed protocols. The conducted mathematical analysis and extensive simulation studies demonstrate the significant improvement in the WLANs' performance in terms of throughput and voice capacity even under the coexistence of best-effort traffic.

In the last chapter, we conclude the overall thesis and discuss the future works.

1.5 Main Contributions

The general contribution of our work is briefly summarized in providing an accurate modeling and efficient enhancement of the voice over both wired and wireless IP-networks. We addressed the unfair packet scheduling currently employed in wired network, the inaccurate voice capacity estimation models and ineffectual AP placement scheme currently utilized in the wireless networks, in addition to the inefficient (excessively overloaded) medium access control protocol presently operated over the later type of networks.

The details of our contributions are as follows:

1. We propose a new packet scheduling architecture that has the capabilities to both meet the tight delay requirement of VoIP applications with nearly the same delay

performance achieved by the current pure Strict-Priority scheduler while allowing a fair bandwidth sharing to other non-voice traffic. The conducted mathematical analysis and extensive simulation studies demonstrate the a graceful and easily controllable trade-off can be initialized between priority and fairness in the future VoIP-capable IP networks.

2. With the consideration of clients spatial distribution (CSD), we develop a new analytical model for a much more accurate estimation of WLANs' average voice capacity instead of the current upper and lower bound estimation models. By properly exploring such CSD information, we further propose a new scheme for access point placement so that the overall voice capacity can be enhanced. We demonstrate through our new model and scheme that the CSD has a significant implication on the achievable voice capacity and thus it should be carefully considered in the AP placement.

3. We enhance the available Bidirectional Transmission-based MAC protocol by proposing a novel packet scheduler and new contention mechanism. The proposed modifications proved to get the full advantages of such protocol while overcoming its drawbacks. We develop analytical models to estimate the maximum gains that we can obtain from the enhanced protocol in terms of throughput and voice capacity. We demonstrate that although the available protocol can only improve the impact of the voice traffic' s downlink problem, this improvement can be much more significant if our proposed modifications are adopted even under the coexistence of best-effort traffic.

4. We propose a new medium access control (MAC) protocol to dramatically increase the voice capacity in the current IEEE 802.11 WLANs. We firstly develop a protocol that suit those WLANs with ideal wireless channel conditions. We further extend it to suit the WLANs that suffer from non-ideal channel conditions. We develop analytical models to estimate both throughput and voice capacity of the proposed protocols. We demonstrate through extensive comparative simulations studies with the available protocols that the proposed ones have significant improved the performance of the wireless IP-networks in terms of throughput and voice capacity even under the coexistence of best-effort traffic.

Chapter 2

Fair Scheduler for Voice Traffic over Wired IP-Networks

2.1 Introduction

With the wide expansion of voice services over the IP networks (VoIP), the volume of this delay sensitive traffic is steadily growing up. The current packet schedulers for IP networks meet the delay constraint of VoIP traffic by simply assigning its packets with the highest priority. This treatment is acceptable as long as the amount of VoIP traffic is relatively very small compared to other non-voice traffic. With the notable expansion of VoIP applications, however, the current packet schedulers will significantly sacrifice the fairness deserved by the non-voice traffic. In this chapter, we extend the conventional Deficit Round-Robin (DRR) scheduler by including a packet classifier, a Token Bucket and a resource reservation scheme and propose an integrated packet scheduler architecture for the growing VoIP traffic. We demonstrate through both theoretical analysis and extensive experimental simulation that the new architecture makes it possible for us to significantly improve the fairness to non-voice traffic while still meeting the tight delay requirement of VoIP applications.

2.2 Voice Traffic Over Wired IP-networks

In the VoIP applications, voice packets are transmitted through IP networks, incurring transmission, queueing and propagation delay at each hop along their paths. Since both delay and delay variation (jitter) may significantly affect the quality of voice services [2, 10, 11], therefore the end-to-end delay and jitter control need to be carefully addressed in the VoIP applications. In general, the end-to-end delay can be regarded as the sum of a constant delay component and a random one. The constant delay component is mainly determined by the signal transmission and propagation time along physical links, so it is in general uncontrollable. The random delay component, on the other hand, is the queueing delay encountered by a voice packet in each IP router along its path. Since the queueing delay of a packet is mainly controlled by the packets scheduling scheme adopted inside the IP routers, so the packet scheduling schemes should be deliberately designed to support the VoIP applications.

The current IP networks are developed mainly for supporting non-real time best effort applications, like file transfer and email, which are not delay or delay jitter sensitive. Thus, the available packet scheduling schemes (e.g. Deficit Round Robin [10] and Weighted Fair Queuing [11]) are designed to provide a fair bandwidth sharing among network traffic without a deliberate consideration about their delay performance. It is notable that current VoIP traffic, although they are delay and delay jitter sensitive, constitutes only a very small fraction of overall traffic in the current IP networks. Therefore, the current packet schedulers guarantee the tight delay and delay jitter requirements of VoIP packets by simply assigning them with the highest priority [12, 7]. This simple priority policy is acceptable as long as the amount of voice traffic is relatively very small in comparison with other non-voice traffic. With the notable expansion of VoIP applications and thus a rapid growth of VoIP traffic, however, the above simple priority policy for VoIP will significantly sacrifice the fairness deserved by the non-voice traffic while adopting the fair schedulers alone, as an alternative solution, will severely degrade the quality of VoIP applications. Therefore, new schedulers should be developed for future IP networks to efficiently handle the impacts that will arise with the expected growth of VoIP traffic and the delay-sensitive traffic in general.

2.3 Related Work

Numerous scheduling algorithms (schedulers) have been proposed to guarantee either the minimum queueing delay or a fair resource sharing for the ongoing traffic in the IP network, see for example [10, 11, 22, 23, 24, 25, 26]. Currently, voice traffic constitutes only a very small portion of total traffic in IP networks, so the simple Strict Priority (SP) scheduling scheme is usually adopted as a low-cost solution to support the delay sensitive voice traffic [12, 7].

In the SP scheme, high priority queue is served first until it is empty, and then the packets in the low priority queue are served. This simple priority policy can easily fulfill the QoS requirement of the delay-sensitive traffic. Unfortunately, the main disadvantage of SP scheme is that the low priority queues may suffer from bandwidth starvation if the higher priority queues saturate the link bandwidth [12], so it may introduce a significant unfairness problem to other non delay-sensitive traffic, in particular with the notable expansion of VoIP applications as expected in the future IP networks. Other priority-based scheme, like the low latency queueing scheduler (LLQ) [27], currently implemented by Cisco in their routers, is also promising to guarantee the delay of real-time services. The LLQ is actually a combination of SP scheme and Class-Based Weighted-Fair Queueing (CBWFQ) [28], and it is currently the recommended queueing function for VoIP applications, and it will also work well with video conferences. However, the LLQ also share the same main drawbacks of priority-based schemes, i.e., they can guarantee this promising performance only when the volume of delay-sensitive traffic is very small otherwise a significant bandwidth starvation (unfairness) problem will occur to other low priority traffic.

The fairness has been the main target in case of scheduling non delay-sensitive traffic (known as the best effort traffic) and many schedulers have been proposed for the fairness purpose with different degree of complexity. The Fair Queueing (FQ) scheme [11] can achieve almost the perfect fairness with a time complexity $O(logN)$, where N is the number of packet streams that are concurrently active at the gateway or router. The Deficit Round-Robin (DRR) scheduler [10] can provide almost the same fairness like FQ but with only a constant complexity $O(1)$. In DRR scheme, we assign a given quantum to each queue depending on its rate and also use a deficit counter to recode the deficit of this queue from previous rounds, and we serve each nonempty queue in a round robin way based both on the quantum assigned to the queue and also the deficit available to this queue. Many variations of DRR have also been proposed to further improve its fairness without sacrificing its attractive constant complexity, such as the nested deficit round-robin scheduler (NDRR) [29] and the dynamic deficit round-robin scheduler (DDRR) [30]. As a matter of fact, the DRR scheduler and its variations all work based on the round robin mechanism, which mainly focuses on the fair resource sharing among all flows regardless whether they are delay sensitive or best effort traffic flows. Therefore, they were unable to efficiently fulfill the tight delay constraint of the delay sensitive VoIP applications, especially with the expected increases of its traffic volume.

It is worth noticing that another class of weight-based schedulers has also been developed to provide fairness according to the weight of the traffic class, like the weighted fair queueing scheduler (WFQ) [11], the worst-case fair weighted fair queueing scheduler (WF2Q) [31] and the delay optimized worst case fair WFQ scheduler (DO-WF2Q) [32]. The main idea of this type of schedulers is to assign a weight to each flow depending on its traffic volume and then achieve the fairness by emulating the General Processor Sharing (GPS) discipline. Although the weight-based schedulers can achieve a relatively better fairness than the DRR, it suffers not only from a high complexity of $O(N)$, which is significantly higher than the constant time complexity provided by DRR, but also can not meet the tight delay requirements of the VoIP applications.

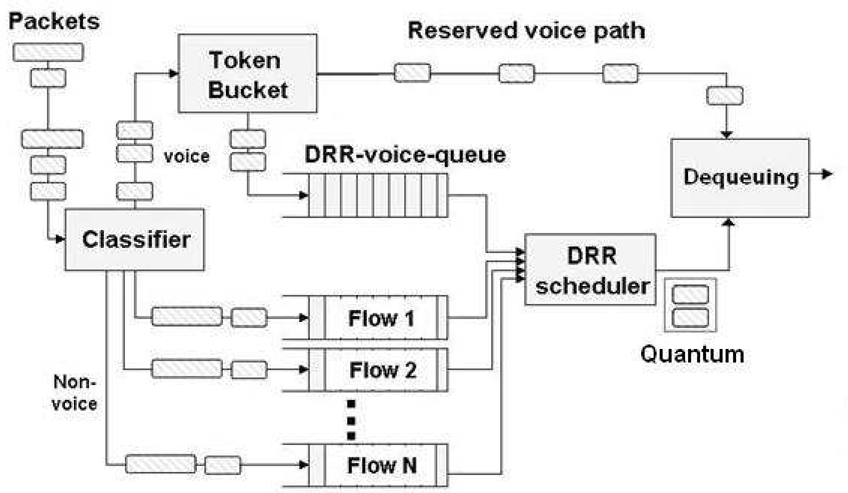


Figure 2.1: The proposed packet scheduler architecture

With the expected expansion of the voice traffic, neither the available SP-Based schemes nor the current fair schedulers will be able to efficiently support both voice and non-voice traffic simultaneously in a way that can easily meet the tight delay requirements of the VoIP applications and also provide fair resource sharing to other low priority traffic. Based on the above observations, we propose in next section a new scheduler for the growing VoIP traffic to achieve a graceful trade-off between delay and fairness in future VoIP-capable networks.

2.4 A New Scheduler For VoIP

The main idea of our new scheduler is to extend a fair scheduler (the DRR scheduler) by combining it with a resource reservation scheme, a traffic shaper (Token Bucket) and a packet classifier into an integrated architecture. The main objective of the proposed scheduler is to guarantee fairness to the non-voice traffic without significantly sacrificing the delay performance of delay sensitive voice traffic.

2.4.1 Scheduler Architecture

The basic architecture of our new scheduler is illustrated in Figure 4.1, where the three main modules (i.e., packet classifier, Token Bucket and DRR) are shown.
Packet Classifier: The packet classifier module in our scheduler is used to classify the incoming packets into voice and non-voice classes. The packet classification function can be easily implemented based on the idea of packet header inspection [33].
Token Bucket: It works as a traffic shaper, where input traffic is divided into two output traffic. the first one is a traffic with an upper bound rate that never exceeds

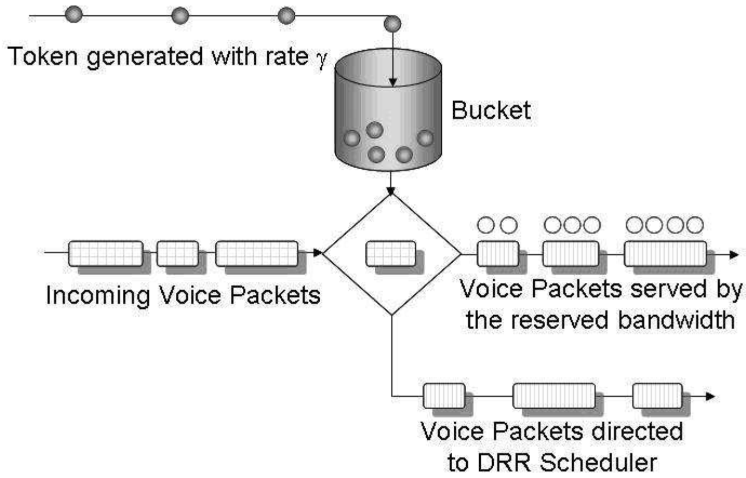


Figure 2.2: The operation of Token Bucket

the token bucket rate R_{token} regardless of it incoming input rate. the second one is the overflow traffic which will be originated due to this limitation. In our scheduler, the main task of this module is to split the incoming voice flows into two parts, as illustrated in Figure 2.2. The first one is a smooth voice traffic composed of a group of voice flows with aggregated rate less than or equal to the R_{token}, and which will be served through a reserved bandwidth equal to this rate. We will refer to this traffic as the reserved voice (RV) traffic. on the other hand, the second voice traffic is composed of the voice flows overflowed from the Token Bucket due to the upper bound rate limitation and they will be re-directed to the DRR module. We will refer to this traffic as the DRR voice (DV) traffic. Within the DRR module, all the voice flows that belong to the DV traffic will compete fairly with other non-voice traffic on the remaining unreserved bandwidth (i.e., the remaining amount of output bandwidth which is not reserved by RV group.)
DRR Scheduler. The DRR module in our scheduler is used to guarantee a max-min fair resource utilization of the remaining unreserved bandwidth among the non-voice traffic and the DV traffic. Actually, the fair resource sharing in our scheduler can also be achieved by adopting other fair schedulers. We choose the DRR scheduler due to its simplicity and its attractive constant time complexity, such that a good scalability can be guaranteed in our proposed scheduler.

2.4.2 Scheduler Control

Upon the arrival of packets, they will be firstly classified into non-voice class and voice class. Those packets that are classified as non-voice will be directly served through the DRR module. On the other hand, the packets that belong to voice class will be forwarded to the Token Bucket module for further classification, partial of them will be served through the reserved bandwidth with a rate limit R_{token} while the remaining packets will be re-directed to the DRR module.

The main task of Token Bucket module is to allow only to a specific amount of voice traffic with an upper bound rate denoted by R_{token} to be served through the reserved bandwidth [34], and to redirect the rest overflow of voice packets to the DRR module. The Token Bucket check the voice packet size and compares it with the tokens accumulated by now (i.e., the arrival time of the voice packet) in the bucket, where token is a unit of byte generated with constant rate (i.e., R_{token}) and accumulated in a buffer called Bucket. If this packet size is less than the tokens accumulation, then an amount of tokens that equals to this packet size will be deleted from the bucket and this packet will be served directly through the reserved bandwidth; otherwise, this packet will be re-directed to the DRR module.

The voice flows that will be directed to the DRR module will be buffered in a dedicated queue called the DRR-voice-queue, in which it will be treated a single competitor flow. In DRR module, each queue i is assigned with a quantum Q_i and is associated with a deficit counter DC_i. The quantum Q_i represents the worth of bits that queue i can send in each round, while the deficit counter DC_i is used to recode the deficit of this queue from its previous rounds. The DRR module can handle variable packet sizes in a fair manner by serving all non-empty queues in turn (round-robin mechanism). Once the queue i got its turn, it begins to send out packets subject to the constraint that the summation of their size is less than or equal to Q_i. If there are no more packets in the queue i after the queue has been serviced, then the queue state variable DC_i is reset to zero. Otherwise, the deficit amount of the Q_i is stored in the state variable DC_i. In subsequent round, the amount of usable bandwidth will be the sum of DC_i (from previous rounds) and Q_i. It is notable that our new scheduler architecture is flexible, since it covers pure DRR and Strict-Priority scheduler as two special cases.

2.4.3 Comparison With Relevant Schedulers

The current packet schedulers like Class-Based Weighted-Fair Queueing (CBWFQ), low latency Queueing scheduler (LLQ) and Strict Priority (SP)scheduler guarantee the tight delay and delay jitter requirements of VoIP packets by simply assigning them with the highest priority where the voice packets are served immediately upon their arrival without imposing any limitation on their maximum rate. This policy is acceptable as long as the voice traffic represents a very small portion of the ongoing traffic across the IP networks. With the notable expansion of VoIP applications as expected in the future IP networks, however, this priority policy will introduce a significant unfairness problem to other low priority traffic since it will saturate the link bandwidth while adopting the fair schedulers alone like Deficit Round-Robin (DRR) or Weighted Fair Queueing (WFQ), as an alternative solution, will severely degrade the quality of VoIP applications.

In our scheduler, we adopt a new splitting mechanism to the VoIP traffic and bandwidth, in which only a specified portion of the voice traffic will be served by an equivalent amount of reserved bandwidth. In this way, we can control not only the amount of resources occupied by the voice traffic but also initiate a fair competition between the remaining portion of the voice traffic and the other non-voice traffic. Actually, the voice traffic will endure an additional amount of delay in our scheduler due to the fair competition with other traffic. However, we showed in this chapter that by properly controlling such splitting mechanism, a nice trade-off between voice packet delay and fairness deserved by other traffics can be achieved. It is notable that such controllable and flexible trade-off is not achievable from using the probabilistic Priority scheme or even the max-min fair schedulers.

In the proposed scheduler, the famous DRR scheduler was proposed to guarantee a fair resources sharing among the competitor traffics. For a while, the WFQ is suppose to be a better choice for this task as it provide a perfect fairness, which is even better than the DRR. Nevertheless, the use of the WFQ will enforce the complexity of the overall proposed scheduling architecture to increase by at least $O(logn)$ (where n is the number of flows), and hence affect its scalability and practicability. The same impact will be overtaken if we change the original DRR scheduling mechanism into an adaptive one. This is why, we are using the original DRR scheduling mechanism with its famous $O(1)$ complexity in order to guarantee a similar level of complexity for the proposed architecture.

Table 2.1: List of important notations in this chapter.

R	Transmission rate of output link (bit/sec).
R_{res}	Reserved transmission rate (bit/sec).
R_{DRR}	Transmission rate handled by the DRR scheduler (bit/sec).
R_v	Voice traffic rate (bit/sec).
R_b	Best effort traffic rate (bit/sec).
R_{vDRR}	offered transmission rate of DRR-voice-queue (bit/sec).
β	Reserved portion of voice traffic (i.e. the splitting ratio) .
R_{token}	Token Bucket rate (bit/sec).
Q_{min}	Minimum quantum value in the DRR scheduler (bit)
Max	Maximum packet size.
$Sent_v(t_1, t_2)$	Amount of voice bits sent in time interval(t_1, t_2).
$Sent_b(t_1, t_2)$	Amount of best-effort bits sent in time interval(t_1, t_2).
N	Number of active queues dedicated to the non-voice traffic in the DRR scheduler .
F	Summation of the active queues' quantum in the DRR scheduler.
Q_b	Summation of all the best effort flows quantum served by the the DRR scheduler (bit)
Q_v	quantum of the DRR-voice-queue.
w_{vDRR}	Weight of DRR-voice-queue.
w_{iDRR}	Weight of flow i inside the DRR scheduler.
w_v	Weight of voice traffic in general.
w_b	Weight of best effort (non-voice)traffic in general.
m	The number of service round received by voice traffic through the DRR module during the interval (t_1, t_2).

2.5 Analytical Analysis

In this section, we conduct a detailed analytical analysis of the proposed scheduler. We develop analytical models to analyze the fairness, packet delay/delay jitter and buffer size requirement of the DRR-voice-queue inside the DRR module. The notations employed in our analysis are summarized in Table 2.1.

2.5.1 Fairness Analysis

In our analysis, we use the same concept of the throughput Fairness Measure (FM) due to Golestani [35], which measures the worst case difference between the normalized services received by different flows that are backlogged during any time interval. Since we want to measure the throughput fairness among the classes and not the flows, so we extend the previous definition to cover the worst case difference between the normalized services received by the voice traffic class and the non-voice traffic class that are backlogged during any time interval.

Definition 1: A class of traffic is backlogged during an interval I if this class is never empty of packets during the interval I.

Clearly, it makes no sense to make this comparison while one of these classes is not backlogged, because the former does not receive any service when it is not backlogged. We assume that each class i has a weight w_i assigned according to its traffic rate. Let $Sent_i(t_1, t_2)$ be the total number of bits sent on the output line by this class interval (t_1, t_2), then the normalized service received by a class i is just $[Sent_i(t_1, t_2)/w_i\ ]$.

In our scheduler, the non-voice packets coming in on different flows are stored in different queues and directly served by the DRR module. For the flow i with rate R_{iDRR}, a weight W_{iDRR} is assigned to it according to the following rule:

$$W_{iDRR} = \frac{R_{iDRR}}{R_{min}} \tag{2.1}$$

where R_{min} is a small rate designated for weight calculation. The assigned quantum Q_i of flow i is calculated depending on its weight as follow:

$$Q_i = W_{iDRR} \cdot Q_{min} \tag{2.2}$$

If the minimum quantum Q_{min} is chosen not less than the maximum packet length Max in the network, then the algorithm time complexity will be $O(1)$. Although the DRR module schedules N of the non-voice flows, each with different rate, weight and quantum, but in our fairness measurement, we refer to them as one aggregated flow (one class) with one rate equal to the summation of their individual rates. We will refer to this class as the Best-effort one with rate R_b, weight w_b and quantum Q_b calculated as follow:

$$R_b = \sum_{i=1}^{N} R_{iDRR} \tag{2.3}$$

$$w_b = \frac{R_b}{R_{min}} = \sum_{i=1}^{N} W_{iDRR} \tag{2.4}$$

$$Q_b = \sum_{i=1}^{N} Q_i \tag{2.5}$$

On the other hand, the voice traffic, with rate R_v, has a weight w_v equal to

$$w_v = \frac{R_v}{R_{min}} \tag{2.6}$$

This traffic is divided into two parts (please refer to Figure 2.2). The first part has an upper bound rate R_{token} that is also equal to $(\beta \cdot R_v)$,where $0 < \beta < 1$. We will refer to β in the rest of this chaper as the *splitting ratio*, and it presents the portion of the voice traffic which is served by the reserved bandwidth. The second part of the voice traffic has a rate R_{vDRR}, which can be expressed as $[(1 - \beta) \cdot R_v]$ and which is served through the DRR module with weight w_{vDRR} and quantum Q_v assigned to the DRR-voice-queue as follow:

$$w_{vDRR} = \frac{R_{vDRR}}{R_{min}} \tag{2.7}$$

$$Q_v = w_{vDRR} \cdot Q_{min} \tag{2.8}$$

We can now express the Fairness Measure of the proposed scheduler as the maximum difference between $[Sent_v(t_1, t_2) \ / w_v]$ and $[Sent_b(t_1, t_2)/w_b]$ in the backlogged time interval (t_1, t_2) .This difference should not depend on the size of the time interval (t_1, t_2). Moreover, If the FM is small, this indicates that the service discipline is closely emulates an ideal fair scheduler.

Lemma 1 *Consider any execution of the proposed scheduling scheme and any interval (t_1, t_2) of any execution such that voice class is backlogged during (t_1, t_2)). Let m be the number of service round received by voice traffic through the DRR module during the interval (t_1, t_2), Then*

$$R_{token} \cdot (t_2 - t_1) + m \cdot Q_v - Max \leq Sent_v(t_1, t_2) \leq R_{token} \cdot (t_2 - t_1) + m \cdot Q_v + Max$$

Since the voice traffic is divided in two parts, the rate of the first part will never excesses the token bucket rate R_{token}. Therefore, the upper bound of service received by this part in time interval (t_1, t_2) is equal to $[R_{token} \cdot (t_2 - t_1)]$. On the other hand, the second part of the voice traffic will be re-directed to the DRR module in which it will be treated as an ordinary flow. It has been proven in [10] that, in the DRR scheduler, the amount of bits sent by any backlogged flow i in time interval (t_1, t_2) is $[m \cdot Q_i \pm Max]$. Therefore, the lemma follows by summing the two parts of the voice traffic. The following theorem proves the fairness of the proposed scheduler, and that this fairness is controllable through the splitting ratio.

Theorem 1 *For an backlogged interval (t_1, t_2) in execution of the proposed scheduler service discipline*

$$\left|\frac{Sent_v(t_1,t_2)}{w_v} - \frac{Sent_b(t_1,t_2)}{w_b}\right| \leq (Q_{min}\cdot(1+\beta)) + \frac{Max}{w_v} + \frac{N\cdot Max}{w_b}$$

by decomposing the voice class from lemma 1 we get the following equation:

$$\begin{aligned} Sent_v(t_1,t_2) &\leq R_{token}\cdot(t_2-t_1)+m\cdot Q_v+Max \\ &= \beta\cdot R_v\cdot(t_2-t_1)+m\cdot Q_v+Max \end{aligned} \tag{2.9}$$

Thus, we can calculate the normalized service received by the voice class as follow:

$$\frac{Sent_v(t_1,t_2)}{w_v} \leq \frac{\beta\cdot R_v\cdot(t_2-t_1)}{w_v} + \frac{m\cdot Q_v}{w_v} + \frac{Max}{w_v} \tag{2.10}$$

From Equation (6), (7) and (8)

$$\frac{Sent_v(t_1,t_2)}{w_v} \leq \beta\cdot R_{min}\cdot(t_2-t_1) + m\cdot(1-\beta)\cdot Q_{min} + \frac{Max}{w_v} \tag{2.11}$$

Using the DRR algorithm invariant which state that the difference in the number of round-robin opportunities given to flow i and flow j in the time interval (t_1, t_2) is $|m - m'| \leq 1$, in addition to the DRR scheduler lemma, we can easily show that the non-voice class can receive the following amount of service

$$Sent_b(t_1,t_2) \geq m'\cdot Q_b - N\cdot Max \tag{2.12}$$

Where N is the number of active queues dedicated to serve the non-voice flows in the DRR module. Then the normalized service received by the non-voice class can be expressed as.

$$\frac{Sent_b(t_1,t_2)}{w_b} \geq \frac{m'\cdot Q_b}{w_b} - \frac{N\cdot Max}{w_b} \tag{2.13}$$

From Equations (4) and (5)

$$\frac{Sent_b(t_1,t_2)}{w_b} \geq m'\cdot Q_{min} - \frac{N\cdot Max}{w_b} \tag{2.14}$$

By combining Equation (11) and (14) we get

$$\left|\frac{Sent_v(t_1,t_2)}{w_v} - \frac{Sent_b(t_1,t_2)}{w_b}\right| \leq \beta[R_{min}\cdot(t_2-t_1) - m\cdot Q_{min}] + Q_{min} + \frac{Max}{w_v} + \frac{N\cdot Max}{w_b} \tag{2.15}$$

The term of $[R_{min}\cdot(t_2-t_1) - m\cdot Q_{min}]$ contains two different operands; the first one

$[R_{min}\cdot(t_2-t_1)]$ is continuous, in term of time (t_2-t_1), and the second operand $(m-Q_{min})$ is a discrete, in term of round m. The relation between these two operands is tight because m represents the number of round-robin rounds that can be accomplished in the time interval (t_1, t_2). In order to simplify this relation, we expressed m in term of time as follow:

$$m = \left\lfloor \frac{(t_2-t_1)}{(Q_v+Q_b)/R_{DRR}} \right\rfloor \tag{2.16}$$

Where R_{DRR} is the bandwidth handled by the DRR module. Then $(m\cdot Q_{min})$ can be now expressed as follow:

$$m\cdot Q_{min} = \left\lfloor \frac{(t_2-t_1)\cdot R_{DRR}}{(Q_v+Q_b)} \right\rfloor \cdot Q_{min} \tag{2.17}$$

From Equation (5), (6), (7) and (8)

$$m\cdot Q_{min} = \left\lfloor \frac{(t_2-t_1)\cdot R_{min}\cdot R_{DRR}}{(R_{vDRR}+R_b)} \right\rfloor \tag{2.18}$$

In the worst case, the R_{vDRR} and the R_b will saturate (equal to) bandwidth R_DRR handled by the DRR module, and then, we can approximate $(m\cdot Q_{min})$ as follow:

$$m\cdot Q_{min} \geq \lfloor (t_2-t_1)\cdot R_{min} \rfloor \tag{2.19}$$

this simplification means that with the increase of time (t_2-t_1) the number of round m equivalently increase too and that $m\cdot Q_{min}$ will always be the floor of $[R_{min}\cdot(t_2-t_1)]$. Therefore, the subtracting of $[R_{min}\cdot(t_2-t_1) - m\cdot Q_{min}]$ will maximally result into one complete round which is equal to Q_{min}. Thereby, the theorem follows

$$\left| \frac{Sent_v(t_1,t_2)}{w_v} - \frac{Sent_b(t_1,t_2)}{w_b} \right| \leq (Q_{min}\cdot(1+\beta)) + \frac{Max}{w_v} + \frac{N\cdot Max}{w_b}$$

The result of this analysis shows that the proposed scheduler can keep this fairness index bounded and independent of both voice traffic amount and backlog time length. Moreover, this result shows that the increases or decreases of β can partly degrades or enhances the fairness respectively, as long as the $(\beta < 1)$; but if $(\beta = 1)$, the fairness will diverge to infinity because of the disappearance of $[m\cdot(1-\beta)\cdot Q_{min}]$ from Equation (15) which means that the whole voice traffic rate will be served by the reserved bandwidth.

2.5.2 The voice Packet Delay and Delay Jitter upper bound

With respect to those voice packets which are served by the reserved bandwidth, their packet delay is in term of μ sec because of the use of the token bucket module, which checks every packet's size and then allowed for only a given rate to be served through reserved bandwidth. For the rest of the voice packets, they will suffer from delay due to the fair competition with other non-voice packets. Generally, the DRR scheduler belongs to the class of LR servers. S. Kanhere and H. Sethu report in [36] a tight upper bound of the packet latency θ_i of any flow i in the DRR scheduler, then later, Anton Kos et al correct this upper bound in [37]. In our analysis we used the former upper bound especially with the lack of accurate delay distribution for this type of scheduler. Equation 20 shows the delay upper bound θ_i for any flow i.

$$\theta_i \leq Q_i\left(\frac{1}{R_i}-\frac{1}{R}\right)+(Max-1)\left(\frac{Q_i}{FR_i}N+\frac{1}{R_i}-\frac{2}{R_{DRR}}\right) \tag{2.20}$$

Since, only $[(1-\beta)\cdot 100]$ percent of the voice traffic may suffer from this delay because of being served through the DRR module, so thanks to the previous equation, we derived the maximum average latency D_{voice} of the voice traffic as follow:

$$D_{voice} \leq (1-\beta)\left[Q_v\left(\frac{1}{R_{vDRR}}-\frac{1}{R_{DRR}}\right)+(Max-1)\left(\frac{Q_v}{FR_{vDRR}}N+\frac{1}{R_{vDRR}}-\frac{2}{R_{DRR}}\right)\right] \tag{2.21}$$

Delay jitter may occur between two packets of the same voice flow if one of them is served directly by the reserved bandwidth (i.e., it will noty suffer from delay) and the other is served by the DRR module (i.e., it will suffer from delay due to fair competition with non-voice traffic). The upper bound of such delay jitter D_{jitter} is equal to maximum packet delay that any packet may encounter in the DRR module as derived in The following equation

$$D_{jitter} \leq \left[Q_v\left(\frac{1}{R_{vDRR}}-\frac{1}{R_{DRR}}\right)+(Max-1)\left(\frac{Q_v}{FR_{vDRR}}N+\frac{1}{R_{vDRR}}-\frac{2}{R_{DRR}}\right)\right] \tag{2.22}$$

In general the packet delay of any flow i in the DRR module is directly related to its relative weight w_i and its assigned quantum Q_i as a result (i.e., the weight of the observed flow with respect to the weight of the other competitor flows); So, a flow with a relative small weight usually suffer from a high packet delay. In our case, the increase of splitting ratio results in a decrease of the amount of voice traffic R_{vDRR} served by the DRR module and hence its relative weight w_{vDRR} and quantum Q_v too. Although the splitting ratio does not appear in the above equation, it actually controll both of R_{vDRR} and Q_v. This is why, the delay-jitter will increase with the increase of the splitting ratio. Finally, We have to notice that only $[(1-\beta)100]$ percent of the voice traffic that may suffer from this delay jitter, and which can be recovered through the jitter buffer at the receiver node.

2.5.3 The DRR-voice queue Buffer size

Generally, the buffer size should be large enough to allow the voice application to endure the normal variance in the scheduler latency without suffering from packet loss [38]. In the proposed scheduler, those voice packets, which are re-directed to the DRR module and buffered in the DRR-voice-queue, will wait for their turn of service according to the DRR mechanism; therefore, we estimate the proper buffer size of the DRR-voice-queue. Due to the availability of the upper delay bound only, we have estimate the maximum buffer size B_{vDRR} of the DRR-voice-queue depending on the maximum delay jitter D_{jitter} as of follow:

$$B_{vDRR} = D_{jitter}(1 - \beta)R_v \tag{2.23}$$

This equation was derived in agreement with the well known thumbing rule [39], which states that the maximum buffer size of any queue is equal to the product of both the maximum delay and the service rate of this queue. This equation is verified through simulation experiments in the next section.

2.6 Simulation Results

In this section, we wish to answer the following questions about the performance of the proposed scheduler:

- Is the fairness independent of the backlogged time interval's length? Is the fairness sensitive to the splitting ratio change? Does the workload variation of both voice and non-voice traffic affect the fairness?
- How does the splitting ratio variation affect the delay and delay-jitter performance of the voice traffic? What is the delay performance of the voice packets served through the DRR module especially with the increase of their workloads?
- does the buffer size model for the DRR-voice-queue reflect the simulation case or not?
- What is end-to-end VoIP QoS performance of the proposed scheduler in practical situation?

2.6.1 Default Simulation Setting

Unless otherwise specified, the default for all the later experiments is as specified here. We simulate the behavior of our scheduler in a single node with OC1 output link capacity (51.89 Mbps). This node receives voice and non-voice flows generated according to a

Poisson distribution. The voice flows were generated according to the G.729A codec specifications with 8kbps for every single voice stream with 20 ms of voice payload. On the other hand, 60% of the non-voice packets sizes were settled as 44 bytes, 20% as 550 bytes and the rest 20% as 1500 bytes. Every point in the following fairness or delay graphs was derived in simulation interval, typically 2000 sec.

2.6.2 The Fairness Analysis

We designed a set of experiments to examine the fairness provided by the proposed scheduler to both voice and non-voice traffic in general. The following experiments are conducted under different backlogged time interval's lengths, various splitting ratio and several workloads. Due to the relation between these parameters, we have devised experiments to isolate each of them, and then we tested every parameter individually.

Table 2.2: The normalized throughput percentage at different backlogged time intervals.

Backlogged time interval	Normalized throughput percentage	
	Voice traffic	Non-voice traffic
10 sec	49.877%	50.123%
100 sec	49.827%	50.173%
1000 sec	49.950%	50.050%

In the first experiment, we examine the difference between the normalize service received by both backlogged voice and non-voice traffic during different backlogged time interval's length. We generate both voice and non-voice traffic with workload equal to 0.4 for each of them, and fix the splitting ratio to 0.5. The normalized throughput of both voice and non-voice traffic has been acquired at different backlogged[1] time lengths typically equal to 10, 100 and 1000 sec. The simulation results, as illustrated Table 2.2, show that the resulting normalized throughput percentage of both voice and non-voice traffic classes are almost equal even with the increase of inspection time period. This result confirms that the fairness provided by the proposed scheduler is independent of time.

In the second experiment, we addressed the impact of the splitting ratio on the difference between the normalize service received by both backlogged voice and non-voice traffic. We fixed the workload of both voice and non-voice traffic to 0.4 for each of them. The normalized throughput of both traffic has been inspected at different splitting ratio. The result, as shown in Figure 2.3, shows an enhancement in the fairness with the decreases of the splitting ratio. This enhancement is due to the decrease of reserved bandwidth offered to the voice traffic, and hence the increase of fair competition between

[1]A flow is called backlogged if it always has a packet waiting for service

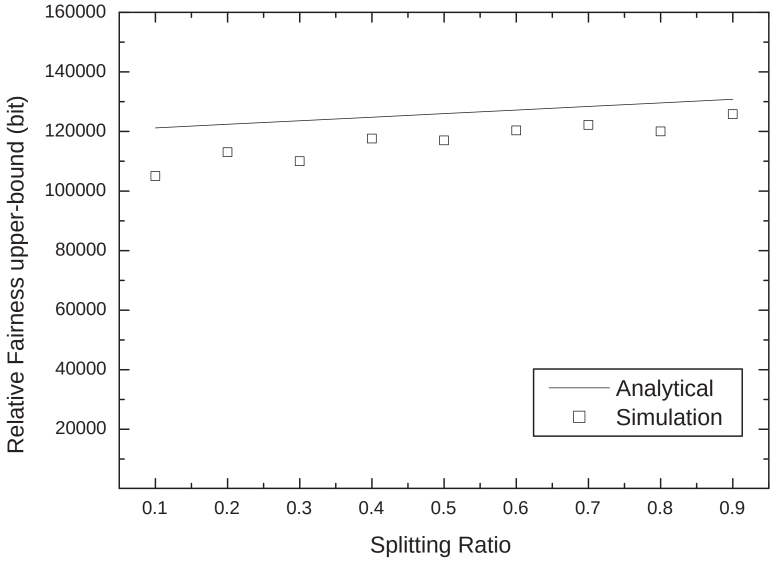


Figure 2.3: The normalized throughput of the voice traffic with respect to non-voice traffic at different splitting ratio.

the overflowed voice traffic and the non-voice traffic in the DRR module. We can also notice that the simulation results match well with the upper bound derived in the previous section. Therefore, we considered the result of this experiment a verification of our theorem.

In the last experiment of this set, we examined the effect of the workload variation of both voice and non-voice traffic on the fairness. We fix the splitting ratio at 0.5, and the simulation time period to its default value, then we acquired the normalized throughput of both voice and non-voice traffic under different combination of their workloads. The results, as shown in Figure 2.4, record the absence of any effect on the difference between the normalized throughputs of these two classes even with the variation of their traffic's workloads. Through these simulation results, we show that the fairness provided by the proposed scheduler is independent of both backlogged time period and workload variation. On the other hand, we show that the splitting ratio is the only metric which affects on the fairness index.

2.6.3 The Voice Packet Delay and Delay-Jitter Analysis

In the following experiments, we inspect the delay and delay-jitter performance of the voice traffic in the proposed scheduler under various splitting ratios and compare it with other scheduling schemes, including Deficit Round-Robin (DRR), Weighted Fair Queueing (WFQ), Surplus Round-Robin (SRR), and Strict Priority (SP) schemes.

In our scheduler, two voice packets that belong to the same flow may be separated due to the splitting mechanism adopted in the proposed scheduler. In the first experiment, we

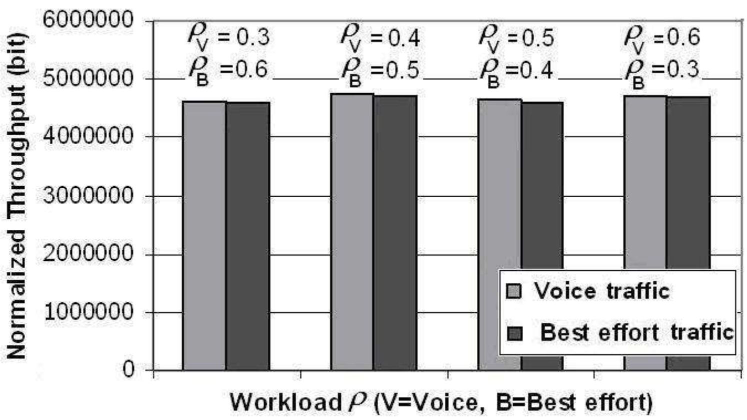


Figure 2.4: The normalized throughput of the voice traffic with respect to non-voice traffic at different workloads.

inspect the delay-jitter between these two packets. We fix the workload of the both voice and non-voice traffic to 0.4 for each of them. We then record the delay-jitter variation with the the increase of splitting ratio from 0.1 up to 0.9. As shown in Figure 2.5, the results indicate that delay-jitter increase with the increase splitting ratio. In general the packet delay of any flow in the DRR module is directly related to its relative weight and hence on its assigned quantum (i.e., the weight of the observed flow with respect to the weight of the other competitor flows); So, a flow with a relative small weight usually suffer from a high packet delay. In our case, the increase of splitting ratio results in a decrease of the amount of voice traffic served by the DRR module and hence its relative weight and quantum too. This is why, the delay-jitter will increase with the with the increase of the splitting ratio. It is also notable that the simulation results of the delay-jitter performance match well with the upper bound derived in the previous section. Therefore, we considered the result of this experiment a verification of our theorem.

In the second experiment, we target to compare the voice packet delay performance of the proposed scheduler in comparison with other scheduling schemes, including Deficit Round-Robin (DRR), Weighted Fair Queueing (WFQ), Surplus Round-Robin (SRR) and Strict Priority (SP) schemes. We fix the workload of the both voice and non-voice traffic to 0.4 for each of them. We then record the delay variation with the the increase of splitting ratio from 0.1 up to 0.9. The splitting mechanism (for both voice traffic and bandwidth) adopted in our scheduler offers us the ability to impose a limit on the priority offered to the voice traffic by controlling the amount of traffic served through the reserved bandwidth, while enhancing the fairness deserved by the non-voice traffic by sacrifice a portion of voice packets delay through the DRR module. As shown in Figure 2.6, the results of this experiment show that the voice packets delay performance of the proposed scheduler lies in between that of the DRR and the SP schemes. Such behavior indicates

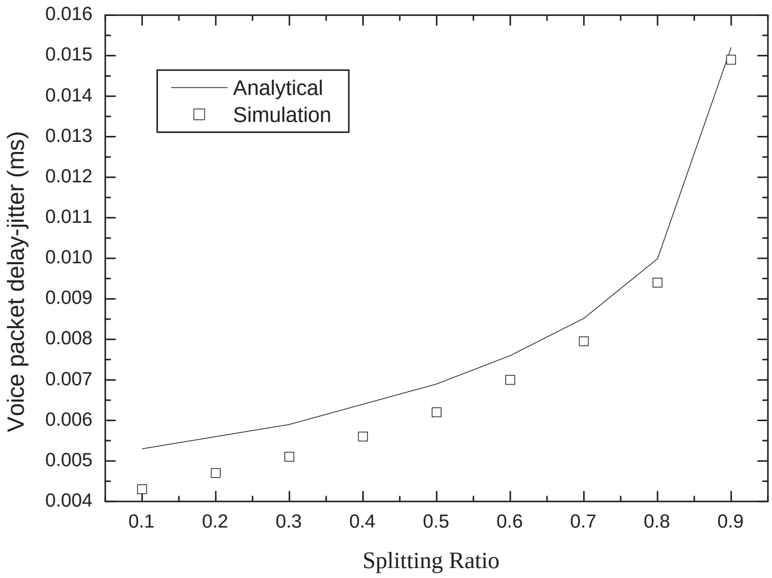


Figure 2.5: The delay-jitter upper bound at different splitting ratios.

the ability of controlling the performance of our scheduler through the splitting ratio to emulate any of the later schemes according to the network policy. We can also notice the performance of the other scheduler, like the WFQ which can report better packet delay performance in compare with the DRR but they still also report a high complexity $O(logn)$.

2.6.4 The Buffer Size of the DRR-voice-queue

In the proposed scheduler, part of the voice traffic is served through the DRR module, in which the voice packet will be temporarily buffered in a queue called the DRR-voice-queue until it get its chance to be served according to the DRR packet scheduling mechanism. We conducted an experiment to verify our analytical maximum length model in Equation (23). In our experiment, we fixed the non-voice workload to 0.4 and then inspected the maximum buffer size of the DRR-voice-queue when the voice traffic workload grows from 0.1 up to 0.4. The experiment results and the corresponding modeling results are summarized in Figure 2.7, which shows clearly that our analytical buffer size model is very efficient and it can be used to dimension the buffer size of DRR-voice-queue without causing any voice packet dropping problem.

2.6.5 The Assessment of VoIP QoS

Due to the existence of two paths for the voice traffic in the proposed scheduler, the amount of voice packets served by the DRR module exert an additional delay (delay-jitter) because of the competition with the best-effort traffic. Nevertheless, the impact

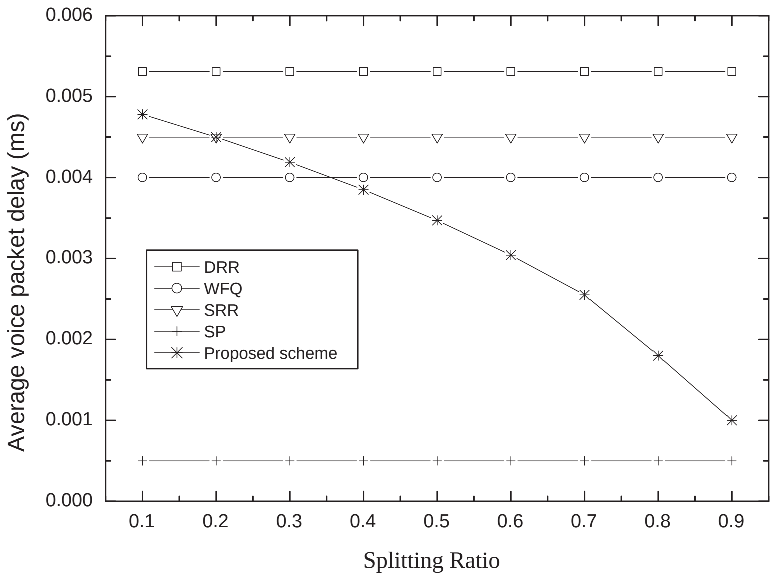


Figure 2.6: Voice packet delay of the proposed scheduler in comparison with relevant schedulers at different splitting ratios.

of such delay jitter can be easily eliminated by adopting a small size jitter-buffer at the receiver node. We now report the results of a new simulation experiment. This experiment is designed to investigate the delay-jitter impact on the end-to-end QoS of the VoIP calls delivered by our scheduler in practical situation. The QoS measurements have been conducted using the famous R-score metric [40].

Simulation Settings: Figure 2.8 shows the network topology used in the experiment. All the links have a bandwidth of $2Mbps$ and propagation delay of $1ms$. We generate 100 VoIP flows (each with a rate of $8Kbps$) from S_0 to D_0. In addition to the 100 observed VoIP flows, the background traffic in the system is as follows. there are 10 best-effort flows with rate $80Kbps$ from each of the following nodes $(S_1, S_2, \cdots, S_5)$ to $(D_1, D_2, \cdots, D_5)$, respectively. The number of end-to-end hopes is chosen based of similar conditions in [41] and [42]. The VoIP and best-effort flows were generated according the default simulation settings

VoIP QoS Measure: The R-score takes into account one way delay, loss rate, and the type of the encoder. For the G.729A encoder [40], the R-score is as follow

$$R = 94.2 - 0.024d - 0.11(d - 177.3)H(d - 177.3) - 11 - 40log(1 + 10e) \qquad (2.24)$$

where

- $d = 25 + d_{jitter_buffer} + d_{network}$ is the total one-way delay in ms comprising of 25ms voice encoder delay, delay in the delay-jitter buffer (30ms, 40ms, 50ms, and 60ms), packet delay.

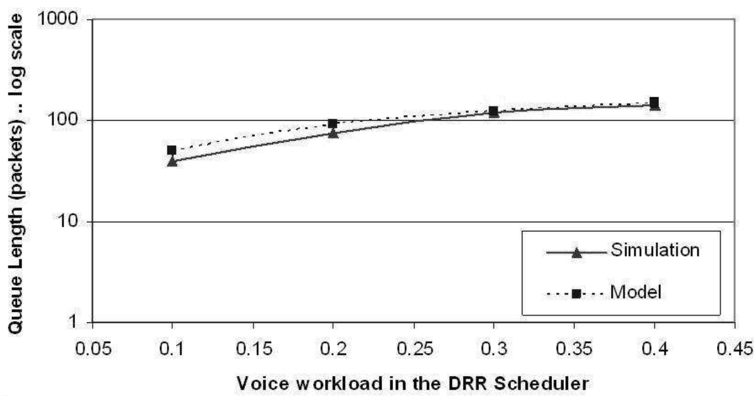


Figure 2.7: A comparison between the simulation and analytical max-length of the DRR-voice-queue.

- $e = e_{network} + (1 - e_{network})e_{jitter}$ is the total loss rate including network and jitter losses

- $H(x) = 1$ if $x \geq 0$, else 0. R-score should be larger than 70 for acceptable call quality.

The quality of VoIP call is sensitive to delay, delay jitter and loss. In order to maintain a good call quality ($R \geq 70$), the one way delay should be less than 200ms and the packet loss rate along the path should be less than 5%. With the increase of delay and packet loss, the VoIP quality is deteriorated.

Simulation Results: Figure 2.9 shows the average R-score for VoIP flows. The R-score of the 40ms jitter-buffer reports an acceptable QoS of the delivered VoIP calls. Such performance was significantly better in comparison with the 30ms jitter-buffer, but almost similar to the 50ms and 60ms jitter-buffer cases. Such results indicates that the jitter-buffer can efficiently conceal the impact of delay-jitter even with a small size. It also worth to note that the performance of the proposed scheduler can be easily controlled through the splitting ratio β, and hence the trade-off between delay and fairness can be easily achieved through the proposed scheduler.

2.7 Summary

In this chapter, we proposed a scheduling architecture for VoIP application. The new architecture is flexible in the sense that we can easily comprise between the packet delay and fairness through only one control parameter (i.e., the splitting ratio of voice traffic). Based on our new architecture, we proved through both the theoretical analysis and experimental simulation that it is possible for us to offer a certain degree of fairness to

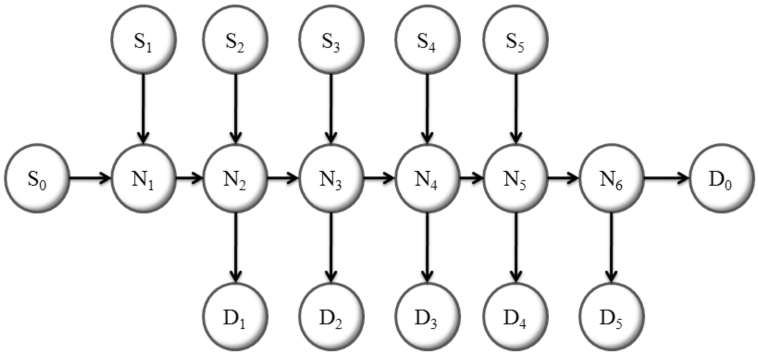


Figure 2.8: Simulated network topology.

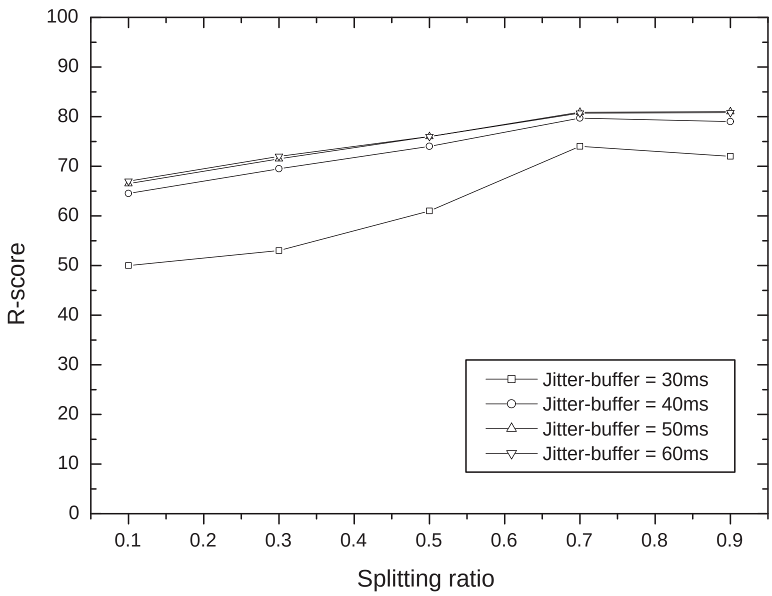


Figure 2.9: End-to-end VoIP QoS inspection at different splitting ratios.

non-voice traffic without significantly sacrificing the performance of delay-sensitive voice traffic. We expect that our proposed architecture can efficiently handle the impacts that will rise up with the expected growth of VoIP traffic in the future voice-intensive IP networks.

Chapter 3

Voice over WLAN

3.1 Introduction

Efficient VoIP support at the wireless access point of a Wireless LAN (WLAN) remains a challenge for the last-mile wireless coverage of IP networks with mobility support. Due to the limited bandwidth available in WLANs, an accurate analysis of voice capacity in such networks is crucial for the efficient utilization of their resources. The available analytical models only provide the upper and lower bounds on voice capacity, which may significantly overestimate or underestimate the WLAN's capability of supporting VoIP and thus are not suitable for above purpose. In this chapter, we focus on the voice capacity analysis of a wireless 802.11(a/b) access point running the distributed coordination function (DCF). In particular, we show that by incorporating the clients' spatial distribution into analysis, we are able to develop a new analytical model for a much more accurate estimation of average voice capacity. By properly exploring this spatial information, we further propose a new scheme for access point placement such that the overall voice capacity can be enhanced. The efficiency of the new voice capacity model and new access point placement scheme is validated through both analytical and simulation studies.

3.2 Voice over Wireless LAN

Driven by huge demands for flexible connectivity and portable access at reduced costs, the Wireless Local Area Networks (WLANs) are increasingly making their way into residential, commercial, industrial and public areas. While the majority of traffic in WLAN is data, it is expected that the voice application will become a significant driver for the deployment of WLANs [43, 44, 45], as evident from the rapid flourish of the VoIP applications in recent years [46].

When designing a Voice over WLAN system, the most important parameter of concern is the voice capacity of the wireless access point (AP), which is defined as the number of voice connections that can be simultaneously supported through the AP. Since the current WLANs have very limited bandwidth and the voice admission control there mainly depends on this parameter to accept or reject new voice calls [47, 48], so a careful voice capacity analysis is crucial for the efficient utilization of WLANs resources.

The available models for voice capacity analysis only provide the upper and lower bounds on voice capacity [17, 16, 18, 49, 50, 51], because they were developed based on the assumption that the transmission rate (R) between the AP and any client in the WLAN is always either the maximum or the minimum achievable rate. In practice, however, the R varies depending on the access distance, shadowing effect and channel fading along the signal path [20]. Thus, the maximum R is only available for those clients who are very close to the AP, while R is sharply stepped down (non-linearly) as the access distance increases, as illustrated in Figure 3.1 for the 802.11a protocol[1]. Therefore, the available simple models may significantly overestimate or underestimate the voice capacity.

In this chapter, we focus on the voice capacity analysis and enhancement of DCF-based IEEE 802.11a/b WLANs. The main contributions of our work are the following:

1. With the consideration of clients spatial distribution (CSD), we develop a new analytical model for a much more accurate estimation of WLANs' voice capacity.

2. By properly exploring the CSD information, we further propose a new scheme for access point placement such that the overall voice capacity can be enhanced.

3. We demonstrate through our new model and scheme that the CSD has a significant implication on the achievable voice capacity and thus it should be carefully considered in the AP placement.

The rest of this chapter is organized as follows. Section II introduces the background and related work. In Section III, we develop an analytical model to estimate the voice capacity based on the CSD. In Section IV, we introduce a new scheme for voice capacity enhancement. Section V presents the validation of our model and scheme. Finally, we conclude this chapter in Section VI.

3.3 Background and related work

In this section, we introduce the IEEE 802.11 DCF-based WLAN and Voice Capacity of WLAN.

[1]In practice, due to other factors like the channel fading, shadowing effect, hidden nodes etc, the signal transmission rate reduction may not exactly follow the very regular pattern shown in Fig. 3.1. The Fig. 3.1 is mainly used for illustration here.

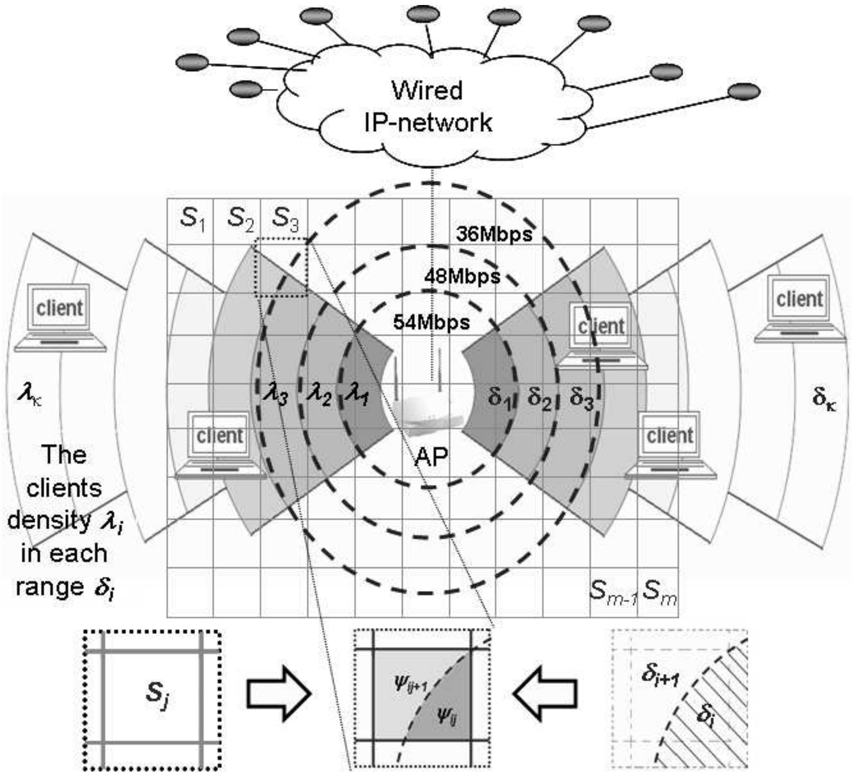


Figure 3.1: Illustration of 802.11a transmission rates.

3.3.1 IEEE 802.11 DCF-based WLAN

The basic access method in the IEEE 802.11 WLANs is the distributed co-ordination function (DCF), which is based on carrier sense multiple accesses with collision avoidance. In the DCF, all clients with packets ready for transmission observe the shared medium before attempting to transmit. If the medium is sensed busy, the clients delay the transmission until the medium is sensed idle for the period of a DCF InterFrame Space (DIFS). The clients then enter the backoff phase, in which every client chose a random backoff counter from [0, CW_{min}] (where CW_{min} is the minimum contention window size). The backoff counter decreases by one for every idle timeslot and freezes if the channel is sensed busy. The decrement procedure is resumed after the channel is sensed idle again for a period of DIFS. The client transmits the packet when the backoff counter reaches zero. If the packet is received successfully, the receiver transmits an ACK following a Short Inter-Frame Space (SIFS). In case of failed transmission due to collision or transmission error, the sender may attempt to retransmit his packet for a specific number of trials before it is dropped, where the contention window size is doubled for each trial until it reaches the maximum value (CW_{max}). Following every successful transmission, contention window size is reset to its initial (minimal) value.

3.3.2 Voice over WLAN's Basics

A VoIP system consists of three indispensable components, namely codec, packetizer and playout buffer [?]. Analog voice signals are first digitized, compressed and encoded into digital voice streams by the codecs. The output digital voice streams are then packed into constant-bit-rate (CBR) voice packets by the packetizer. Each voice packet has a 40-B real-time transport protocol (RTP)/user datagram protocol (UDP)/IP header followed by the voice payload. After voice packets are delivered through an IP network, the decoding and depacketizing are implemented at the receiver. Due to packets delay jitter, a playout buffer is usually used to smooth the speech at receiver.

Different codecs use different compression algorithms, resulting in different bit rates [52, 53, 54]. The G.711 and The G.729 are popular codecs used by VoIP applications, where the G.711 is the international standard for encoding telephone audio with a fixed bit rate of 64 kb/s. If the packetization interval is 10 ms (i.e. the time between two adjacent packets is 10ms, which corresponds to a rate of 100 packets/s), the payload size will be $64000/(100 \times 8) = 80$ Byte. If the packetization interval is increased to 20 ms, which corresponds to a rate of 50 packets/s, the payload size will be reduced to $64000/(50 \times 8) = 160$ Byte and so on. The G.729, on the other hand, is a low bit-rate codec (8 kb/s) at the expense of higher codec complexity. It is used by some available 802.11 VoIP phones (such as the Zyxel Prestige and Senao S7800H). We chose both of these widely used voice codecs in our analysis and simulations throughout this chapter.

3.3.3 Voice Capacity of WLAN

The available analysis on the WLANs' voice capacity of has been conducted via both experiments in [49, 50, 51] and analytical models in [17, 16, 18]. Based on a testbed, chapter [49] shows that the 802.11b can only support ten voice connections by adopting G.711 voice codec, 10ms audio payload and silence suppression. A measurement experiment without silence suppression was carried out in chapter [50], which indicates that only six voice connections can be accommodated. In chapter [51], it has been shown that the 802.11b can support up to ten G.711 and eighteen G.723.1 voice connections with 20 ms and 30 ms of audio payload, respectively. In general, the available measurement results indicate that the silence suppression and audio payload interval heavily affect the voice capacity. These experiments, however, can be only considered as cases study and may not be generally applicable.

On the analytical model side, the model in [17] considers the overheads (e.g., the voice packet header, DIFS, SIFS, ACK packet, and the random backoff) and simplifies the voice capacity analysis with the assumption that there are no collisions. In [16] another better analytical model was proposed by assuming that there are always two active stations competing for the wireless channel. In [18], the authors improved the previous models by considering both the details of collision avoidance mechanism and the practical AP-bottleneck effect induced by unbalanced traffic of AP/clients. Although the available models successfully quantified the factors that affect the voice capacity like overheads, unbalance traffic load and collision, they can only provide the theoretical upper or lower bounds on the voice capacity since they always assume a constant transmission rate (either the maximum or the minimum) in their analysis.

3.4 Analytical Modeling for Voice Capacity

In this section, we introduce the voice capacity analysis and also our new approach for it.

3.4.1 Voice Capacity Analysis

To analyze the AP's voice capacity, we first need to understand the maximum channel throughput that can be achieved in the AP as modeled in [17, 16]:

$$\frac{T_P \cdot R_{avg}}{T_{voice} + T_{Ack} + T_{SIFS} + T_{DIFS} + T_{Backoff}} \tag{3.1}$$

The above equations indicates that the channel throughput is a function of the AP's average transmission rate R_{avg} and other parameters, like the time T_P needed to transmit the voice payload, the time T_{voice} needed to transmit the whole voice packet, the SIFS time T_{SIFS}, the DIFS time T_{DIFS}, the time T_{Ack} needed to transmit the acknowledgment packet, and the backoff time overhead $T_{Backoff}$.

For an IEEE 802.11 WLAN, T_{SIFS} and T_{DIFS} in (1) are usually constants, while $T_P = S_P/R_{avg}$, $T_{voice} = T_{PHY} + ((S_{MAC} + S_h + S_P)/R_{avg})$, and $T_{Ack} = T_{PHY} + (S_A/R_{avg})$ are determined by other basic parameters. The S_{MAC}, S_h and S_p are the sizes of the voice packet' MAC header, RTP/UDP/IP header, and voice payload, respectively. The S_A and T_{PHY} are the Ack. packet size and the time needed to process the physical layer overhead, respectively. Finally, $T_{Backoff}$ in (1) can be evaluated based on the following formula with the consideration of collision [16].

$$T_{backoff} = \begin{cases} 4.5\times 9 + T_w \times 0.06\ \mu s & \text{for the IEEE 802.11a Wlan} \\ 8.5\times 20 + T_w \times 0.03\ \mu s & \text{for the IEEE 802.11b Wlan} \end{cases} \tag{3.2}$$

where T_w is defined in terms of the basic parameters as follow

$$T_w = T_{SIFS} + T_{DIFS} + 2T_{PHY} + \frac{1}{R_{avg}}(S_H + S_P + S_A + S_{MAC}) \tag{3.3}$$

The typical values for above basic parameters (except R_{avg}) are shown in the Table 3.1 (part a)) when the IEEE 802.11a/b standards are considered [55, 56].

Table 3.1: The IEEE 802.11(a/b) DCF-based WLAN Parameters & clients spatial distribution patterns

Part	Item	802.11b	802.11a
a)	Transmission rates R	11, 5.5, 2, and 1Mbps	54, 48, 36, 24, 18, 12, 9, and 6Mbps
	Slot Time	20μs	9μs
	T_{SIFS}	10μs	16μs
	T_{DIFS}	50μs	34μs
	CW_{min}	32	16
	CW_{max}	1024	1024
	Retry Limit	7	7
	T_{PHY}	192μs	24μs
	S_{MAC}	(34 Byte $\times$ 8)	
	S_H	(40 Byte $\times$ 8)	
	S_P	(payload $\times$ 8)	
	S_A	(14 Byte $\times$ 8)	
b)	Pattern1	$(\lambda_1,...,\lambda_4)$= (79,52,39,30)	$(\lambda_1,...,\lambda_8)$=(25,54, 28,24,20,19,16,14)
	Pattern2	$(\lambda_1,...,\lambda_4)$= (10,46,95,49)	$(\lambda_1,...,\lambda_8)$=(2,8,11, 35,60,35,27,22)
	Pattern3	$(\lambda_1,...,\lambda_4)$= (8,14,25,153)	$(\lambda_1,...,\lambda_8)$=(2,6,6,8, 10,15,53,100)

Based on the above channel throughput model, the voice capacity C of the AP (i.e., the number of VoIP connections the AP can simultaneously support) can be estimated as follow:

$$C = \left\lfloor \frac{S_P \cdot R_{avg}}{2L \cdot ((S_H + S_P + S_A + S_{MAC}) + R_{avg}(2T_{PHY} + T_{SIFS} + T_{DIFS} + T_{Backoff}))} \right\rfloor \tag{3.4}$$

where all voice calls are assumed to use the same voice codec with bit rate L b/s.

Based on (2), the maximum number of voice calls that a single AP can support under different transmission rates and payload size are illustrated in Figure 3.2. The Figure indicates clearly that the voice capacity C is significantly affected by the variation of R_{avg}. For example, the IEEE 802.11a WLAN can support up to ninety two G.729 voice connections (with 30ms payload size) if R_{avg} is equal to 54Mbps. This capacity is dramatically decreased to forty eight if R_{avg} decrease to 6Mbps. Therefore, an accurate model of R_{avg} is crucial for the efficient estimation of the voice capacity in a WLAN.

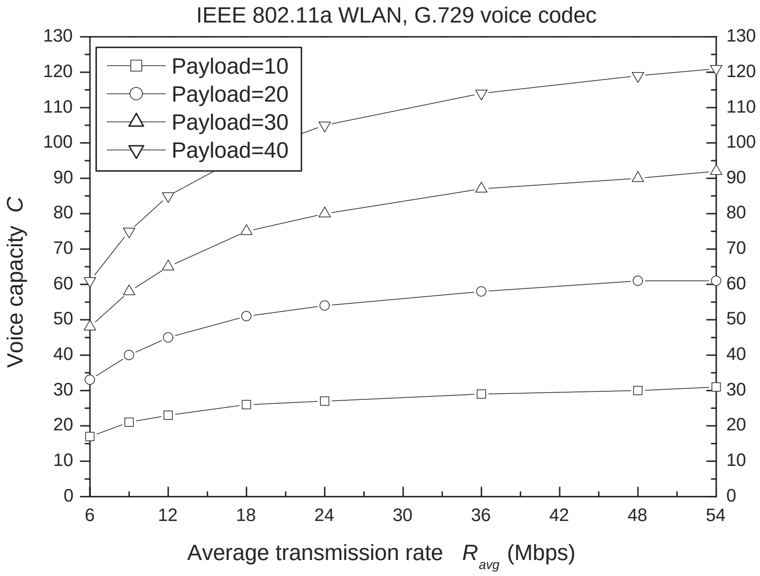


Figure 3.2: The voice capacity of 802.11a under different R_{avg}

3.4.2 A New Model for R_{avg}

For an accurate estimation for R_{avg}, we first need to characterize the spatial distribution of the clients. Given a WLAN coverage area, we use grid to divided it equally into m squares (S_1,S_2,..., S_m), as illustrated in Figure 3.1. Suppose that the number of clients N_{S_j} in square S_j ($j = 1, 2, \ldots, m$) follows the Poisson distribution[2] with mean value λ_{S_j} and these clients N_{S_j} are uniformly distributed inside the square[3], then the spatial distribution of clients in the WLAN area can be described in terms of $(\lambda_{S_1}, \lambda_{S_2}, \cdots, \lambda_{S_m})$.

The step-down of transmission rate in one AP's coverage area virtually creates multiple mutually exclusive rate ranges $(\delta_1, \ldots, \delta_k)$ around it (please refer to Fig. 3.1), where each range δ_i has a particular transmission rate r_i, $i = 1, ..., k$. A number of models have been developed to predict the distance limit of each r_i for a given AP, see, for example [59, 60]. Thus, the area size of each delta range δ_i can be easily calculated by treating these distance limits as the radius of multiple overlapped circles sharing the same center point (i.e., the AP).

Let ψ_{ij} denote the intersection between δ_i and S_j, that is

$$\psi_{ij} = \ S_j \cap \delta_i, \ \ \text{for } 1 \le i \le k, \text{ and } 1 \le j \le m \tag{3.5}$$

[2]In practice, due to the mobility of clients in WLAN, they may randomly enter and leave a square independently of each other. Therefore, the number of clients in a square can be generally regarded as a random process similar to the random occurrence of an event per a specified region, which can be nicely described by Poisson distribution. That is why the Poisson distribution assumption has been widely adopted to describe the number of mobile clients in a specified area [57, 58].

[3]When a square is small enough, the clients distribution in it can be nicely approximated as uniform.

Then each δ_i can be expressed in terms of ψ_{ij} as follow

$$\delta_i = \cup_{j=1}^{m} \psi_{ij}, \quad \text{for } i = 1, \ldots, k \tag{3.6}$$

We denote the number of clients in ψ_{ij} by $N_{\psi_{ij}}$. Since the clients are uniformly distributed inside a square S_j, then $N_{\psi_{ij}}$ also follows the Poisson distribution with mean value $\lambda_{\psi_{ij}}$ determined as:

$$\lambda_{\psi_{ij}} = \frac{|\psi_{ij}|}{|S_j|} \times \lambda_{S_j} \tag{3.7}$$

where $|S|$ denotes the size of area S. We denote the total number of clients in range δ_i by $N_i = \sum_{j=1}^{m} N_{\psi_{ij}}$. Then we have the following Lemma regarding the distribution of N_i.

Lemma 2 *The random variable N_i follows the Poisson distribution with mean value $\lambda_i = \sum_{j=1}^{m} \lambda_{\psi_{ij}}$.*

We can easily see that

$$N_i = \sum_{j=1}^{m} N_{\psi_{ij}} \tag{3.8}$$

Let $M_x(e^t)$ denote the Moment Generating Function (MGF) of a random variable x, then we have the following formula based on (6) and the independence among $N_{\psi_{ij}}$.

$$M_{N_i}(e^t) = \prod_{j=1}^{m} M_{N_{\psi_{ij}}}(e^t) \tag{3.9}$$

For a Poisson random variable x with mean value λ_x, its MGF is given by

$$M_x(e^t) = e^{(t-1)\lambda_x} \tag{3.10}$$

Thus, we have

$$M_{N_i}(e^t) = \prod_{j=1}^{m} e^{(t-1)\cdot\lambda_{\psi_{ij}}} = e^{(t-1)\cdot\sum_{j=1}^{m}\lambda_{\psi_{ij}}} \tag{3.11}$$

Based on the one-to-one mapping property between a random variable and its MGF, we can see that N_i follows the Poisson distribution with mean value $\lambda_i = \sum_{j=1}^{m} \lambda_{\psi_{ij}}$.

Now, the spatial distribution defined in terms of $(\lambda_{S_1}, \lambda_{S_2}, \cdots, \lambda_{S_m})$ is now transformed into the terms of $(\lambda_1, \lambda_2, \cdots, \lambda_k)$.

If we further denote the total number of clients in the coverage area of the access point by N ($N = \sum_{i=1}^{k} N_i$), then using Lemma 1, we can easily prove that the random variable N also follows the Poisson distribution with mean value $\lambda = \sum_{i=1}^{k} \lambda_i$. Before presenting the main result about the R_{avg} of an access point, we first establish the following Lemma about a property of the random variable N.

Lemma 3 *For any set of non-negative variables $x_i \geq 0$, $i = 1, \ldots, k$, let $X = \sum_{i=1}^{k} x_i$. Then we have*

$$Pr(N_1 = x_1, \cdots, N_k = x_k | N = X) = \frac{X!}{x_1! \cdots x_k!} \prod_{i=1}^{k} \left(\frac{\lambda_i}{\lambda}\right)^{x_i} \tag{3.12}$$

Notice that N_i follows the Poisson distribution with mean value $\lambda_i = \sum_{j=1}^{m} \lambda_{\psi_{ij}}$ ($i = 1, \ldots, k$), then N_i is distributed as:

$$Pr(N_i = x_i) = \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!} \tag{3.13}$$

The mutual independence of N_i's indicates that their joint probability distribution is given by

$$Pr(N_1 = x_1, \cdots, N_k = x_k) = \prod_{i=1}^{k} \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!} = e^{-\lambda} \prod_{i=1}^{k} \frac{\lambda_i^{x_i}}{x_i!} \tag{3.14}$$

Where $\lambda = \sum_{i=1}^{k} \lambda_i$. Since the random variable N also follows the Poisson distribution with mean value λ, then we have

$$Pr(N_1 = x_1, \cdots, N_k = x_k | N = X) = \frac{Pr(N_1 = x_1, \cdots, N_k = x_k)}{Pr(N = X)} \tag{3.15}$$

$$= \left(e^{-\lambda} \prod_{i=1}^{k} \frac{\lambda_i^{x_i}}{x_i!}\right) / \left(\frac{e^{-\lambda} \lambda^X}{X!}\right) = \frac{X!}{x_1! \cdots x_k!} \prod_{i=1}^{k} \left(\frac{\lambda_i}{\lambda}\right)^{x_i} \tag{3.16}$$

Based on the above two Lemmas, we now can establish the following Theorem regarding the evaluation of R_{avg} for an access point.

Theorem 2 *Given a WLAN access point with k mutually exclusive rate ranges δ_i ($i = 1, \ldots, k$). Suppose that the transmission rate and expected number of clients for range δ_i are r_i and λ_i, respectively. Then the average transmission rate R_{avg} of this access point can be estimated as*

$$R_{avg} = \sum_{X=1}^{\infty} \frac{e^{-\lambda} \lambda^X}{X!} \sum_{\substack{x_1, \cdots, x_k \geq 0 \\ x_1 + \cdots + x_k = X}} \left(\frac{X!}{x_1! \cdots x_k!} \prod_{i=1}^{k} \left(\frac{\lambda_i}{\lambda}\right)^{x_i}\right) \cdot \left(\frac{1}{X} \sum_{i=1}^{k} x_i \cdot r_i\right) \tag{3.17}$$

where $\lambda = \sum_{i=1}^{k} \lambda_i$.

Based on the definition of random variable N, we have:

$$R_{avg} = \sum_{X=1}^{\infty} F(R_{avg} | N = X) \cdot Pr(N = X) \tag{3.18}$$

where $F(R_{avg} | N = X)$ is the average transmission rate of the access point when there are in total X users in its coverage area. The $F(R_{avg}|N = X)$ can be further expressed as:

$$F(R_{avg}|N=X)=\sum_{\substack{x_1,\cdots,x_k\geq 0\\ x_1+\cdots+x_k=X}} \begin{array}{l} F(R_{avg}|N_1=x_1,\cdots,N_k=x_k) \\ \quad \cdot Pr(N_1=x_1,\cdots,N_k=x_k|N=X) \end{array} \tag{3.19}$$

Based on the assumption that each client, regardless of its location, has the same chance to establish a connection with the AP, then the average transmit rate R_{avg} can be easily evaluated by the following formula given that $N_1=x_1,\cdots,N_k=x_k$.

$$F(R_{avg}|N_1=x_1,\cdots,N_k=x_k)=\sum_{i=1}^{k}\frac{x_i}{x_1+\cdots+x_k}\cdot r_i=\frac{1}{X}\sum_{i=1}^{k}x_i\cdot r_i \tag{3.20}$$

Based on the Lemma 1 and (12), we can see that $F(R_{avg}|N=X)$ is given by

$$F(R_{avg}|N=X)=\sum_{\substack{x_1,\cdots,x_k\geq 0\\ x_1+\cdots+x_k=X}}\left(\frac{1}{X}\sum_{i=1}^{k}x_i\cdot r_i\right)\cdot\left(\frac{X!}{x_1!\cdots x_k!}\prod_{i=1}^{k}\left(\frac{\lambda_i}{\lambda}\right)^{x_i}\right) \tag{3.21}$$

Substituting the (3.21) into (3.18) we can see that R_{avg} is determined by (3.17).

3.5 A new scheme for access point placement

With the help of clients spatial distribution, we propose here an access point placement scheme for voice capacity enhancement.

3.5.1 Motivation

At the time of building a new WLAN, the network designers usually select the geometric center of the considered area to place the AP. As we will show in next Section that R_{avg} is heavily affected by the CSD. Notice that in practice the clients in an area may be non-uniformly distributed (please refer to Figure 3.6 for illustration), so its geometric center may be far from the squares with high clients density. Therefore, placing the AP at the geometric center of such area may significantly degrade R_{avg} and thus result in a waste of the WLAN's limited resources. It is notable, however, that some recent studies [61, 62] indicate clearly that despite the diversity of individual's mobility, the humans mobility pattern in an area (like urban) is actually predictable and stable. Based on the above observations, we propose here a new scheme for AP placement based on a careful consideration of the CSD.

3.5.2 The Proposed Scheme

The main idea of the new scheme is to place the AP as near as possible to the squares with higher clients' densities, such that average voice capacity can be enhanced as much as possible. By regarding each square's density as a mass point located at the center of the square, this AP placement problem is actually similar to the typical problem of finding the Center of Gravity (CG) for a set of point masses [63]. The CG for a collection of masses is the point where all the weight of these masses can be considered to be concentrated and this CG usually does not coincide with their intuitive geometric center. It is notable that to balance the forces that act on all masses simultaneously, the CG is naturally nearer to the points with heavier masses.

For a collection of point masses $M_1, \ldots, M_m$ located at (x_1, y_1), $\ldots$,(x_m, y_m), respectively, let (x_{cg}, y_{cg}) denote the coordinate of their CG, then the summation of all the gravitational torque created by these masses must be equal to the opposite torque at (x_{cg}, y_{cg}) in order to balance the forces that act on all masses simultaneously. Denotes by g the constant of gravity, so the torque at a point (x, y) with mass M is equal to the force $M \cdot g$ times the distance from the axis of rotation [63]. Thus, we have :

$$
\begin{aligned}
(\sum_{j=1}^{m} M_j) g x_{cg} &= \sum_{j=1}^{m} M_j g x_j \\
(\sum_{j=1}^{m} M_j) g y_{cg} &= \sum_{j=1}^{m} M_j g y_j
\end{aligned}
\tag{3.22}
$$

The above equalities imply that the torque about the origin would be the same if the entire weight acted through the center of gravity instead of acting through the individual masses. Solving the x and y coordinates of the center of gravity, we have:

$$
x_{cg} = \frac{\sum_{j=1}^{m} M_j x_j}{\sum_{j=1}^{m} M_j}, y_{cg} = \frac{\sum_{j=1}^{m} M_j y_j}{\sum_{j=1}^{m} M_j}
$$

Inspired by the above typical CG problem, we propose here the following scheme for access point placement (or equally for determining the coordinate (x_{AP}, y_{AP}) for the AP):

In next Section, the voice capacity enhancement from using the above scheme will be demonstrated.

3.6 Numerical Results

In this section, we first verify the efficiency of the new voice capacity model through simulation, then we demonstrate the voice capacity enhancement through using the new AP placement scheme.

Table 3.2: The access point placement scheme

AP Placement (x_{AP}, y_{AP})
1) Use grid to divide the WLAN area equally into squares S_j, $j = 1, \ldots, m$.
2) Acquire the expected average number of clients λ_j in square S_j.
3) Determine the coordinate (x_j, y_j) for the center of S_j.
4) Determine (x_{AP}) and (y_{AP}) as
$x_{AP} = \sum_{j=1}^{m} \lambda_j x_j / \sum_{j=1}^{m} \lambda_j$
$y_{AP} = \sum_{j=1}^{m} \lambda_j y_j / \sum_{j=1}^{m} \lambda_j$
5) Place AP at (x_{AP}, y_{AP}).

3.6.1 Simulation Settings

Our simulation is based on the topology shown in Figure 3.1. We developed simulators for DCF-based IEEE 802.11a/b WLANs with the basic parameters defined in Table 3.1 (part a)). The three clients spatial distribution (CSB) patterns considered in our simulation are illustrated in Figure 3.6, where each pattern is stored in a database in the form of a 2-D array of cells[4]. Their corresponding data of the average number of clients λ_i in each range δ_i, which are derived based on the method in section III, are summarized in Table 3.1 (part b)). The transmission rate R in IEEE 802.11a and IEEE 802.11b WLANs is stepped-down every 10 and 20 meters, respectively. With respect to the voice codec, we adopt G.711 and G.729 with a variety of payload sizes 10ms, 20ms, 30ms, and 40ms.

3.6.2 Simulation Process

Based on the above settings, the simulation was conducted as follows: Given a clients spatial distribution pattern (i.e. λ_i, $i = 1, \ldots, k$) and a specific voice codec with a constant payload size.

a) Generate the number of clients x_i for each range δ_i, $i = 1, \ldots, k$.

b) Randomly select one client to initiate a full-duplex VoIP call between him and a wired node through the AP.

c) Add the selected call to the C ongoing calls, while monitoring the QoS of the ongoing calls. If the new call results in QoS degradation of the ongoing calls, the simulation is stopped and C is reported.

d) Repeat Steps b) and c).

e) Repeat Steps a) and d).

[4]The cells are small enough such that each cell contains maximum one client.

f) Estimate the voice capacity by the average of C's from Step e).

3.6.3 The Effect of the Square Sizing

To understand how the grid's square size affect the estimation efficiency of our new voice capacity model, we conducted both simulation and analytical analysis on an IEEE 802.11a WLAN when different square size is adopted. The G.729 voice codec with 30ms payload size was used and the clients were spatially distributed in the WLAN according to Pattern 1 and Pattern 2 illustrated in Figure 3.6. We conducted our analysis by first regarding the whole WLAN's coverage area A as one big square ($m = 1$) and then we gradually divide it into multiple equal-size squares (m= 4,9,16,25,...,etc), where the analytical estimation of the voice capacity is performed for each value of m until the analytical and numerical results are matched. The corresponding results are summarized in Figure 3.3.

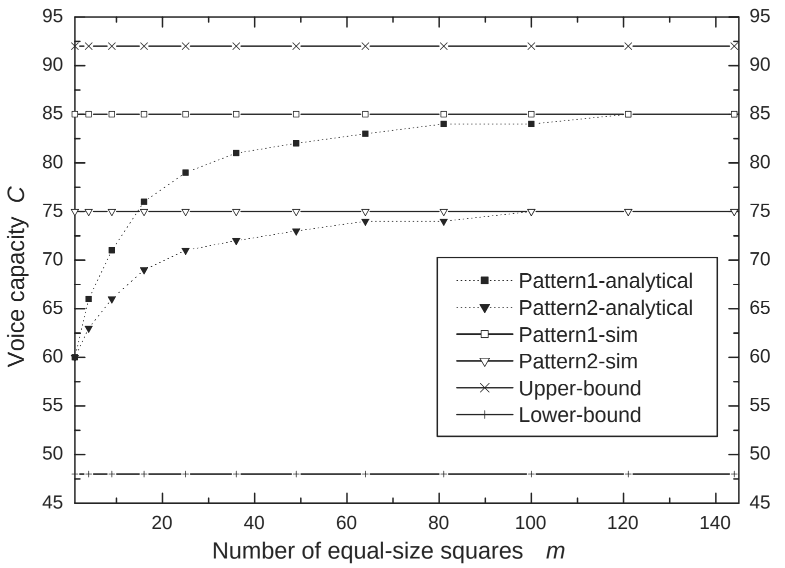


Figure 3.3: The voice capacity of 802.11a under different square size

We can easily see from Figure 3.3 that the number of squares m (or equally the square size A/m) can significantly affect the efficiency of our model for the estimation of voice capacity. For example, for the pattern 1, where the actual voice capacity is 85, the estimation of our model is enhanced from 60 to 82 as m increases from 1 to 49 (i.e., A is divided into $7 \times 7 = 49$ equal-size squares). The results in this Figure also indicate clearly that when square size A/m is small enough (or equally when the number of squares m is larger enough), the number of clients can be regarded as uniformly distributed in a square and thus our model can always result in a very efficient estimation of actual voice capacity. However, for a given estimation error of voice capacity, how to find the minimum number of required m is a complex issue since it is determined by many factors, like the overall area A, the total number of users in the WLAN, how these users are distributed in this area, etc. Nevertheless, compared to the old upper and lower bound models the estimation efficiency of voice capacity can be significantly improved by adopting our model even with a rough partition of A.

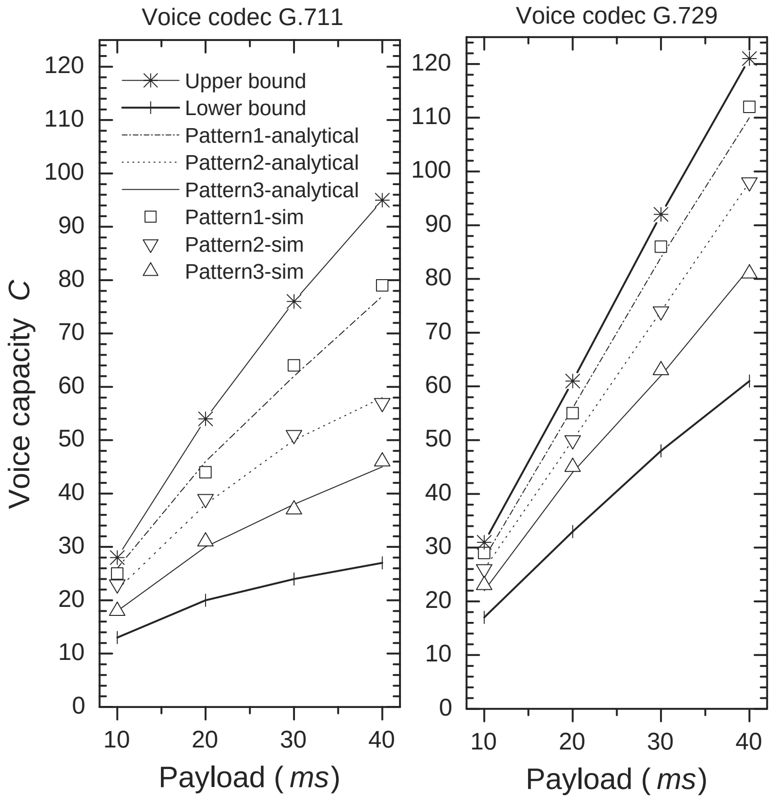


Figure 3.4: Voice capacity comparison between analytical and simulation results for 802.11a WLAN.

3.6.4 Model Verification

Figure 3.4 and Figure 3.5 show the numerical and the analytical results for IEEE 802.11a/b WLANs' voice capacity under different CSD patterns. For comparison, we also show in both Figures the results of the upper and lower bounds. The two Figures show clearly that with the help of the CSD information, our analytical model is able to provide a very efficient estimation of real voice capacity, while the corresponding estimations from the available upper and lower bounds are too far from the real case. For example, for the IEEE 802.11a WLAN with G.711 voice codec and 30 ms payload, the voice capacity's upper bound and lower bound are 76 and 26, respectively, while the numerical and analytical results under the first CSD pattern are 65 and 63, respectively. Similarly, in case of the IEEE 802.11b WLAN with G.729 voice codec and 30 ms payload, the voice capacity's upper and lower bounds are 22 and 9, respectively, while the numerical and analytical results under the third CSD pattern are 12 and 13, respectively. The above results indicate

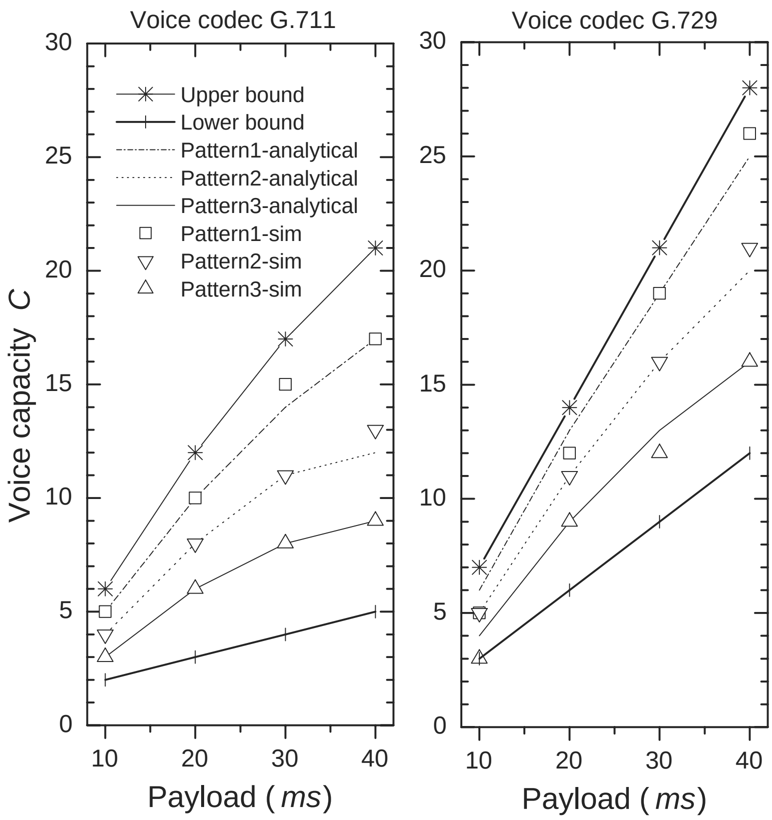


Figure 3.5: Voice capacity comparison between analytical and simulation results for 802.11b WLAN.

that the CSD can significantly affect the R_{avg}, so it should be carefully considered for an accurate estimation of actual voice capacity.

3.6.5 Voice Capacity Enhancement

To demonstrate the efficiency of the proposed scheme for AP placement, we studied the three different CSD patterns illustrated in Figure 3.6, where the accumulation of the clients density with respect to the X and Y axises is also shown. Without losing the generality, we conducted our study based on an IEEE 802.11a WLAN with G.729 voice codec and 30ms payload. The implementation results of our proposed scheme are summarized in Figure 3.6. For comparison, we also include in that Figure the corresponding results when AP is always placed at its geometric center regardless of the CSD. It is notable that the proposed scheme can always result in a voice capacity enhancement for any CSD pattern here, especially when clients are unsymmetrical distributed in the WLAN coverage area. For example, for the first and second (asymmetric) patterns, the voice capacity improvements are 40% and 18%, respectively. On the other hand, for the third (symmetric) pattern, the enhancement is not so significant. The above results indicate that the CSD should be carefully considered in AP placement, especially for WLANs that suffer from resources limitation and the clients are unsymmetrically distributed in the area under consideration.

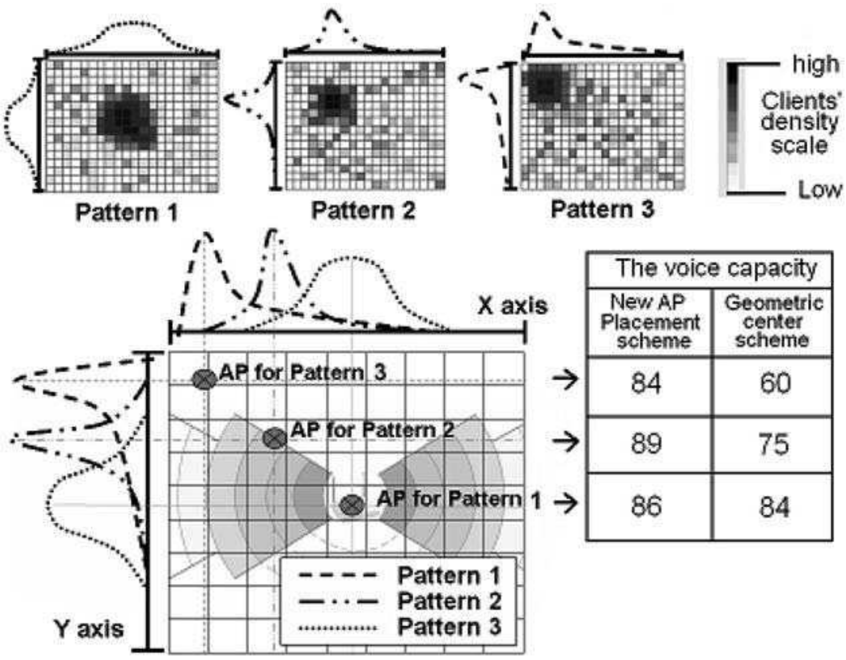


Figure 3.6: Access point placement under different clients spatial distributions.

3.7 Summary

In this chapter, we developed a new model for R_{avg} estimation and also a new scheme for WLAN access point placement with the consideration of clients' spatial distribution. We showed through both simulation and analytical studies that the proposed model can provide a very efficient estimation for WLAN voice capacity, and this capacity can be significantly enhanced if we place the access point properly by using our new placement scheme, especially when clients are unsymmetrically distributed in the WLAN area. It is expected that the work in this chapter will contribute to the network planning and protocol design of future VoIP over WLANs.

Chapter 4

An Enhanced Bidirectional-Transmission Protocol

4.1 Introduction

The available bidirectional transmission (BDT) MAC protocol promises to eliminate the VoIP traffic downlink bottleneck at the WLAN's AP by allowing the each transmission session to afford two packets delivery, from sender to receiver and then from receiver to sender in it has a packet to send at such instance. Nevertheless, the current first-in-first-out scheduler adopted in the WLANs' AP does not support the BDT protocol to fulfill its promises. This is because, in the AP, the scheduling of the packet for transmission does not consider whether the receivers have also a packet to send-back at such instance or not. In this chapter, we enhance the BDT protocol through a novel packet scheduler and a new contention mechanism. In the proposed scheduler, we used a probabilistic-based scheme to help the AP to schedule its packets according to the higher BDT chance first out rule, while the new contention mechanism is proposed to provide the highly expected BDT sessions with the highest channel access priority. We demonstrate through both analytical analysis and computer simulations that the proposed modifications can significantly enhance the performance of the BDT protocol in terms of throughput, voice capacity. We also show that the enhanced BDT protocol become able to provide a better support to the VoIP applications in the WLANs even under the coexistence of Best-effort traffic.

4.2 Bidirectional Transmission MAC Protocol

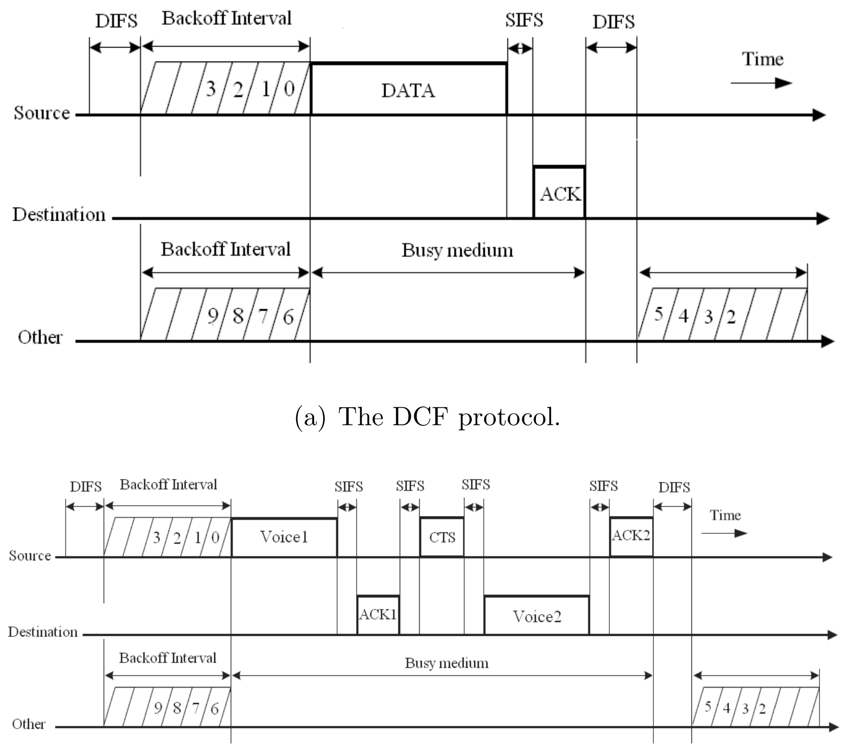
VoIP is an established and rapidly developing technology. It has been proved to be less expensive to install and maintenance in compare with the traditional public switched telephone network (PSTN). Recently the Wireless local area network (WLAN) is also growing dramatically and become a ubiquitous networking technology around the world. Thus, supporting voice over Wireless Local Area Network (VoWLAN) has drawn extensive attentions for fulfilling the last-mile wireless coverage of IP networks with mobility support. However, VoWLAN still poses many challenges since the WLANs were not originally designed for supporting delay-sensitive voice traffic [64, 65, 66].

The basic access method in the IEEE 802.11 medium access control (MAC) is the distributed co-ordination function (DCF). In that protocol, the channel access involves a long unbiased competition among all the active stations[1] including the WLAN's AP that may be even ended with a collision. However, once a successful transmission session is initiated, although it is established between two nodes, the sender node is the only one who is allowed to transmit its packet while the receiver node can only acknowledge the reception of such packet as illustrated in Fig. 4.1(a). Based on the fact that the VoIP traffic is divided half-by-half upon the downlink (AP-stations) and uplink (stations-AP), and because of such unbiased channel access opportunity, the current DCF protocol usually cause a VoIP traffic downlink bottleneck at the AP and hence restrain the VoIP applications over the WLANs [67, 68]. In general, this protocol can be efficient for best effort applications with big size packets and low sampling rate[2], but for a delay sensitive traffic like VoIP with very small packet size and very fast sampling rate (e.g., 50 packets per second for a payload of 20ms per packet) , is ultimately a bottle-neck.

Recently a bidirectional transmission-based MAC protocols were proposed to enhance the performance of reliable transporter over WLAN [69, 1, 70]. The BDT protocol main idea is to allow the receiver node to send back a packet to the sender in the same transmission session just after the acknowledging the reception of sender's packet as illustrated in Fig. 4.1(b). Such simple idea promises to eliminate the VoIP problems because of two main reasons. First, since the WLAN's AP must take part of any transmission session, so its chance to transmit its buffered packets will be multiplied by the number of the active stations. Second, since any voice call is simply a two-way voice streams between AP and the mobile station, so the bidirectional transmission mechanism will decrease the packet loss and collision probability. Nevertheless, the simple first-in-first-out packet scheduler adopted at the network AP does not support the above protocol to fulfill its promises since the selection of the packet to be served does not consider whether the receiver node has also a packet to send at such instance or not. Hence, such careless selection of the packet to be served may result in the cost of losing a lot of BDT chances. Furthermore, based on a careful observation of the expected performance of the BDT protocol, the number of competition instances will be significantly reduced because each contention instance will

[1]Active station means that it has a packet to send at that instance

[2]the number of generated packets per second



(a) The DCF protocol.

(b) The BDT protocol.

Figure 4.1: The Illustration of the DCF and BDT protocols.

hold two packets not only one as before. Therefore, the contention mechanism adopted in the current BDT protocol, which is typically inherited from DCF protocol, must be modified in order to take the full advantage of the new enhanced BDT protocol.

In this chapter, we enhance the BDT protocol by introducing a novel packet scheduler and a new contention mechanism. The novel packet scheduler uses a probabilistic-based scheme to help the AP in selecting the packet and hence its receiver station with the highest probability of possessing a packet ready to send at that instance. The new contention mechanism, on the other side, is proposed to provide highly expected BDT chances with highest priority for channel access in order to get the full advantages of proposed scheduler and increase the WLAN voice capacity.

In summary, the main contributions of this work are as follows

1. We propose a novel packet scheduler for the Bidirectional Transmission-based MAC protocol in order to enhance its performance and hence increase the voice capacity in the current IEEE 802.11 WLANs.

2. We proposed a new contention mechanism for the same protocol to provide the highly expected BDT chances with the highest priority for channel access in order to get the full advantages of proposed scheduler

3. We develop analytical models to estimate the throughput and voice capacity maximum gain, that can obtain because of the proposed packet scheduler.

4. We demonstrate that although the available protocol can slightly improve the voice capacity, this improvement can be much more significant if our proposed modifications are adopted even under the coexistence of best-effort traffic.

The rest of chapter is organized as follows. In Section II, we review the Bidirectional Transmission-based protocol and the related works. In Section III, we introduce our novel packet scheduler and the new contention mechanism. Analytical models of throughput and voice capacity are presented in Section IV. Section V includes the simulations results, and finally we conclude this chapter in Section VI.

4.3 Background and Related Work

In this section, we introduce the bidirectional transmission approach background, protocol and limitations.

4.3.1 The Bidirectional Transmission approach background

Bi-directional transmission approach in single transmission session has been previously investigated in literature [69, 1, 70]. The concept was first introduced in [69], where the receiver is allowed to append one of his packets to the traditional acknowledgment packet and transmit it on the same transmission session, regardless of its destination address. However, this protocol requires that all stations must be always ready to receive packets at any time, which is not backward compatible with the original protocol.

Wu et al [1] proposed 'DCF+' to restrict the bidirectional packets transmission to be only between the sender and the receiver node. The receiver appropriately sets the duration field of the ACK packet if it holds a data packet for the sender. In the sender side, the clear-to-send (CTS) packet is employed to reserve the channel for reverse data transmission, and is sent in response to the ACK. Later in [70], the author further proposed to implement the bi-direction transmission approach at AP only. he also proposes that AP can directly send a Self-Clear-To-Send (Self-CTS) packet to reserve the channel if it has a packet to send. However, such modification did not result in any enhancement in terms of voice capacity or throughput. In the rest of the chapter we will mainly focus on the 'DCF+'as the basic Bidirectional Transmission-Based Protocol that we target to enhance.

4.3.2 The Bidirectional Transmission-based protocol

As illustrated in Fig. 4.1(b), the DCF+ protocol (i.e., the BDT protocol) works as follow: all stations with packets ready for transmission observe the shared medium before attempting to transmit. If the medium is sensed busy, the station delay transmission until the medium is sensed idle for a period of time equal to a DCF InterFrame Space (DIFS). After a DIFS medium idle time, the station enters the backoff phase in which it sets a backoff counter randomly chosen from [0, CW_{min}], where CW_{min} is the minimum contention window size. The backoff counter decreases by one for every idle slot and freezes if the channel is busy. The decrement procedure resumes after the channel is sensed idle again for a DIFS. The station transmits the packet when the backoff counter reaches zero. If there is no acknowledgment (Ack) received due to collision or transmission errors, the contention window size doubles after each unsuccessful transmission trial until it reaches the maximum value (CW_{max}), and the sender reschedules the transmission according to the aforementioned backoff rule. The frame is dropped when the retransmission limit is reached. Given that the frame is received successfully, if the receiver holds a data packet for the sender at such instance, it appropriately sets the duration field of the ACK packet and transmits Ack following a Short InterFrame Space (SIFS), otherwise the receiver just send a regular Ack packet. On the sender side, a clear-to-send (CTS) packet is employed to reserve the channel for reverse data transmission, and is sent in response to the aforementioned modified Ack. finally, the receiver starts to transmit its packet and receives AcK from the original sender similarly as before. After every successful transmission, contention window size is reset to its initial (minimal) value.

4.3.3 The Bidirectional Transmission-based protocol limitations

Since the main problem of the VoIP applications is the congestion of the voice packets at the AP due to the long waiting for a successful channel access chance. The above protocol promises to eliminate such bottleneck as it offers the AP the chance to transmit a packet at each transmission session.

Although the workload of the AP is almost equal to the aggregated workload of all the other wireless stations in the WLAN, the AP is not supported by any priority to access the wireless channel, this is why many packets are accumulated (buffered) in the AP waiting to be transmitted based on the first-in-first-out scheduling scheme. Such simple scheduler results in the cost of continuous waste of a successful bidirectional transmission chances because in that case the AP does not care whether the receiver station, to which the selected packet in the AP will be transmitted, has also a packet to send at that instance or not. Actually, such continuous missing of bidirectional transmission chances will turn the overall bidirectional transmission protocol performance to be as similar as the old

DCF-based one, which suffers from a severe wasting of resources.

Nevertheless, the simple first-in-first-out scheduler adopted at the WLANs' AP does not support the above protocol to fulfill its promises since the selection of the packet to be served does not consider whether the receiver node has also a packet to send at such instance or not. Therefore, a non careful consideration of such scheduling issue at AP will be always in the cost of losing many Bidirectional Transmission chances[3] and hence turn the performance of such protocol to be as similar as the traditional DCF protocol, which suffers from a severe wasting of resources.

Furthermore, the backoff mechanism adopted in the above protocol was typical inherited from the DCF-based protocol despite of the radical differences between them in terms of the packet exchanging method. Based on a careful observation of the expected performance of the BDT protocol, the number of competition instances will be significantly reduced because each contention instance will be followed by the transmission of two packets not only one as before. In addition, a highly expected BDT chances must receive a higher priority for channel access. Therefore the contention mechanism adopted in the above protocol must be accordingly modified in order to take the full advantage of such protocol.

4.4 The Enhanced Bidirectional Transmission-Based Protocol

In this section, we present our proposed enhancement to the bidirectional transmission based protocol. We firstly introduce the novel packet scheduler scheme and then the proposed contention mechanism.

[3]bidirectional transmission chance means that both sender and receiver have a packet to transmit to each other in the transmission session

Table 4.1: The proposed packet scheduler

We use a list of active calls so called "ActiveList", with standard operations like InsertActiveList(), which adds new call to the ActiveList. UpdateActiveList(i) which update the existing call entity i at the arrival of any of its packets.

```
Once the AP receive any packet p
I=ExtractCall(p);
If(ExistInActiveList(i)==FALSE) then
   InsertActiveList(i);
Else
   UpdateActiveList(i);

Once the AP access the channel to send a packet
N=CountPackInAP(); /*count the buffered packets in AP*/
For (j=0; j<N; j++)
   i=ExtractCall(p_j);
   If(ExistInActiveList(i)==FALSE) then
      P_j.priority = 1;
   Else
      P_j.priority = Eqation1(i);
return(arg(max_{j=1}^{N}(P_j)))

Each Call Entity (i) in the ActiveList is as follows:
   i.SenderIP;
   i.ReceiverIP;
   i.PacketCounter; /*the number of received packets*/
   i.LastArrival; /*the arrival time of the last packet*/
   i.InterArrival; /*the last inter-arrival period*/
   i.AvgIntArrTime; /*average inter-arrival period*/

void UpdateActiveList(i)
{
   int n = i.PacketCounter;
   i.InterArrival = CurrentTime - i.LastArrival;
   i.LastArrival = CurrentTime;
   i.AvgIntArrTime = (i.InterArrival + (n*i.AvgIntArrTime)) / (n+1);
   i.PacketCounter += 1;
}

Float Equation1(CallID)
{
   return ((CurrentTime - CallID.Last arrival) / CallID.AvgIntArrTime)
}
```

4.4.1 The proposed packet scheduler

The main target of the proposed scheduler is this chapter is to overcome the aforementioned limitation of the BDT protocol by helping the AP to select the packet and hence the receiver station with the highest probability of possessing a packet ready to send at such instance, thus a complete bidirectional transmission can be achieved. The main idea of our scheduler is based on the fact that any voice codec uses a constant packet generating rate during the voice call (e.g., one packet every 20ms), so if the AP keep-track of the ongoing calls by recording their arrival times and inter-arrival periods in a look-up table, it will be able to estimate the generation time of the next packets in their sources (i.e., their original source stations) using a very simple recursive formula. So based on this idea, once the AP got the chance to access a priority value is instantly computed for each buffered packets inside the AP with the help of a look-up. Such priority value will help the AP to decide which packet to be send at that instance so that a bidirectional transmission can be achieved with a high probability. Next we will explain our scheduler in details.

As shown in Table 4.1, in the proposed scheduler, once the AP receive any voice packet from any station, it first check whether such packet belong to an ongoing active call or not. If not, it creates a new entity for such new call and store its source and destination IP addresses in addition to its arrival time. Otherwise, if the received packets belong to an ongoing voice call, it updates the arrival time, inter-arrival period, packets' counter and average packets' inter-arrival periods of the corresponding call entity in the ActiveList. As we can show in the update function, the average packets' inter-arrival periods is always updated at the arrival of any packet from the other-side station in order to provide an accurate estimation of the priority value when it is computed in Equation1.

On the other side, when the AP got the chance to transmit a packet, a priority value is computed for all the packets in the AP at that instance, which reflects the possibility that the corresponding receiver node has a packet ready to send at that instance. The priority value is computed using Equation1 which result in a value greater than 1 if the corresponding receiver has a packet ready to send at that instance. Such priority value mainly depends on the previously collected information stored in the ActiveList. For a while if there is no information about such packet in the ActiveList due to any reason, such packet will be treated with the priority value equal to one, which for instance reflects a high priority.

Obviously, if the AP start to consider whether the receiver node has also a packet to send or not using the proposed packet scheduler, the possibility of achieving a complete bidirectional transmission will increase. Concurrently, the collision and packet loss probability will decrease as well. This is because once the AP sends its packet to a station which may strongly posses a packet ready to send, it will eliminate such station to participate in the channel access competition in the following rounds since it will be able to send its packet in the same transmission session initiated by the AP according to the bidirectional transmission protocol.

4.4.2 The proposed contention mechanism

As we explained above, the use of the proposed scheduler will decrease the number of contention periods. Therefore, adopting the old contention mechanism will not be convenient anymore, contrarily, it will decrease the throughput and the channel utilization's efficiency. This is because the old contention mechanism was originally designed to provide an equal-chance of channel access to all the wireless stations including AP using a relatively big CW_{min} to decrease the possibility that two stations select the same random CW and hence collied.

In our proposed contention mechanism, the AP is supported with the highest priority for channel access only and only if it has a packet with a priority value greater than one. Such value indicates that one of the buffered packet in the AP has a corresponding receiver node that also hold a packet waiting to be sent at such instance. In this case, once the current transmission session is finished, the AP does not wait for a complete DIFS period to be elapsed, it directly start a new transmission session just after an SIFS period (i.e., it does not give the chance for a complete DIFS period to elapse because SIFS is less than DIFS). By this way, the currently frozen stations will not be able to resume their contention and they will be kept frozen to the end of the new initiated transmission session initiated by the AP.

Moreover, with respect to the value of the CW_{win}, we propose to reduced it to the half of it original value. This reduction is based on a similar modification adopted in the Enhanced DCF protocol [71] , which reduce the CW_{win} to the half of it original value for the real-time traffic only based on the fact that such reduction will give the real-time traffic a kind of priority among the other traffic. In fact, adopting such reduction in our protocol is more reasonable than the EDCF case because the actual number of contention period will be decrease by half and hence reducing the CW_{min} will be convenient if we sake to get the maximum advantage of the BDT protocol.

4.5 Network Throughput Analysis

In this section, we present models to estimate the WLAN's gain in terms of the throughput and voice capacity due to the adoption of the proposed scheduler. We first evaluate the maximum channel throughput U_{max} that can be achieved in an AP of a usual DCF protocol. Using the same derivation method mentioned in [64, 16, 17], the U_{max} can be modeled as follows

$$U_{max}=\frac{T_P \times R}{T_{voice}+T_{Ack}+T_{SIFS}+T_{DIFS}+T_{Backoff}} \tag{4.1}$$

The above equation indicates that the channel throughput is a function of the AP's transmission rate R and other parameters, like the time T_P needed to transmit the voice payload, the time T_{voice} needed to transmit the whole voice packet, the SIFS time T_{SIFS}, the DIFS time T_{DIFS}, the time T_{Ack} needed to transmit the acknowledgment packet, and the backoff time overhead $T_{Backoff}$. The $T_{Backoff}$ can be evaluated as follows based on the assumption that there is always two active stations (one is the AP and the other is one of the mobile stations) [16]

$$T_{backoff}=\begin{cases} 4.5\times 9+T_w\times 0.06\,\mu s & \text{for the 802.11a} \\ 8.5\times 20+T_w\times 0.03\,\mu s & \text{for the 802.11b} \end{cases} \tag{4.2}$$

where T_w is defined in terms of some basic parameters as

$$T_w = T_{SIFS} + T_{DIFS} + T_{voice} + T_{Ack} \tag{4.3}$$

By adopting the proposed scheduler, the maximum throughput can be achieved if and only if each transmission session can hold a bidirectional transmission (i.e., the receiver has a packet to send) . In this case half of the voice traffic will be transmitted without any contention. Thus, the the maximum channel throughput U_{max} that can be achieved is such case can be expressed as follows

$$U_{max}=\frac{2\cdot T_P \times R}{2\cdot T_{voice}+2\cdot T_{Ack}+4\cdot T_{SIFS}+T_{DIFS}+T_{Backoff}+T_{CTS}} \tag{4.4}$$

Based on the above channel throughput model, the voice capacity C of the AP (i.e., the number of VoIP connections the AP can simultaneously support) can be expressed as follow:

$$C = \left\lfloor \frac{U_{max}}{2L} \right\rfloor \tag{4.5}$$

where all voice calls are assumed to use the same voice codec with bit rate L bps.

Table 4.2: The IEEE 802.11(a/b) DCF-based WLAN Parameters

		802.11b	802.11a
Available Transmission Rates (R)		11 Mbps	54 Mbps
Slot Time		20μs	9μs
SIFS		10μs	16μs
DIFS		50μs	34μs
CW_{min}		32	16
CW_{max}		1024	1024
Retry Limit		7	7
T_{voice}	PLCP & Preamble	192μs	24μs
	MAC Header + FCS	24.7μs	5μs
	RTP/UDP/IP Header	29.1μs	6μs
	Voice Payload (T_v)	(payload $\times 8/R)\mu$s	
T_{Ack}	PLCP & Preamble	192μs	24μs
	Ack Frame	10.2μs	2.1μs

The typical values for above basic parameters are shown in the Table 4.2 when the 802.11(a/b) standards are considered.

Table 4.3: The WLAN's throughput and voice capacity gain

WLANS	IEEE 802.11b			IEEE 802.11a		
	before	after	gain	before	after	gain
$U_{max}(Mbps)$	0.226	0.264	16%	0.991	1.2	21%
C	14	16	15%	64	75	18%

We use Equations (1-3) to compute the maximum throughput and voice calls gain that can be obtained in a single IEEE (802.11b/802.11a) WLAN. These calculations are done using the famous voice codec G729 with 20 ms of audio data in 20 bytes of payload, without any compression. A single G729 VoIP stream therefore constitutes 8Kbps. Table 4.3 tabulates the maximum throughput and voice capacity gain that can be achieved before and after the adoption of the proposed scheduler. In these calculations we assumed that each transmission is performed with the maximum possible transmission rate. Here, we have to mention that the results listed in Table 4.3 reflect only the expected enhancement due to the adoption of the proposed scheduler. In the next section we will show that this enhancement will be much more significant if the proposed scheduler is jointly combined with the proposed contention mechanism.

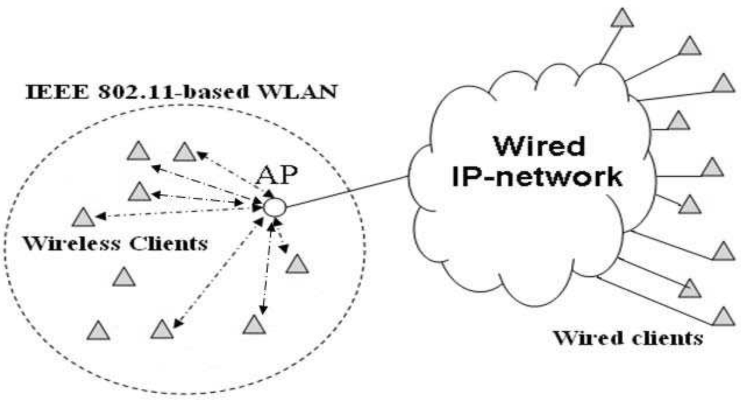


Figure 4.2: Network topology for simulations

4.6 Simulation Results

In this section we introduce the simulation analysis for the proposed scheduler and contention mechanism.

4.6.1 Simulation setting and scenario

Our simulation is based on the topology shown in Figure 4.2. We developed simulators for both IEEE 802.11b and IEEE 802.11a with BDT-based protocol. The basic parameters used in our simulation are summarized in Table 4.2. For voice traffic, the standard voice codec G.729 is considered with payload size 20 bytes per packet (i.e., 20 ms). For performance comparisons, we also developed the above simulators for DCF and DCF+ protocols too where the DCF+ stand for the available BDT-based protocol). In our simulation results we will refer to our enhanced BDT-based protocol by DCFnew.

In our experiment, n wireless client stations were associated with an AP. This setting was used to make full-duplex VoIP calls between wireless client stations and wired nodes in the wired network counterpart through the AP. We are interested in only connections between the AP and the wireless stations. It is assumed that two voice connections of each voice call are independent of each other. Through the chapter, all the simulation results are average over 100 simulations each for 200 seconds length and with a random starting time for each call.

4.6.2 Voice capacity

First we investigate the enhancement of the voice capacity. In general the voice capacity of any WLAN is inspected based on the voice packets delay of the ongoing voice calls in such WLAN. A voice call is added to the ongoing calls, and once the WLAN exceeds it actual capacity the voice packets delay is suddenly increased. For instance, as shown in Fig4.3, the IEEE 802.11b WLAN can support only 14 simultaneous G.729 voice calls, while the IEEE 802.11b WLAN can support 66 G.729 voice calls. We can also notice that the voice capacity has been meaningfully increased in comparison with the original DCF protocol and also in comparison with the DCF+ protocol (i.e., the regular BDT-based protocol). For example the voice capacity is increase from 14 to 17 in the IEEE 802.11b WLAN and from from 66 to 83 in the IEEE 802.11a WLAN due to the implementation of the proposed scheduler at the AP in addition to the adoption of the proposed contention mechanism.

4.6.3 Throughput

We here seek to find the maximum throughput that can be achieved by the proposed enhanced protocol and then compare it with the available ones. In this simulation we add the voice calls one by one while monitoring the network throughput only. We notice that the throughput increase gradually up to a certain limit, which we defined as H_{max} and then it start to decrease again due to the increase of contention between nodes and the increase of packet collision probability. As illustrated in Fig. 4.4, we can easily notice the throughput enhancement achieved by the DCF+ in compare to the traditional DCF. However, DCFnew still report the best performance among all the schemes in terms of throughput thanks to the efficient packet scheduler and contention mechanism adopted in the DCFnew protocol.

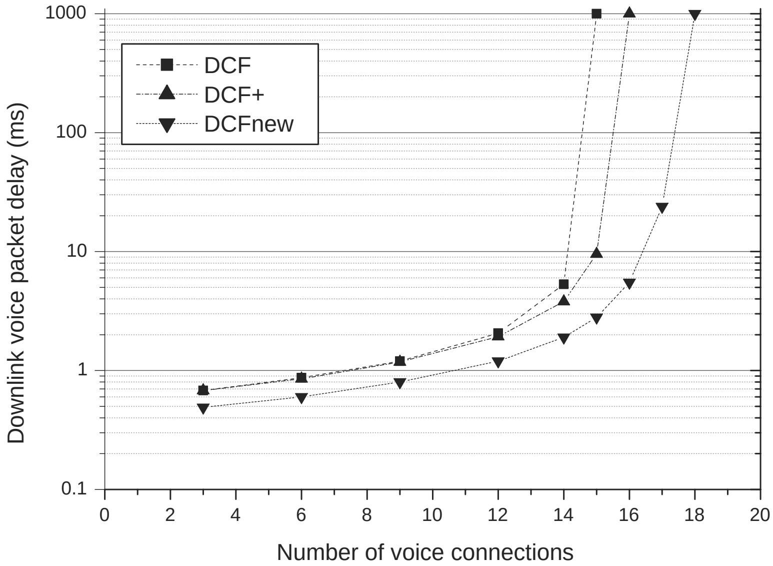
4.6.4 VoIP support under coexistence of Best-Effort traffic

One of the main objective of our protocol is to provide efficient support the VoIP traffic even under the coexistence of best-effort (BE) traffic. Therefore, we here create a new simulation scenario to evaluate their performance under such realistic condition.

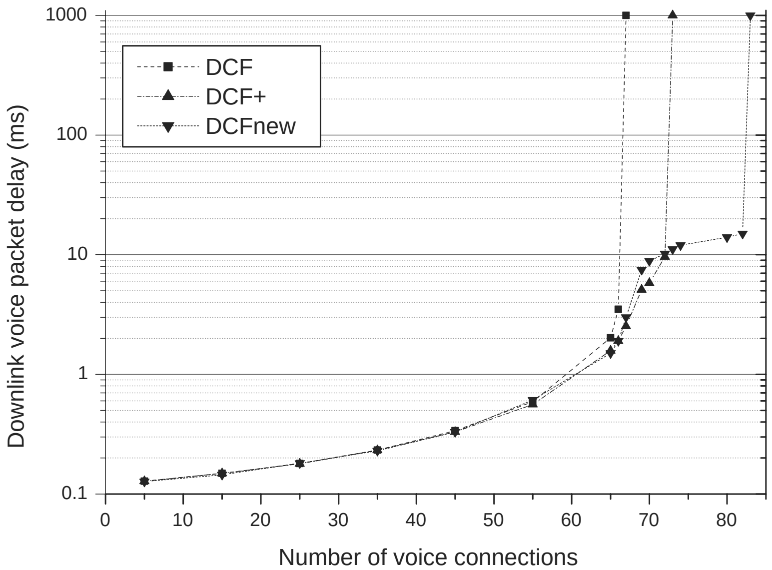
Since in most of the cases, data traffic is heavily put on the downlink direction [15] for downloading purpose, in our simulation we create a BE flow which flows from the AP to one of the wireless stations. Concurrently, we report the average voice packet delay and the corespondent voice capacity under the gradual increases of the BE flows number. Each BE flow is generated at the form of constant bit rate (CBR) with packet size of 440 bytes[4], and its load is set to 1Mbps.

As illustrated in Fig. 4.5, we can easily notice that the coexistence of BE flows reduces the voice capacity under the DCF-based protocol. For example, the voice capacities in the DCF-based IEEE 802.11b WLANs is reduced from 12 to 8 and then to 4, due to the coexistence of one and then two BE flows, respectively. Although the voice capacity is

[4]Since the packet size of 60%, 20%, and 20% of the BE traffic is 44, 550, and 1500 bytes [72], respectively, then the packet size of the BE traffic in our simulation is settled to the average size.

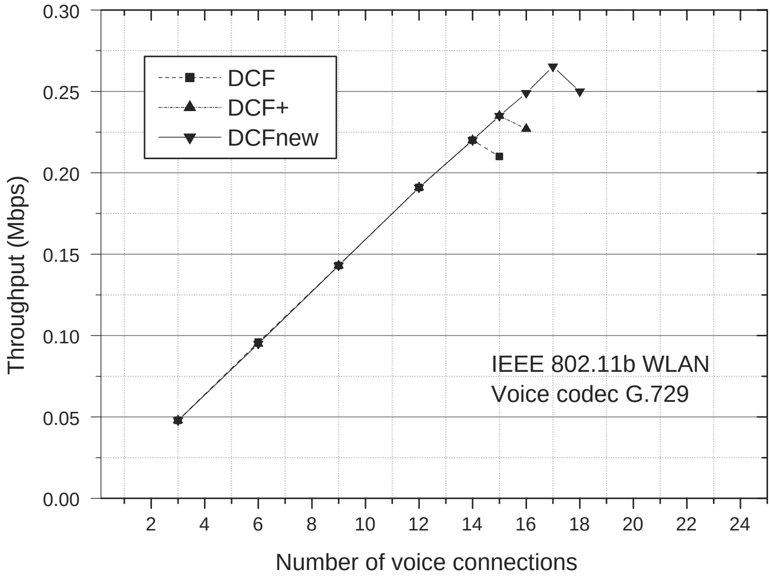


(a) IEEE 802.11b WLAN / voice codec G.729.

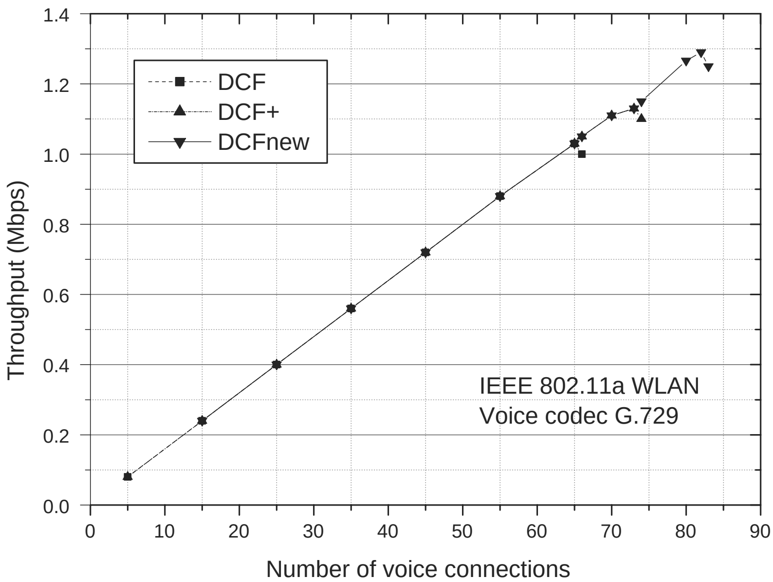


(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 4.3: The voice packet delays under DCF, DCF+, and DCFnew protocols.

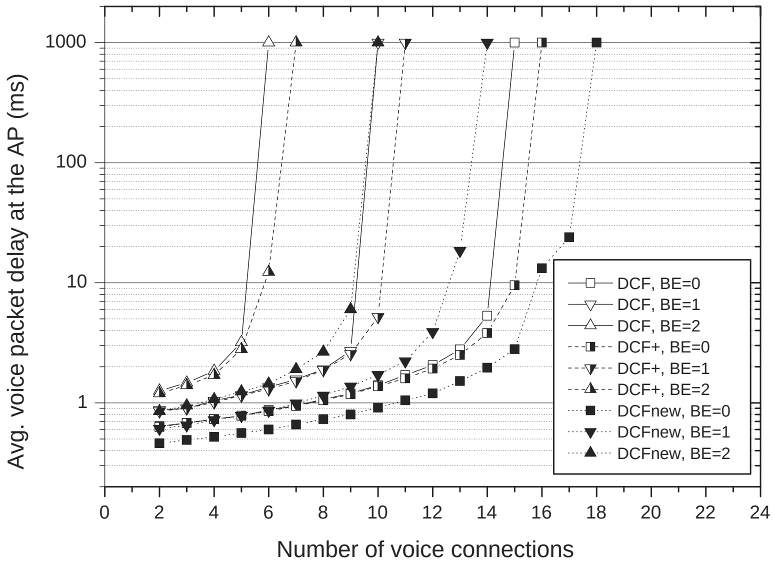


(a) IEEE 802.11b WLAN / voice codec G.729.

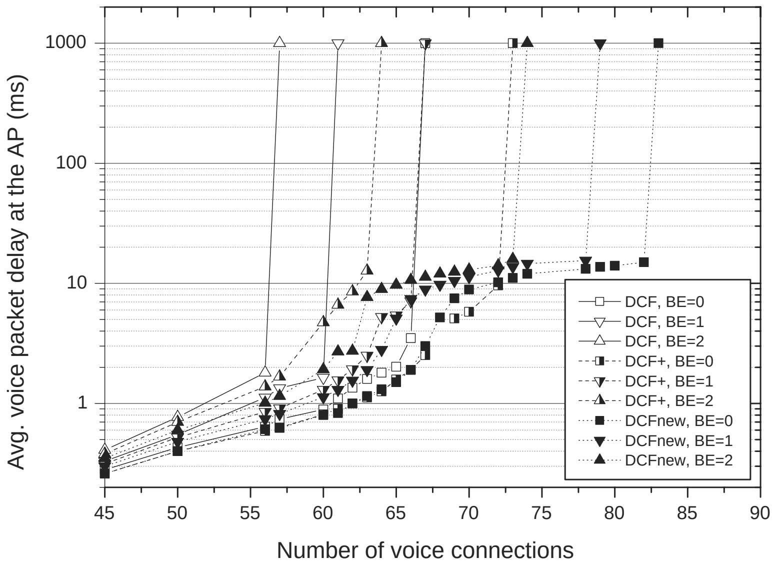


(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 4.4: The WLAN's throughput under the DCF, DCF+, and DCFnew protocls.



(a) IEEE 802.11b WLAN / voice codec G.729.



(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 4.5: The WLAN's voice capacity under the coexistence of best-effort traffic using the DCF, DCF+ and DCFnew protocols.

also slightly reduced in the DCFnew protocol, it still report a significant support to the VoIP traffic in comparison with the DCF+ protocol. The results also indicate that the voice capacity in the DCFnew-based IEEE 802.11b WLAN is only reduced from 17 to 14 and then to 10, due to the coexistence of one and then two BE flows, respectively. On the IEEE 802.11a WLAN side, the DCFnew protocol can support up to 74 G.729 calls under the coexistence of two BE flows, while the DCF+ protocol can only support 73 G.729 calls even without any BE flows.

4.7 Conclusion

In this chapter, we focus on the bidirectional transmission-based protocol which promise to eliminate the downlink bottleneck of the VoIP traffic. We proposed a novel packet scheduler to maximize the chance bidirectional transmission in each transmission session. We then modify the old inherited DCF-based contention mechanism in order to achieve the maximum benefit of the proposed scheduler.

Extensive simulation results demonstrated that the available bidirectional transmission protocol can slightly improve the WLAN voice capacity as well as the network throughput, but this improvement can be much more significant when the proposed packet scheduler and contention mechanism is jointly adopted with it. For example, in an IEEE 802.11a WLAN/G.729 voice codec with $20ms$ payload size, the available bidirectional approach can slightly improve the voice capacity from 66 to 73 while this improvement can be increased to 83 with the help of our proposed modifications.

The results in this chapter also indicate clearly that voice capacity is severely degradated with the coexistence of the best-effort traffic, while the proposed enhanced bidirectional transmission protocol can still guarantee a very efficient support to the VoIP traffic under the same condition.

Chapter 5

A Novel MAC Protocol for VoIP support over WLANs

5.1 Introduction

The current IEEE 802.11 MAC protocol is not suitable for efficient support of VoIP, because of both its downlink bottleneck and also the too large packet overhead required by the ACK mechanism even in the transmission of a small packet. Based on the observations that voice packets are of small size, bi-directional and can tolerant certain of packet loss, we propose a new MAC for efficient support of VoIP over WLANs. The main idea of the new MAC is to remove the downlink bottleneck by adopting the bidirectional instead of unidirectional transmission mechanism, and then enhancing the channel utilization efficiency by compacting the packets exchanging processes into a fewer number of steps. We further extend this protocol to the WLANs that may suffer from non-ideal channel conditions. We demonstrate through both analytical analysis and computer simulations that the new MAC can not only increase the voice capacity and throughput dramatically but also guarantee a much better fairness even under the coexistence of best effort traffic.

5.2 Medium Access Control Protocol

The basic access method in the IEEE 802.11 medium access control (MAC) is the distributed co-ordination function (DCF). It is basically based on CSMA/CA (carrier sense multiple access/collision avoidance) principle, where channel access is gained by equal competition among the single CSMA instance of downlink (AP-to-station) and multiple CSMA instances of uplink (stations-to-AP) despite of their huge workload difference. This asymmetric behavior leads to unfairness in terms of throughput and delay, and thus significantly affects the bi-directional voice traffic that is usually divided half-by-half between the downlink and uplink. Ultimately, downlink voice connections become a bottle-neck to restrain the voice capacity of IEEE 802.11 WLANs [73, 74, 75]. Moreover, based on a careful observation of DCF, we can easily see that even the transmission of a small voice packet involves a huge overhead. For example, it takes 550 μs to transmit an 60-byte voice packet over an IEEE 802.11b WLAN (11Mbps), among which only less than 50 μs is used for the transmission of voice packet itself while as long as 500 μs is used for overhead (almost half of that is for acknowledgment only)[76, 77]. Thus, the current DCF does not really provide an efficient support of VoIP.

A number of previous studies are available on how to increase the voice capacity in the IEEE 802.11 WLANs by alleviating the unfairness between downlink and uplink [78]. In [67, 79], the packet bursting technique is implemented to downlink flows where a group of packets are aggregated into one burst and sent together at a time. However, large burst size that matches with the ongoing voice calls may have an adverse impact on other real-time traffic like video applications. In [80, 81, 82], direction-based prioritization of channel access has been proposed, where different EDCA parameters are applied to different traffic direction and traffic types. Although this approach can alleviate the unfairness by assigning higher priority to downlink traffic for channel access, it is difficult to implement and may not always lead to a higher voice capacity. Our work is most closely related to the technique introduced by Tourrilhes [69], which was later enhanced in [1] and [70]. In these works, the authors introduce the bidirectional packet transmission mechanism in a single channel access. Based on this mechanism, the receiver station is allowed to send a data packet in the same CSMA instance through a specific packets exchange scenario. Although this transmission mechanism can significantly improve the fairness between the downlink and uplink, it only results a slight improvement in WLAN voice capacity.

In this chapter, we propose a novel medium access protocol for supporting VoIP over 802.11 WLAN, so called DCFvs (DCF voice support). The basic idea of our protocol is to first remove the downlink bottleneck through adopting the bidirectional instead of unidirectional transmission mechanism, and then enhance the channel utilization efficiency by accomplishing the same packets exchanging processes but in a fewer number of steps.

In summary, the main contributions of this work are as follows:

1. We propose a new medium access protocol (DCFvs) to dramatically increase the voice capacity in the current IEEE 802.11 WLANs with ideal wireless channel condition.

2. We further extend DCFvs to a secured version (DCFsvs) suitable for WLANs that suffer from non-ideal channel condition.

3. We develop analytical models for estimating both the throughput and voice capacity

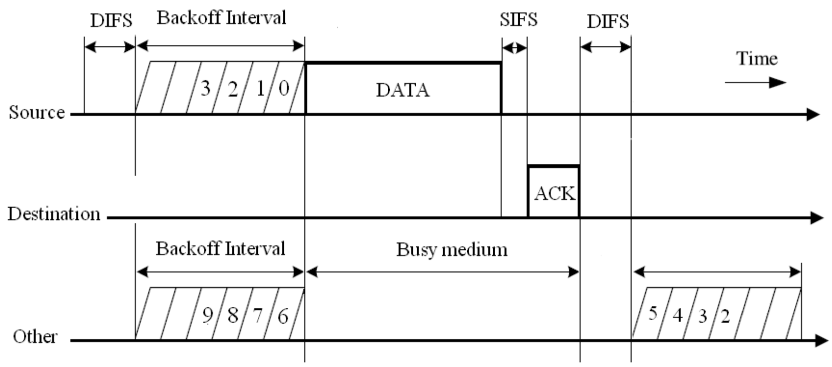


Figure 5.1: IEEE 802.11 DCF basic access scheme.

of the proposed protocols.

4. We demonstrate that although the available works can slightly improve the voice capacity, this improvement can be much more significant if our proposed transmission mechanism is adopted even under the coexistence of best-effort traffic.

The chapter is organized as follows. In Section II, we review the IEEE 802.11 MAC DCF-based protocol, the VoWLAN basics and the related works. In Section III, we introduce our proposed DCFvs/DCFsvs Protocols and also the related implementation issues. Analytical models of throughput and voice capacity are presented in Section IV. Section V includes the simulations results, and finally we conclude this chapter in Section VI.

5.3 Background and Related Work

In this section, we introduce the IEEE 802.11 MAC DCF-based WLAN, the VoWLAN basics and also the related works on the MAC enhancement for VoWLAN support.

5.3.1 IEEE 802.11 DCF-based WLAN

There is a variety of standards defined in the IEEE 802.11 family, like 802.11a and 802.11b, which can provide up to 54Mbps and 11Mbps raw data rate, respectively. The basic access method in the IEEE 802.11 medium access control (MAC) is the distributed co-ordination function (DCF), which is based on carrier sense multiple accesses with collision avoidance (CSMA/CA).

As illustrated in Fig. 5.1, the DCF works as follow: all stations with packets ready for transmission observe the shared medium before attempting to transmit. If the medium is sensed busy, the station delay transmission until the medium is sensed idle for a period

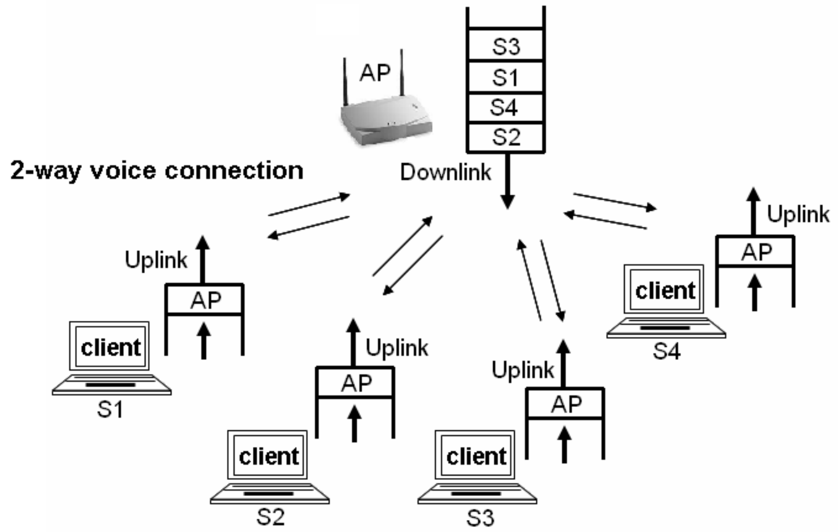


Figure 5.2: Unfairness between downlink and uplink traffic.

of time equal to a DCF InterFrame Space (DIFS). After a DIFS medium idle time, the station enters the backoff phase in which it sets a backoff counter randomly chosen from [0, CW_{min}], where CW_{min} is the minimum contention window size. The backoff counter decreases by one for every idle slot and freezes if the channel is busy. The decrement procedure resumes after the channel is sensed idle again for a DIFS. The station transmits the packet when the backoff counter reaches zero. If there is no acknowledgment (Ack) received due to collision or transmission errors, the contention window size doubles after each unsuccessful transmission trial until it reaches the maximum value (CW_{max}), and the sender reschedules the transmission according to the aforementioned backoff rule. The frame is dropped when the retransmission limit is reached. If a frame is received successfully, the receiver transmits an Ack following a Short InterFrame Space (SIFS). After every successful transmission, contention window size is reset to its initial (minimal) value.

5.3.2 VoWLAN Basics and Limitation

In a VoWLAN, analog voice signals are first digitized, compressed and encoded into digital streams by the voice codecs. The output digital voice streams are then packed into constant-bit-rate (CBR) voice packets by the packetizer. Each voice packet has a 40-bytes real-time transport protocol (RTP)/user datagram protocol (UDP)/IP header followed by the voice payload. Before the transmission over wireless medium, the voice packet is further overloaded by two types of header, namely 1) 24-byte physical layer (PHY) header, which is transmitted with a very low transmission rate (e.g., in the IEEE 802.11b the PHY header is transmitted with rate 1Mbps), and 2) 34-bytes MAC layer header, which is transmitted with the same transmission rate of the voice packet. After a successful transmission, the destination node must reply with an 14-bytes acknowledgment (Ack) packet overloaded by the same PHY header. Once the Ack packet is received by the source node, the transmission session is ended. We can easily notice that this traditional Packet-Ack mechanism involves a lot of overheads and results in a very low channel utilization, especially in case of transmitting a huge amount of small size packets like voice.

The current DCF transmission mechanism in VoWLAN suffers from severe unfairness between downlink (AP-to-Station) and uplink (Station-to-AP). As illustrated in Fig.5.2, suppose that there are n voice conversations over WLAN, so downlink will be shared by the n voice connections, while uplink will be only occupied by a single connection. Thus, the volume of entire downlink voice traffic is approximately equal to the aggregated traffic of all uplink voice connections. The downlink may be further overloaded due to the coexistence of other background traffic [83]. Since the AP and all stations have only a single CSMA/CA instance and they equally contend to gain the chance to send their packets [84]. This contention-based mechanism leads to an equal transmission opportunity to all stations in the WLAN despite of the difference between their workloads. Due to the bi-directional characteristic of voice traffic, the difference in throughput and packet delay between downlink and uplink are extremely enlarged with the increase of ongoing voice connections. Thus, the downlink voice connection becomes one of the major bottlenecks to restrain the voice capacity of current IEEE 802.11 WLAN.

5.3.3 Related Works

There have been a number of previous works addressing the unfairness challenge of 802.11 WLAN. In [67], Clifford et al applied the packet bursting technique in the channel access of AP to improve the fairness. Its main concept is that once the AP gains access to the wireless medium through ordinary contention, consecutive transmissions of multiple packet are allowed without contending for the medium again. However, this packet bursting enforces a much longer channel occupation time, which may have a significant impact on the other real-time traffic like video.

In [68], Casetti and Chiasserini proposed direction-based prioritization scheme to solve the unfairness problem between downlink and uplink traffic. It assigns higher priority EDCA parameters to downlink traffic, while assigns lower priority parameters to uplink traffic. Although this approach can improve fairness, it may not always lead to a higher voice capacity.

Bi-directional transmission approach in single CSMA transmission session has been previously investigated in literature [69, 1, 70]. The concept was first introduced in [69],

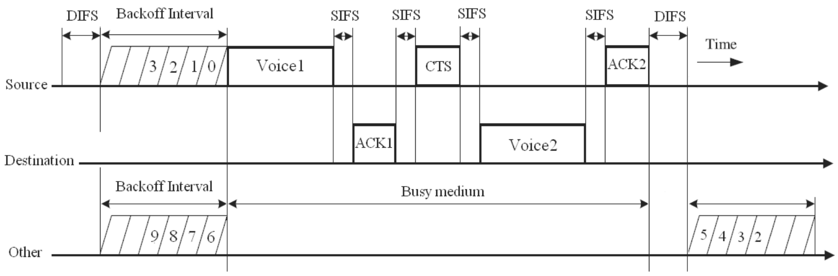


Figure 5.3: Illustration of the DCF+ protocol by Wu et al [1].

where the receiver is allowed to append one of his packets to the traditional Ack and transmit it on the same CSMA instance, regardless of its destination address. However, this protocol requires that all stations must be always ready to receive packets at any time, which is not backward compatible with the original protocol.

Wu et al [1] proposed 'DCF+' to restrict the bidirectional transmission to be between the sender and the receiver only. As illustrated in Fig. 5.3, the receiver appropriately sets the duration field of the ACK packet if it holds a data packet for the sender. In the sender side, the clear-to-send (CTS) packet is employed to reserve the channel for reverse data transmission, and is sent in response to the (ACK1). Although this approach can significantly improve fairness, it can only slightly increase the voice capacity due to the new overheads introduced by exchanging ACK and (CTS) packets between the sender and receiver.

Based on 'DCFmm' scheme proposed in [70], the bi-direction transmission is implemented at the AP only. In addition, the AP in this scheme sends by itself a Self-Clear-To-Send (Self-CTS) packet to reserve the channel if it has a packet to send. Unlike the 'DCF+' scheme, the 'DCFmm' can be easily deployed in the existing 802.11 WLANs, because it needs to modify the AP only. However, it shares with the 'DCF+' scheme the same drawback of inefficient resource utilization and thus a low voice capacity improvement. Moreover, it also introduces unfairness but this time it is for the uplink.

5.4 Illustration of the Proposed MAC Protocols

In this section, we introduce our proposed MAC protocols for VoIP support over IEEE 802.11 WLANs. First, we present our proposed DCFvs (DCF voice support) protocol, which work efficiently under idea channel conditions. Second we present our proposed DCFsvs (DCF secured voice support) protocol, which suites the WLANs that suffer from non-ideal channel conditions.

5.4.1 DCFvs

It is notable that in the current DCF, the two main factors that restrain the voice capacity are the downlink-uplink unfairness and the inefficient Packet-Ack transmission mechanism. The basic idea of the DCFvs is to first remove the downlink bottleneck by adopting the bidirectional transmission mechanism instead of unidirectional one, and then enhance the channel utilization efficiency by removing the dispensable overheads in the transmission process.

To remove the downlink bottleneck, we allow the bidirectional transmission between the AP and mobile stations in each transmission chance (from Station to AP or vice versa). As illustrated in Fig. 5.3, the bidirectional transmission permits the receiver to send a packet to the sender within the same transmission session based on the 'Packet-Ack-CTS-Packet-Ack' mechanism. By this way, the downlink will gain the same number of transmission chances as the uplink on the long run, so the voice packets in the downlink direction will not suffer from a long queuing delay in AP. The channel utilization efficiency of current DCF is simply due to the huge overheads involved at the transmission of a packet. These overheads can be divided into two parts, namely indispensable overheads (like PHY and MAC attached to a voice packet) and dispensable overheads (like Ack and CTS packets). The Ack packet is usually used for two main purposes: 1) to inform the sender that its packet has been received successfully, and 2) to declare when the wireless medium channel will be free again. This declaration is done by updating the duration field in the header of the Ack packet [70]. The CTS packet, on the other hand, is used to reserve the channel and avoid collision due to simultaneous transmission over the same wireless channel. Based on the observation that for the small size voice packets the above tasks of Ack and CTS can be accomplished in a much simple way, we propose the following DCFvs protocol.

Table 5.1: The DCFvs protocol

At the Source node Send (Voice1)

At the Destination node
if (Receive(Voice1))* **then**
{
  if (Exist(Voice2))** **then**
    Send(Voice2)
  else
    Send(Ack1)
}
else freeze***

At the Source node **if**(((Receive(Voice2)) or (Receive(Ack)))== FALSE) **then**   Failed(Voice1)****
* Receive(Packet): Boolean function that indicates the success or failure of Packet reception. * Exist(Packet): Boolean function that indicates the existence of Packet which is ready to be sent to the same Source only. ** Freeze: no response until Ack timeout pass. *** Failed(Packet): Duplicate the CW size and retransmit it again in the next chance. If the packet is not successfully transmitted withing a certain limit of sending trials (e.g. 4 trials in case of the voice packet), it will be discarded.

As shown in Table 1, DCFvs works as follows: Suppose that the source station got the chance to access the wireless channel. It starts to send its voice packet (Voice1) as shown in Fig. 5.4. If the destination has a packet (Voice2) to be sent to such source, it starts to send Voice2 once Voice1 is received successfully. Otherwise, it will send a usual Ack packet if the destination has nothing to send to that source. We have to notice that if the destination didn't receive Voice1 successfully, it will freeze until the ACK-timeout passes and new contention starts again among the stations. As we can notice in the above scenario, both Ack and CTS are not used, this is because of the following two reasons.

- First, since Voice2 will not be transmitted unless Voice1 is successfully received, therefore, the transmission of (Voice2) will be enough to declare the successful reception of Voice1, which is the first purpose of the Ack packet.

- Second, by updating the duration field in the MAC header of Voice2 with the corresponding time needed for transmitting it, the second purpose of the Ack packet will be accomplished too. Moreover, we have to highlight that the frozen stations

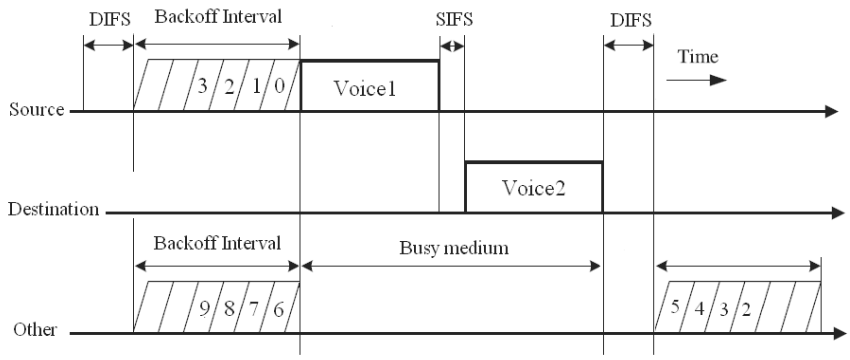


Figure 5.4: Illustration of the proposed DCFvs protocol.

will not resume their contention again unless the medium is sensed to be free for DIFS period [55, 56], which will not happen until Voice2 is fully transmitted.

Furthermore, the transmission session between the source and the destination will be ended by the reception of Voice2. Again, we have to notice that the destination node does not need an Ack2 from the source node (i.e., as a response to Vocie2) because of two reasons.

- First, the probability that Voice2 will be dropped due to collision is zero, because Voice2 will not be transmitted unless the channel is already reserved by the source station while all the other stations are frozen.

- Second, inspired by the fact that the Ack packet in the original DCF protocol does not need another Ack from the other side, we argue as follows: Given that Voice1 was already successfully transmitted, the probability that Voice2 will be dropped due to as sudden change in the channel conditions is very low. This is because the size of the voice packet is very small (almost equal to the size of the Ack packet) and also because the time needed to transmit the voice packet itself is very short (in term of several tens of micro seconds), where the channel condition can not change dramatically in such a short time period unless the speed of the mobile station is very high [85].

5.4.2 DCFsvs

The DCFvs protocol will work well for those WLANs with ideal wireless channel condition. For WLANs that suffer from non-ideal channel condition, we argue that there is still no need of Ack1 because Voice2 will be used to acknowledge the successful reception of Voice1. However, since there is no way to guarantee the successful delivery of Voice2 in a non-ideal channel condition, so we extend the DCFvs protocol by forcing the source to

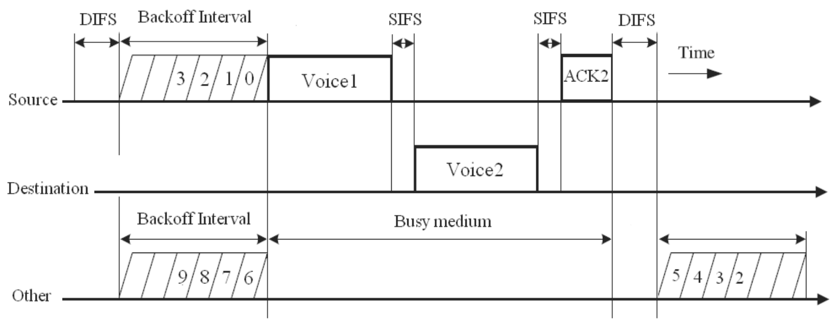


Figure 5.5: Illustration of the proposed DCFsvs protocol.

send an Ack packet to the destination as response to Voice2, as shown in Fig. 5.5. We refer to this extended version of the DCFvs protocol as DCFsvs (DCF for secured voice support).

5.5 Analytical Model for Throughput and Voice Capacity

In this section, we present models for throughput and voice capacity estimation in a DCFvs or DCFsvs-based IEEE 802.11 WLAN. To analyze the DCFvs, we first evaluate the maximum channel throughput H_{max} that can be achieved in AP. The H_{max} can be easily modeled as in [16] and [17]:

$$H_{max} = \frac{T_P \times R}{T_{voice} + T_{Ack} + T_{SIFS} + T_{DIFS} + T_{Backoff}} \tag{5.1}$$

The above equation indicates that the channel throughput is a function of the AP's transmission rate R and other parameters, like the time T_P needed to transmit the voice payload, the time T_{voice} needed to transmit the whole voice packet, the SIFS time T_{SIFS}, the DIFS time T_{DIFS}, the time T_{Ack} needed to transmit the acknowledgment packet, and the backoff time overhead $T_{Backoff}$. The $T_{Backoff}$ can be evaluated as follows based on the assumption that there is always two active stations (one is the AP and the other is one of the mobile stations) [16]

$$T_{backoff} = \begin{cases} 4.5 \times 9 + T_w \times 0.06\, \mu s & \text{for the 802.11a WLAN} \\ 8.5 \times 20 + T_w \times 0.03\, \mu s & \text{for the 802.11b WLAN} \end{cases} \tag{5.2}$$

where T_w is defined in terms of some basic parameters as

$$T_w = T_{SIFS} + T_{DIFS} + T_{voice} + T_{Ack} \tag{5.3}$$

Table 5.2: The IEEE 802.11(a/b) DCF-based WLAN Parameters

		802.11b	802.11a
Available Transmission Rates		11 Mbps	54 Mbps
Slot Time		20μs	9μs
SIFS		10μs	16μs
DIFS		50μs	34μs
CW_{min}		32	16
CW_{max}		1024	1024
Retry Limit		7	7
T_{voice}	PLCP & Preamble	192μs	24μs
	MAC Header + FCS	24.7μs	5μs
	RTP/UDP/IP Header	29.1μs	6μs
	Voice Payload (T_v)	(payload $\times$8/transmission rate)μs	
T_{Ack}	PLCP & Preamble	192μs	24μs
	Ack Frame	10.2μs	2.1μs

Using the same method, we can model the maximum channel throughput H_{max} that can be achieved in case of DCF+, DCFvs, and DCFsvs-based IEEE 802.11 WLAN as follows

$$H_{max} = \begin{cases} \frac{2\cdot T_P \times R}{2\cdot T_{voice}+2\cdot T_{Ack}+4\cdot T_{SIFS}+T_{DIFS}+T_{Backoff}+T_{CTS}} & \text{DCF+} \\ \frac{2\cdot T_P \times R}{2\cdot T_{voice}+T_{SIFS}+T_{DIFS}+T_{Backoff}} & \text{DCFvs} \\ \frac{2\cdot T_P \times R}{2\cdot T_{voice}+T_{SIFS}+T_{Ack}+T_{DIFS}+T_{Backoff}} & \text{DCFsvs} \end{cases} \tag{5.4}$$

Based on the above channel throughput model, the voice capacity C of the WLANs AP (i.e., the number of VoIP connections the AP can simultaneously support) can be expressed as follow:

$$C = \left\lfloor \frac{H_{max}}{2L} \right\rfloor \tag{5.5}$$

where all voice calls are assumed to use the same voice codec with bit rate L bps.

The typical values for above basic parameters are shown in the Table II when the 802.11(a/b) standards are considered.

5.6 Simulation Results

In this section, we present simulation results to show the effectiveness of the proposed DCFvs scheme for VoIP support over IEEE 802.11a and 11b WLANs.

5.6.1 Simulation Setting and Scenario

Our simulation is based on the topology shown in Fig. 5.6. We developed simulators for both IEEE 802.11a and 802.11b when DCFvs or DCFsvs-based is applied. The basic

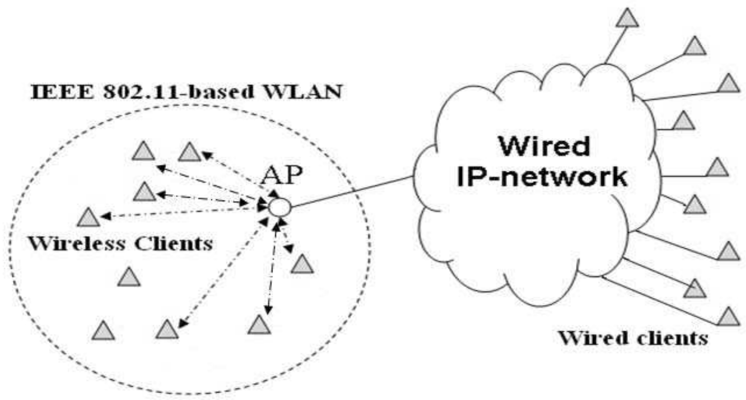


Figure 5.6: Network topology for simulations

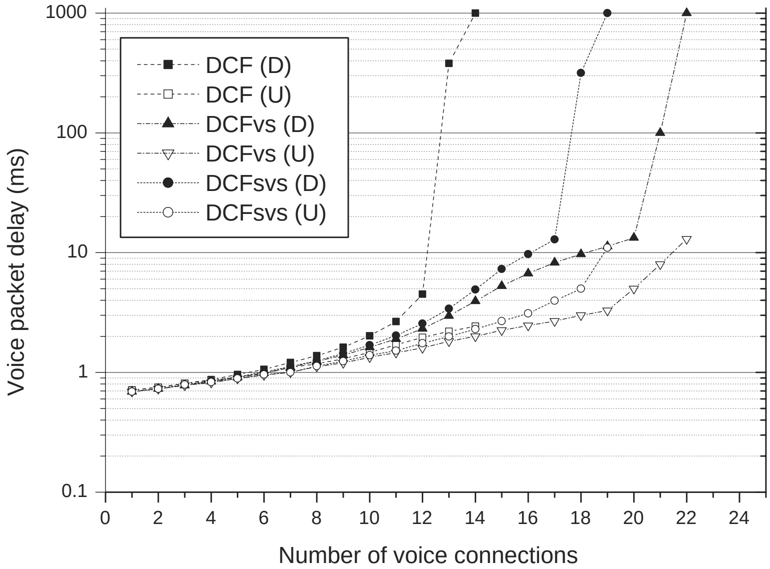
parameters used in our simulation are summarized in Table II. For voice traffic, the standard voice codecs G.711 and G.729 are considered with payload size 160 bytes and 20 bytes, respectively. For performance comparisons, we also developed the above simulators for DCF, DCF+ and DCFmm protocols too.

In our experiment, n wireless client stations were associated with an AP. This setting was used to make full-duplex VoIP calls between wireless client stations and wired nodes in the wired network counterpart through the AP. We are interested in only connections between the AP and the wireless stations. It is assumed that two voice connections of each voice call are independent of each other. Through the chapter, all the simulation results are average over 100 simulations each for 100 sec length and with a random starting time for each flow.

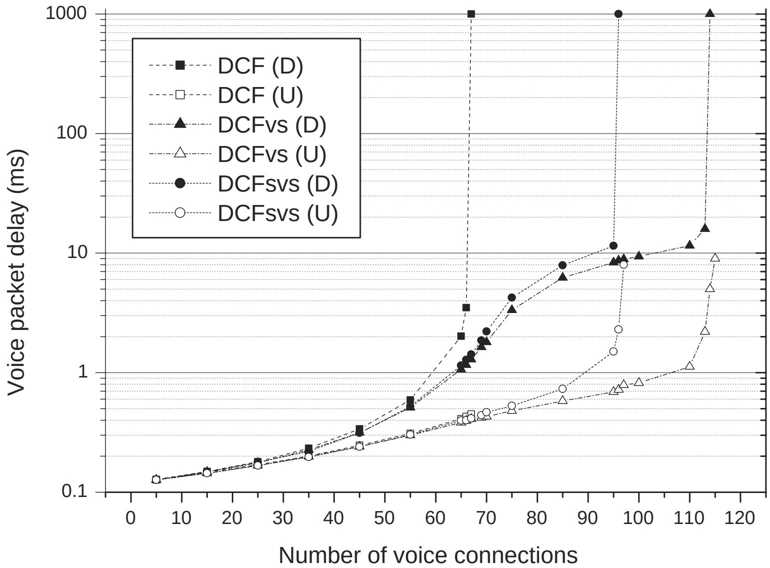
5.6.2 Voice Capacity

In our simulation of voice capacity, we consider the delay of the voice packet in both downlink (D) and uplink (U) traffic, and compare both DCFvs and DCFsvs protocols with other protocols. Fig. 5.7 compares the voice packet delay between DCF, DCFvs, and DCFsvs under different network settings. In DCF, the difference between voice packet delay of both uplink and downlink is increasing exponentially once the number of voice calls exceed certain limit, which is considered to be the voice capacity of the WLAN. The proposed DCFvs, on the other hand, reduces the delay of both downlink and uplink and significantly improve the fairness and the voice capacity. For example, in the IEEE 802.11b/G711 and 802.11a/G729, the proposed DCFvs succeeds to increase the voice capacity from 11 to 20, and from 65 to 114, respectively (i.e., more than 80% improvement in both cases).

Figure 5.8 and 5.9 compare the performance of 5 different schemes for the downlink and uplink, respectively. We can easily notice the very low improvement in voice capacity due



(a) IEEE 802.11b WLAN / voice codec G.711.



(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 5.7: The downlink and uplink voice packet delays under DCF, DCFvs, and DCFsvs protocols.

Table 5.3: Maximum number of VoIP connections for different MAC protocols

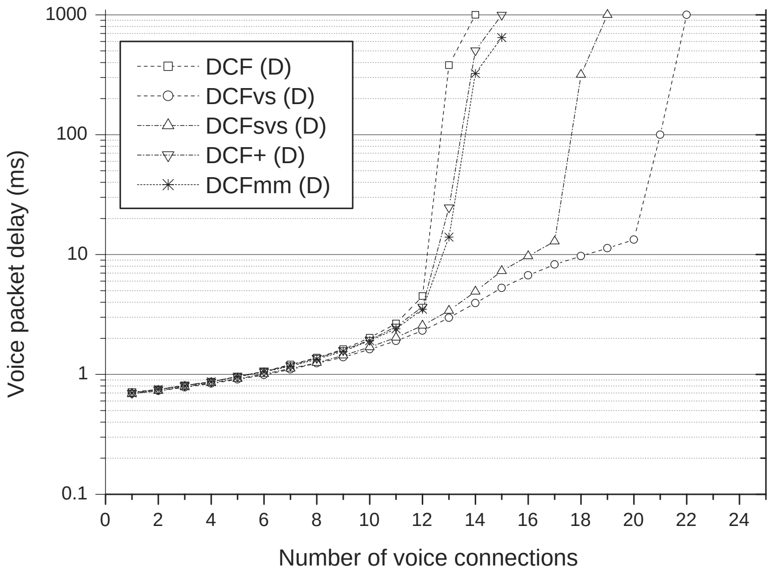
Parameter	DCF	DCF+	DCFvs	DCFsvs
H_{max} (Mbps)	1.570036	1.586354	2.6235	2.17243
C	12	13	20	16

to the adoption of bidirectional packet transmission in the DCF+ and DCFmm schemes in comparison with the original DCF scheme. On the other side, it is interesting to notice the remarkable improvement in voice capacity thanks to the DCFsvs and DCFvs schemes. it is also quite clear that the DCFvs scheme is still reporting a much better voice capacity due to minimum overhead.

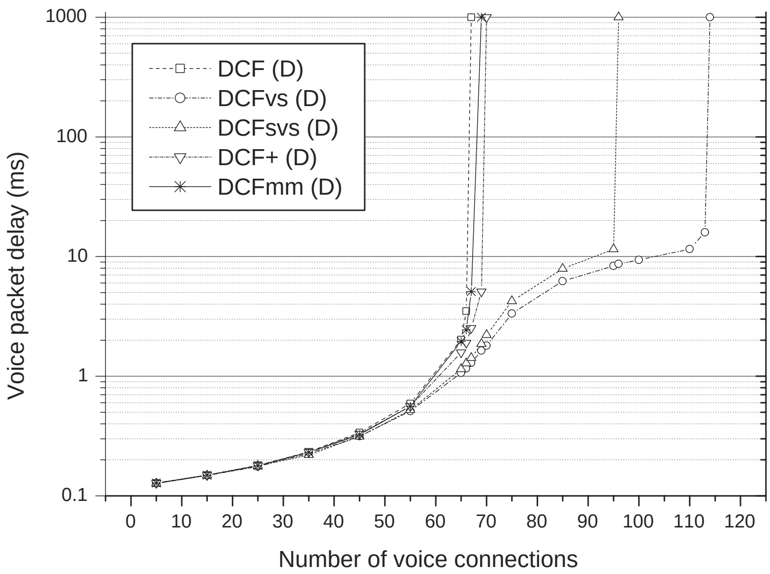
We use Equations (1-3) to compute the maximum throughput and the maximum number of VoIP connections that a single 802.11b access point can support. These calculations are done using the famous voice codec G711 with 20 ms of audio data in 160 bytes of payload, without any compression. Table III tabulates the maximum throughput and voice capacity that can be achieved under different MAC protocols: DCF, DCF+, DCFvs, and DCFsvs-based IEEE 802.11b WLAN. In these calculations we assumed that each transmission is performed with the maximum possible transmission rate. By comparing the results illustrated in Fig. 5.8.a) and Table III, we can easily notice that the voice capacity of all the protocols matches nicely with the analytical results.

5.6.3 Throughput

We here seek to find the maximum throughput that can be achieved by the proposed DCFvs protocol and then compare it with the available ones. In this simulation we add the voice calls one by one while monitoring the network throughput only. We notice that the throughput increase gradually up to a certain limit, which we defined as H_{max} and then it starts to decrease again due to the increase of contention between nodes and the increase of packet collision probability. As illustrated in Fig. 5.10, we can easily notice the throughput's enhancement achieved by the DCF+ and DCFmm in comparison with the traditional DCF scheme. However, DCFvs still reports the best performance among all the schemes in term of throughput because of the shorter duration of channel access session in compare to the other schemes. Such reduction increases the chance of accommodating more transmission sessions (CSMA) on the long-term. We should also highlight that the theoretical maximum throughput achieved by each scheme match well with the obtained numerical results tabulated in Table III.

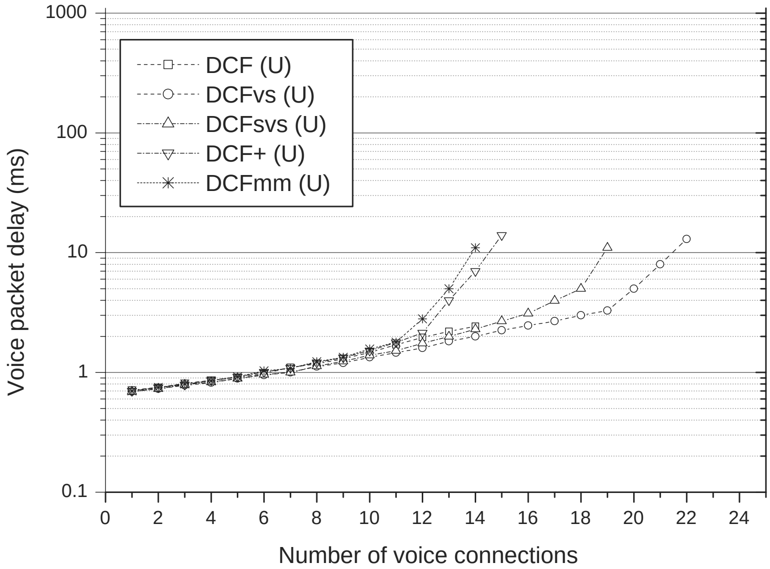


(a) IEEE 802.11b WLAN / voice codec G.711.

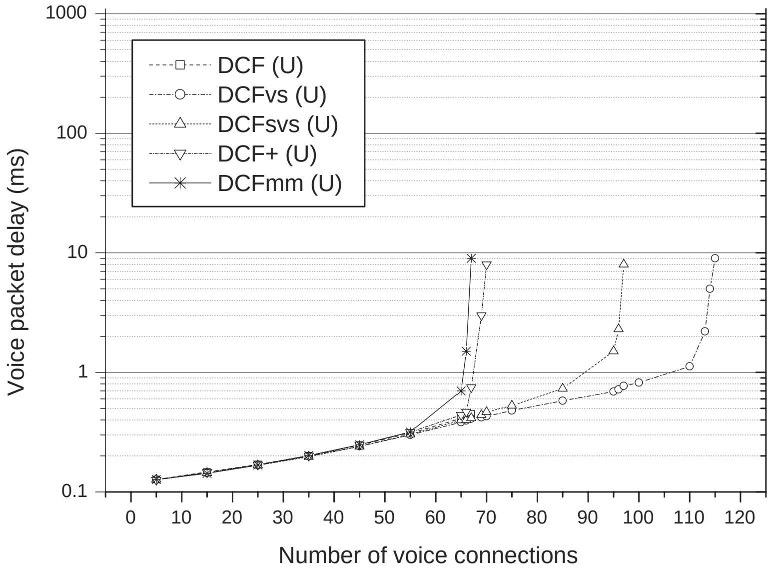


(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 5.8: Comparison between the downlink voice packet delays under different protocols.



(a) IEEE 802.11b WLAN / voice codec G.711.



(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 5.9: Comparison between the uplink voice packet delays under different protocols.

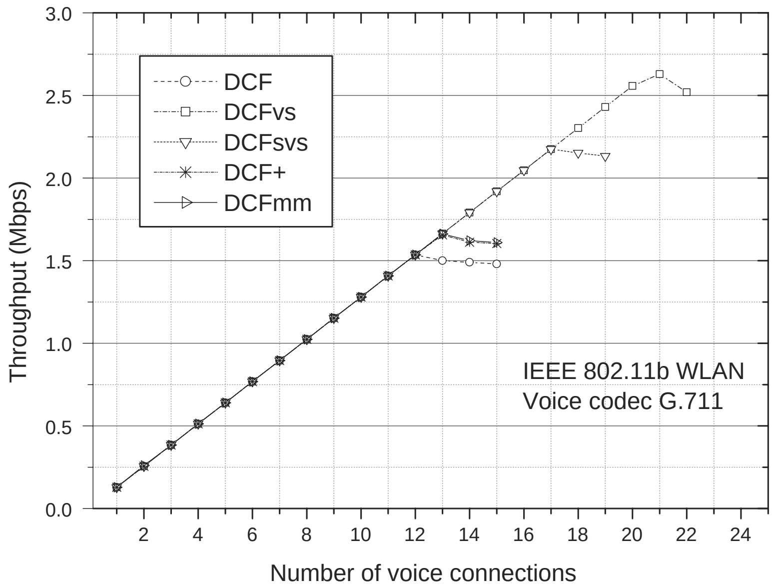


Figure 5.10: Throughput achieved by DCF, DCF+, DCFmm, DCFvs, and DCFsvs for voice only scenario.

5.6.4 Fairness Between Downlink and Uplink

Here we use a metric called fairness index [86] with a slide modification to measure the fairness between downlink and uplink. The fairness index, f is defined as follows: if the goodput[1] of the downlink and uplink are G_1 and G_2, respectively, then

$$f = (\sum_{i=1}^{2} G_i)^2/(2\sum_{i=1}^{2} G_i^2) \tag{5.6}$$

The fairness index is always a non-negative value, which lies between 0 and 1. The closer the value to 1, the better the fairness is. Fig. 5.11 reports the fairness index for the goodput achieved by the downlink and the uplink using DCF, DCFvs and DCFsvs protocols. All of the bidirectional transmission approach scheme family (i.e.,DCFmm, DCF+, DCFsvs and DCFvs) can easily provide a better fairness index in comparison with the traditional DCF scheme. Even though, the DCFsvs and the DCFvs are still able to report the best performance among the other schemes due to their efficient transmission mechanism, which supports both downlink and uplink in each single channel access session.

5.6.5 VoIP Support Under Coexistence of Best-Effort Traffic

As presented in this chapter, DCFvs and DCFsvs aim to support the VoIP traffic even under the coexistence of best-effort (BE) traffic. Therefore, we here create a new simula-

[1]The goodput is the throughput without taking the retransmitted traffic into account.

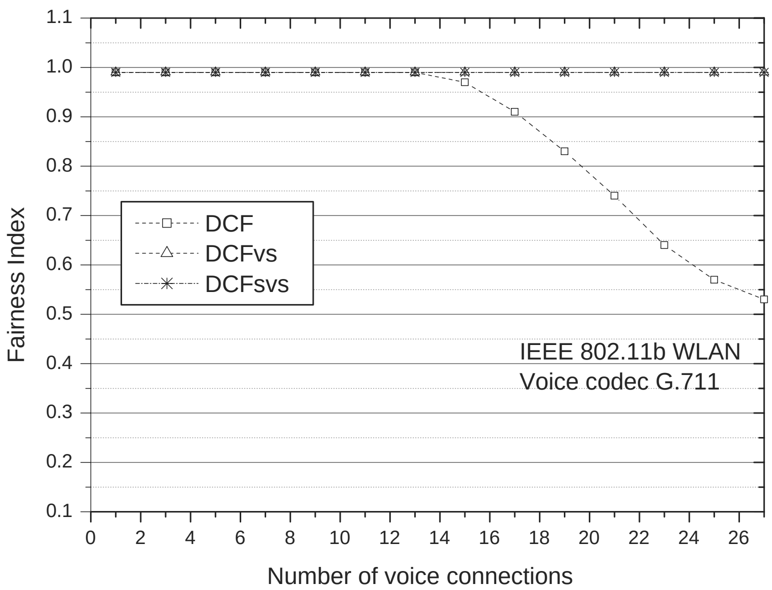


Figure 5.11: Fairness index of the downlink and uplink traffic under the DCF, DCFvs, and DCFsvs protocols.

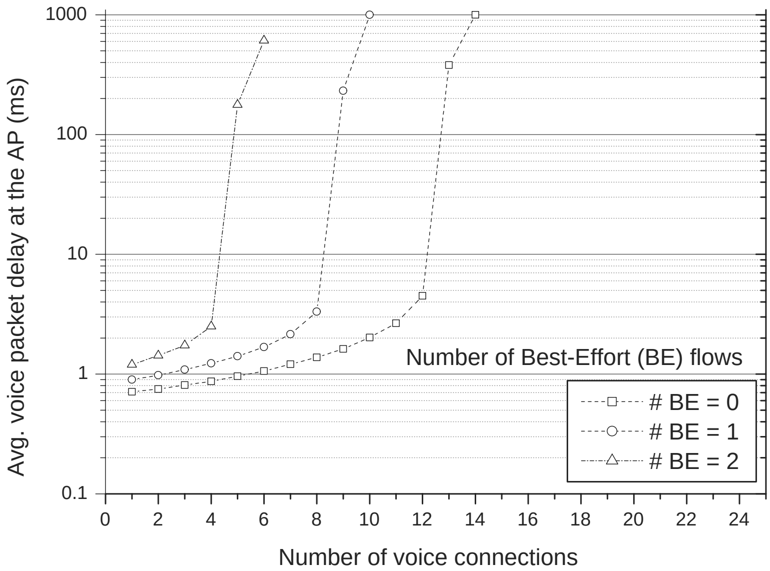
tion scenario to evaluate their performance under such realistic condition. Since in most of the cases, data traffic is heavily loaded on the downlink direction [16] for downloading purpose, in our simulation we create a BE flow which flows from the AP to one of the wireless stations. Concurrently, we report the average voice packet delay and the correspondent voice capacity under the gradual increases of the BE flows number. Each BE flow is generated at the form of constant bit rate (CBR) with packet size of 440 bytes[2], and its load is set to 1Mbps.

As illustrated in Fig. 5.12, 5.13, and 5.14, we can easily notice that the coexistence of BE-flows reduces the voice capacity under the DCF-based protocol. For example, the voice capacities in the DCF-based IEEE 802.11b WLANs is reduced from 12 to 8 and then to 4, due to the coexistence of one and two BE flows, respectively. Although the voice capacity is also slightly reduced in the DCFvs and DCFsvs protocols, they still report a significant support to the VoIP traffic. We can easily notice from the results that the voice capacity in the DCFvs-based IEEE 802.11b WLANs is only reduced from 20 to 15 and then to 10, due to the coexistence of one and then two (1 Mbps BE-flows), respectively.

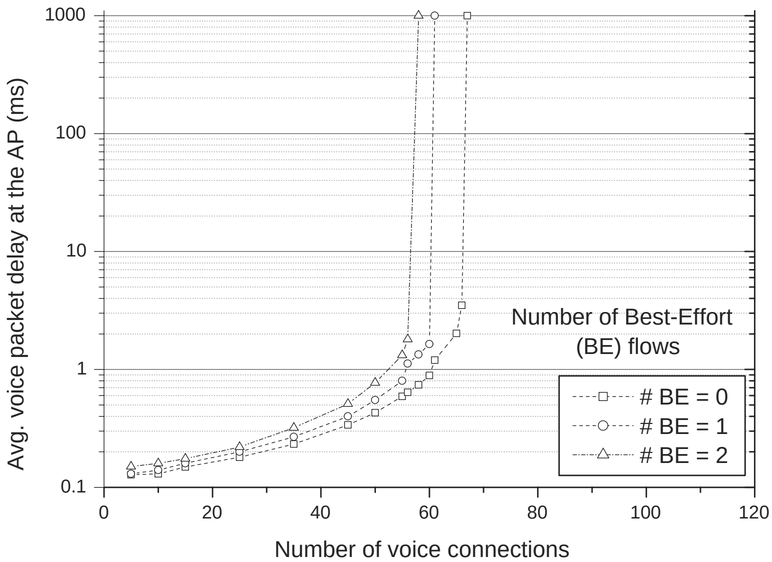
5.6.6 Best-Effort Flows as Dominant Traffic

Actually, we found that conducting the same experiment in the opposite direction will match well with the real case, where the best-effort flows are the dominant traffic and the VoIP traffic is the minor one. This is why we conduct a new simulation experiment to

[2]Since the packet size of 60%, 20%, and 20% of the BE traffic is 44, 550, and 1500 bytes [72], respectively, then the packet size of the BE traffic in our simulation is settled to the average size.

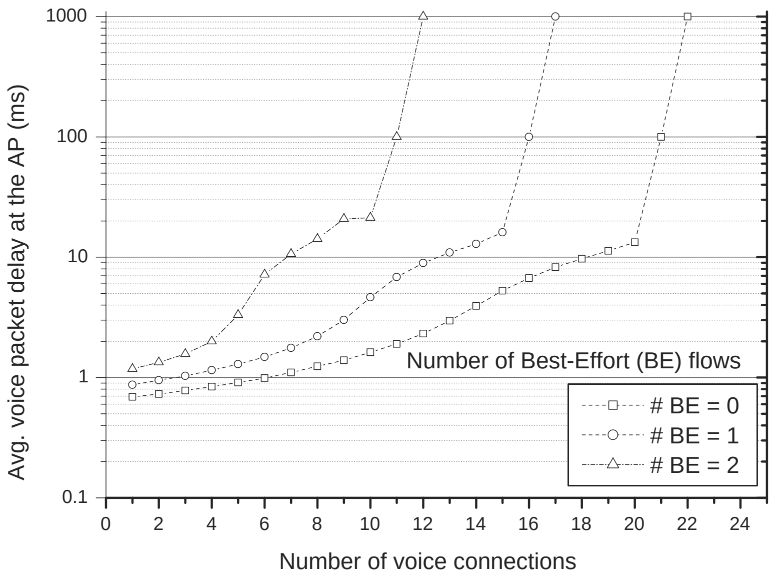


(a) IEEE 802.11b WLAN / voice codec G.711.

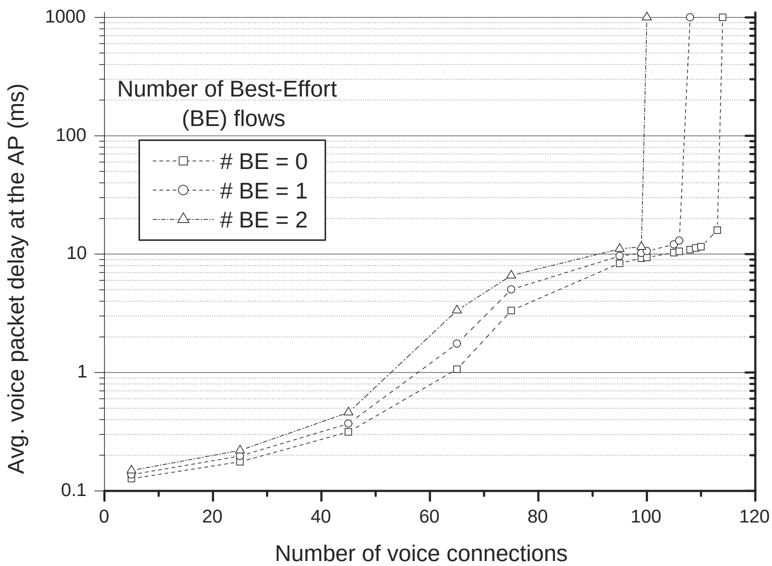


(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 5.12: Voice capacity under the coexistence of best-effort traffic in a DCF-based WLAN.

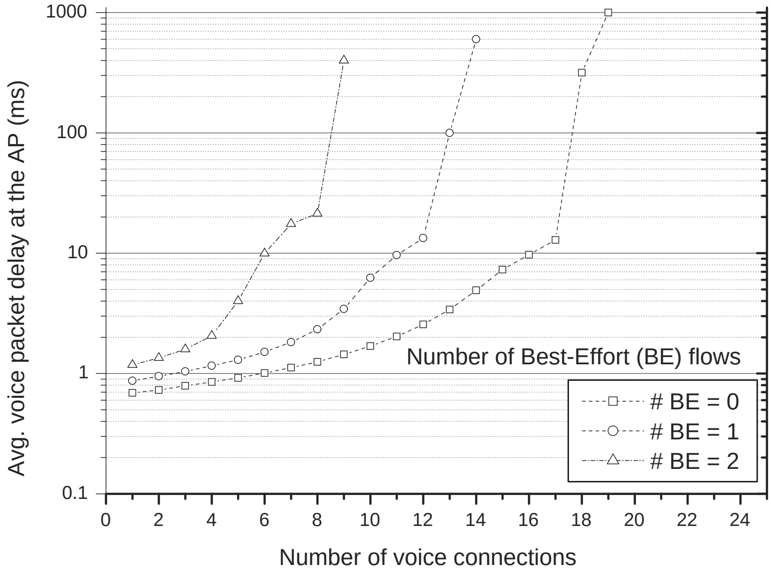


(a) IEEE 802.11b WLAN / voice codec G.711.

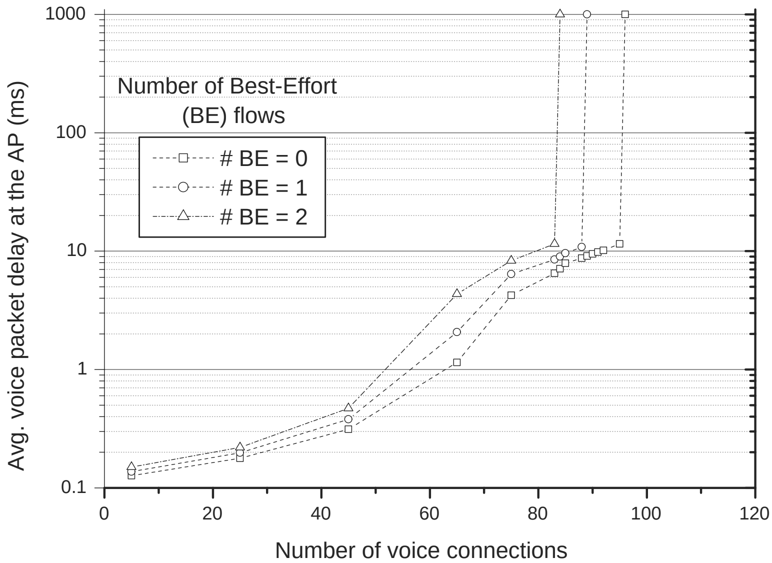


(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 5.13: Voice capacity under the coexistence of best-effort traffic in a DCFvs-based WLAN.



(a) IEEE 802.11b WLAN / voice codec G.711.



(b) IEEE 802.11a WLAN / voice codec G.729.

Figure 5.14: Voice capacity under the coexistence of best-effort traffic in a DCFsvs-based WLAN.

inspect the maximum number of voice calls a wireless access point can support under the coexistence of a big volume of best-effort traffic.

In our simulation, we generate 5 different rates of best effort flows (1Mbps, 600Kbps, 300Kbps, 200Kbps, and 100Kbps), which flows from the AP to one of the wireless stations. We use G.711 and G.729 voice codecs to generate the voice calls with 20ms payload size. The G.711 was jointly incorporated with the IEEE 802.11b WLAN, while the G.729 with the IEEE 802.11a WLAN.

In our simulation, a specific number of BE-flows is generated as a dominant traffic, and maximum number of voice calls that can be supported simultaneously by the access point is examined. For example, as tabulated in the second row of Table 5.4, if there exist four BE flows (each of 600 Kbps) the DCF-based IEEE 802.11b WLAN can support only three G.711 voice calls while the DCFvs-based and the DCFsvs-based WLANs can support six and five calls, respectively. In general, we can easily notice from the obtained simulation results that the proposed protocols can efficiently provide a better support to the VoIP applications over the IEEE 802.11 WLANs under the coexistence of a dominant Best-effort traffic in comparison with the available protocols.

Table 5.4: The WLAN's voice capacity under the existence of dominant Best-effort traffic

Best-effort flows		WLAN's voice capacity IEEE 802.11b G.711 voice codec		
Rate	flows	DCF	DCFvs	DCFsvs
1 Mbps	2	5	10	8
600 Kbps	4	3	6	5
300 Kbps	7	4	8	6
200 Kbps	13	3	5	4
100 Kbps	17	5	10	8
Best-effort flows		WLAN's voice capacity IEEE 802.11a G.729 voice codec		
Rate	flows	DCF	DCFvs	DCFsvs
1 Mbps	13	10	20	16
600 Kbps	22	9	17	14
300 Kbps	43	10	19	16
200 Kbps	65	10	19	15
100 Kbps	100	22	41	33

5.7 Summary

In this chapter, we addressed the downlink bottleneck and the overloaded Packet-Ack transmission mechanism that result in a non efficient voice support in the current IEEE

802.11 DCF-based protocol. We proposed a novel medium access control scheme that overcome downlink problem with the help of bidirectional transmission approach and then we adjust transmission mechanism from Packet-Ack to Packet-Packet based on the unique characteristics of the voice traffic. We then extend the proposed protocol to a more secured version in order to suit those WLANs that suffer from non-ideal channel conditions.

Extensive simulation results demonstrated that bidirectional transmission approach can improve the WLAN voice capacity as well as fairness between downlink and uplink traffic, but this improvement can be more significant when the proposed transmission mechanism is jointly adopted with it. For example, in an IEEE 802.11b WLAN/ voice codec G.711 (20ms payload), the available bidirectional approach can slightly improve the voice capacity while this improvement can be significantly high (more than 80%) with the help of the proposed transmission mechanism. The results in this chapter also indicate clearly that voice capacity is severely degraded with the coexistence of the best-effort traffic, while the proposed protocols still guarantee a very efficient support to the VoIP traffic under the same condition.

Chapter 6

Conclusion

6.1 Summary and Discussions

In this dissertation we have addressed a group of problems which significantly affect the performance of the voice applications over the wired and wireless IP-networks. Our research mainly focuses on four problems. First, the voice packet scheduling problem especially with the expected immense migration of such traffic from the PSTN to the Internet. Second, the inaccurate available WLANs' voice capacity estimation models ,in addition to the ineffectual AP placement scheme currently utilized in such type of networks. Third, the loose and inefficient MAC protocol presently deployed in the Wireless IP-networks. for each of those problems, we provide novel contributions, which are supported with intensive mathematical analysis and extensive simulation studies. Precisely, our contributions as listed as follows:

1. In chapter 2, we proposed a new scheduling architecture for the wired IP-networks. The new architecture is flexible in the sense that we can easily comprise between the packet delay of the voice traffic and fairness deserved by the other non-voice traffic through one control parameter only (i.e., the splitting ratio of voice traffic). we proved through both theoretical analysis and experimental simulations that it is possible for us to offer fairness to non-voice traffic without significantly sacrificing the performance of delay-sensitive voice traffic. We expect that our proposed architecture can efficiently handle the impacts that will rise up with the expected growth of VoIP traffic in the future voice-intensive IP networks.

2. In chapter 3, we developed a new model for Wireless IP-networks' voice capacity estimation and also a new scheme for access point placement with the consideration of clients' spatial distribution. We showed through both simulation and analytical studies that the proposed model can provide a very accurate estimation of the average voice capacity instead of the current upper and lower bounds estimation

models. Moreover, we showed that this capacity can be significantly enhanced if we place the access point properly by using our new placement scheme, especially when clients are unsymmetrically distributed in the WLAN area. It is expected that our work will contribute to the network planning and protocol design of future VoIP over WLANs.

3. In chapter 4, we focused on the bidirectional transmission-based protocol which promise to eliminate the downlink bottleneck of the VoIP traffic. We proposed a novel packet scheduler to maximize the bidirectional transmission chances in each transmission session. We then modify the old inherited DCF-based contention mechanism in order to achieve the maximum benefit of the proposed scheduler. The conducted theoretical analysis and simulation studies demonstrate that the enhanced bidirectional transmission protocol outperforms the available ones in terms of voice capacity as well as throughput even under the coexistence of the best-effort traffic.

4. In chapter 5, we proposed a novel medium access control (MAC) protocol that overcome most of the current protocols' drawbacks. We first propose a protocol which can be efficiently used over the WLAN that enjoy ideal channel conditions. We then extend it to suit those WLANs that suffer from non-ideal channel conditions. The conducted theoretical analysis and the extensive simulation studies demonstrated the WLANs performance has been dramatically improved (over 80% improvement in comparison with the available ones).The results in this chapter also indicate clearly that voice capacity is severely degraded with the coexistence of the best-effort traffic if the available protocols are adopted, while the proposed protocols still guarantee a significant support to the VoIP under the same condition.

6.2 Future Works

In this dissertation, we have proposed a group of approaches and models to enhance the voice applications performance over the wired and wireless IP-networks. Each of these approaches and models can be further extended to cover more complex conditions. The potential future works are as follows:

- The proposed packet scheduler (in chapter 2) was mainly designed to support the voice traffic only under the coexistence of the best-effort traffic (i.e., non-delay sensitive traffic). Such scheduler can be further extended to accommodate another type of delay sensitive traffic like video in addition to the voice. Although such new considered traffic will make our architecture more complex, but it is still simple to maintain in the meaning that the same traffic splitting approach will used in a more general paradigm.

- Notice that the regular rate region pattern considered in chapter 3 did not incorporates some factors like channel fading, shadowing effect, and hidden nodes. So, one future work is to examine the efficiency of our model for R_{avg} estimation based on a more realistic rate region pattern that incorporates the effect of those factors.

- Another interesting future work in the same chapter is to find a way to determine the required number of equal-size squares for a given maximum allowed estimation error of our model in voice capacity estimation.

- It also worth our effort to study how to reduce the overhead in the voice packet header used in the current MAC protocols especially the part which is dedicated to coop with the MAC and physical layer of the WLAN, so that more packets can be accommodated and hence more voice calls.

- A possible and useful future work is the evaluation the performance of the proposed MAC protocol in a wireless mesh network scenario [87]. The evaluations in terms of throughput and voice capacity in a multi-hop conditions can reveal the level of its practicability.

Bibliography

[1] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE802.11 wireless LAN: Analysis and enhancement ," in *IEEE INFOCOM*, 2002, pp. 599 – 607.

[2] P. Mockapetris, "Telephony's next act," in *IEEE Spectrum*, vol. 43, 2006, pp. 28–32.

[3] V. Soares, P. Neves, and J. Rodrigues, "Past, Present and Future of IP Telephony," in *Int. Conf. on Communication Theory, Reliability, and Quality of Service*, 2008, pp. 19–24.

[4] J. Conti, "Talk about a change [voip replacing pstn]," *IEE Review*, vol. 50, no. 12, p. 27, 2004.

[5] M. ALAkhras, "Voice over IP (VoIP)," in *Int. Conf. on Communications, Computers and Applications*, 2008, pp. xiii – xiii.

[6] H. M. Chong and H. Matthews, "Comparative analysis of traditional telephone and voice-over-Internet protocol (VoIP) systems," in *IEEE Int. Symposium on Electronics and the Environment*, 2004.

[7] P. Markopoulou, A. Tobagi, and J. Karam, "Assessing the quality of voice communications over internet backbones," *IEEE/ACM Trans. on Networking*, vol. 11, no. 5, pp. 747–760, 2003.

[8] C. Chu, H. Pant, S. Richman, and P. Wu, "Enterprise VoIP Reliability," in *Int. Symposium on Telecommunications Network Strategy and Planning*, 2006, pp. 1 – 6.

[9] P. Mehta and S. Udani, "Voice over ip," *IEEE Potentials*, vol. 20, no. 4, pp. 36 – 40, 2001.

[10] M. Shreedhar and G. Varghese, "Efficient fair queuing using deficit round-robin," *IEEE/ACM Trans. on Networking*, vol. 4, no. 3, pp. 375–385, 1996.

[11] A. demers, S. Keshav, and S. Shenker, "Design and analysis of a fair queuing algorithm," in *ACM SIGCOMM*, 1989, pp. 1–12.

[12] M. Karam and F. Tobagi, "Analysis of the delay and jitter of voice traffic over the Interne," in *IEEE INFOCOM*, vol. 2, 2001, pp. 824–833.

[13] D. G. et al, "Improving wlan voip capacity through service differentiation," *IEEE Trans. on Vehicular Technology*, vol. 57, no. 1, pp. 465 – 474, 2008.

[14] K. Brian and D. Mark, "An Experimental Analysis of the Call Capacity of IEEE 802.11b Wireless Local Area Networks for VoIP Telephony," in *Int. Conf. on Irish Signals and Systems*, 2006, pp. 283 – 287.

[15] S. Garg and M. Kappes, "Admission control for VoIP traffic in IEEE 802.11 networks," in *IEEE GLOBECOM*, 2003, pp. 3514 – 3518.

[16] ——, "Can I add a VoIP call," in *IEEE ICC*, 2003.

[17] D. Hole and F. Tobagi, "Capacity of an IEEE 802.11b Wireless LAN Supporting VoIP," in *IEEE ICC*, 2004.

[18] L. Cai, X. Shen, J. Mark, L. Cai, and Y. Xiao, "Voice capacity analysis of wlan with unbalanced traffic," *IEEE Trans. on Vehicular Technology*, vol. 55, no. 3, pp. 752–761, 2006.

[19] O. Awoniyi and F. Tobagi, "Effect of fading on the Performance of VOIP in IEEE 802.11 a WLANs," in *IEEE Int. Conf. on Communications*, vol. 6, 2004, pp. 3712 – 3717.

[20] Y. Zhang and B.-H. Soong, "Performance of mobile networks with wireless channel unreliability and resource insufficiency," *IEEE Trans. on Wireless Communications*, vol. 5, pp. 990 – 995, 2006.

[21] P. Dini, O. Font-Bach, and J. Mangues-Bafalluy, "Experimental analysis of VoIP call quality support in IEEE 802.11 DCF," in *Int. Symposium on Communication Systems, Networks and Digital Signal Processing*, 2008, pp. 443 – 447.

[22] S. Kanhere, H.Sethu, and A.Parekh, "Fair and efficient packet scheduling using elastic round robin," *IEEE Trans. on Parallel and Distributed Systems*, vol. 13, no. 3, pp. 324–336, 2002.

[23] A. Striegel and G. Manimaran, "Dynamic class-based queue management for scalable media servers," in *Real-Time Technology and Applications Symposium*, vol. 2, 2000, pp. 228–236.

[24] A. Kortebi, S. Oueslati, and J. Roberts, "An Implicit Service Differentiation using Deficit Round Robin," in *Proc. of ITC19*, Beijing, Aug. 2005.

[25] L. Lenzini, E. Mingozzi, and G. Stea, "Aliquem: a Novel DRR Implementation to Achieve Better Latency and Fairness at O(1) Complexity," in *Proc. of IWQoS*, 2002.

[26] S. Ramabhadran and J. Pasquale, "Stratified Round Robin: a low Complexity Packet Scheduler with Bandwidth Fairness and Bounded delay," in *ACM SIGCOMMM*, 2003, pp. 239–249.

[27] "The low latency queueing," http://www.cisco.com/warp/latencyqueueing.html.

[28] R. E. Abdouni and Khayari, "Class-based weighted fair queueing: validation and comparison by trace-driven simulation," *Int. Journal of Communication Systems*, vol. 18, pp. 975–994, 2005.

[29] S. Kanhere and H. Sethu, "Fair, Efficient and low-latency packet scheduling using nested deficit round robin ," in *IEEE Workshop on High-Performance Switching and Routing*, 2001.

[30] K. Yamakoshi, K. Nakai, E. Oki, and N. Yamanaka, "Dynamic deficit round-robin scheduling scheme for variable-length packets," *Electronics Letters*, vol. 38, no. 3, pp. 1480–1493, 2002.

[31] M. Karam and F. Tobagi, "WF2Q: Worst-case fair weighted fair queueing ," in *IEEE INFOCOM*, 1996.

[32] X. Fei and A. Marshall, "Delay optimized worst case fair WFQ (WF2Q) packet scheduling," in *IEEE ICC*, 2002.

[33] S. Oueslati and J. Roberts, "A new direction for quality of service: Flow-aware networking," in *Proc. of NGI*, Rome, 2005.

[34] N. Saberi and M. Coates, "Bandwidth Reservation in. Optical WDM/TDM Star Networks," in *22nd Bienn. Symp. on Comm*, 2004, pp. 219–221.

[35] S. Golestani, "A self clocked fair queueing scheme for broadband applications ," in *INFOCOM*, 1994.

[36] S. Kanhere and H. Sethu, "On the Latency Bound of Deficit Round Robin ," in *Int. Conf. on Computer Communications and Networks*, 2002.

[37] A. Kos and J. Miletic, "Enhanced Latency Analysis of DRR Algorithm," 2005, http://alas.matf.bg.ac.yu/mv00023/papers.htm.

[38] G. Appenzeller, I. Keslassy, and N. Mckeown, "Sizing router Buffer," in *ACM SIGCOM*, 2004.

[39] S. Gorinsky, A. Kantawala, and J. turner, "Link Buffer Sizing: A New Look at the Old Problem ," in *Proceedings of ISCC*, 2005.

[40] R. Cle and J. Rosenbluth, "Voice over ip performance monotoring," *ACM Computer Communication Review*, vol. 31, 2001.

[41] S. Henning, F. James, and T. Don, " Congestion control for real-time traffic in high-speed networks," in *IEEE INFOCOM*, 1990, pp. 543 – 550.

[42] A. Bhargava, J. Kurose, D. Towsley, and L. Van, "Performance comparison of error control schemes in high speed computer communication networks ," in *IEEE INFOCOM*, 1988, pp. 694 – 703.

[43] G. F. Mao, A. Talevski, and E. Chang, "Voice over Internet Protocol on Mobile Devices," in *IEEE/ACIS Int. Conf. on Computer and Information Science*, 2007, pp. 163 – 169.

[44] W. Wang, S. C. Liew, and V. Li, "Solutions to performance problems in voip over a 802.11 wireless lan," *IEEE Trans. on Vehicular Technology*, vol. 54, no. 1, pp. 366 – 384, 2005.

[45] P. Perini, "Wireless applications using Internet protocol and voice over IP," in *IEEE Proceedings on Aerospace Conference*, vol. 2, 2001, pp. 1043 – 1050.

[46] G. Me and D. Verdone, "An overview of some techniques to exploit VoIP over WLAN," in *Int. Conf. on Digital Telecommunications*, 2006.

[47] A. Kashyap, S. Ganguly, S. Das, and S. Banerjee, "VoIP on Wireless Meshes: Models, Algorithms and Evaluation ," in *IEEE INFOCOM*, 2007, pp. 2036 – 2044.

[48] S. Wang, Z. Mai, D. Xuan, and W. Zhao, "Design and implementation of qos-provisioning system for voice over ip," *IEEE Trans. on Parallel and Distributed Systems*, vol. 17, no. 3, pp. 276 – 288, 2006.

[49] F. Anjum, M. Elaoud, D. Famolari, A. Ghosh, R. Vaidyanathan, A. Dutta, P. Agrawal, T. Kodama, and Y. Katsube, "Voice Performance in WLAN Networks an Experimental Study ," in *IEEE GLOBECOM*, vol. 6, 2003, pp. 3504 – 3508.

[50] S. Garg and M. Kappes, "An Experimental Study of Throughput for UDP and VoIP Traffic in IEEE 802.11b Networks ," in *IEEE WCNC*, vol. 3, 2003, pp. 1748 – 1753.

[51] P. T. et al, "Capacity Estimation of VoIP Channels on Wireless Networks," 2003, http://www.ece.utexas.edu/wireless/EE381K11Spring03/projects/11.3.pdf.

[52] B. Ngamwongwattana, "Effect of packetization on VoIP performance," in *IEEE int. Conf. on Computer, Telecommunications and Information Technology*, vol. 1, 2008, pp. 373–376.

[53] A. Sfairopoulou, B. Bellalta, and C. Macian, "How to tune voip codec selection in wlans?" *IEEE Communications Letters*, vol. 12, no. 8, pp. 551 – 553, 2008.

[54] M. Narbutt and M. Davis, "An assessment of the audio codec performance in voice over WLAN (VoWLAN) systems," in *Int. Conf. on Mobile and Ubiquitous Systems: Networking and Services*, 2005, pp. 461–467.

[55] "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer extension in the 2.4/5 GHz band," ANSI/IEEE Std 802.11b," 1999.

[56] "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer extension in the 2.4/5 GHz band," ANSI/IEEE Std 802.11a," 1999.

[57] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network ," in *3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2002, pp. 80 – 91.

[58] C. Tuduce and T. Gross, "A mobility model based on WLAN traces and its validation," in *IEEE INFOCOM*, vol. 1, 2005, pp. 664 – 674.

[59] M. Hope and N. Linge, "Determining the propagation range of IEEE 802.11 radio LANs for outdoor applications ," in *Local Computer Networks*, 1999, pp. 49 – 50.

[60] H. Bertoni, *Radio Propagation for Modern Wireless Systems.* Printice Hall, 2000.

[61] G. Marta, A. Cesar, and L. Albert, "Understanding individual human mobility patterns," *Nature*, vol. 453, June 2008.

[62] M. Heger, "Human travel patterns surprisingly predictable," *IEEE Spectrum*, june 2008.

[63] R. P. et al, *The Feynman Lectures on Physics.* Addison Wesley, 2005.

[64] L. Cai, Y. Xiao, X. Shen, L. Cai, and J. W. Mark, "Voip over wlan: Voice capacity, admission control, qos, and mac," *Communication Systems*, vol. 19, pp. 491 – 508, 2006.

[65] D. Niculescu, S. Ganguly, K. Kim, and R. Izmailov, "Performance of VoIP in a 802.11 Wireless Mesh Network ," in *IEEE INFOCOM*, 2006.

[66] F. A. et al, "Voice Performance in WLAN Networks: An Experimental Study ," in *IEEE GLOBECOM*, 2003, pp. 3504 – 3508.

[67] P. Clifford, K. Duffy, D. Leith, and D. Malone, "On Improving voice capacity in 802.11 infrastructure networks ," in *WirelessCom*, 2005.

[68] C. Casetti and C. Chiasserini, "improving fairness and throughput for voice traffic in 802.11e EDCA ," in *IEEE PIMRC*, 2004, pp. 525 – 530.

[69] J. Tourrilhes, "PiggyData: Reducing CSMA/CA collisions for multimedia and TCP connections ," in *IEEE VTC'99 Fall*, 1999, pp. 1675 – 1679.

[70] J. Kim, H. Tode, and K. Murakami, "Medium access control protocol for voice traffic in ieee 802.11 wlans," *IEICE*, pp. 1545 – 1553, 2006.

[71] "IEEE 802.11 WG. Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: MAC enhancements for quality of service (QoS)," 2003.

[72] K. Thompson, G. Miller, and R. Wilder, "Wide-area internet. traffic patterns and characteristics," *IEEE Network*, vol. 11, pp. 10 – 23, 1997.

[73] A. F. et al, "Voice performance in WLAN networks - an experimental study," in *IEEE GLOBECOM*, vol. 6, 2003, pp. 3504 – 3508.

[74] Z. Becvar, J. Zelenka, and M. Brada, "Impact of saturation on speech quality in VoIP," in *Int. Conf. on Systems, Signals and Image Processing*, 2008, pp. 381–384.

[75] A. Takahashi, A. Kurashima, and H. Yoshino, "Objective assessment methodology for estimating conversational quality in voip," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1984 – 1993, 2006.

[76] W. Quan and D. M. Hui, "Improving the performance of WLAN to support VoIP application," in *Int. Conf. on Mobile Technology, Applications and Systems*, 2005, p. 5.

[77] M. Gardner, V. Frost, and D. Petr, "Using optimization to achieve efficient quality of service in voice over IP networks," in *IEEE Int. conf. on Performance, Computing, and Communications*, 2003, pp. 475 – 480.

[78] E.-C. Park, E. Park, D. Kim, and J. So, "Improving quality of service and assuring fairness in wlan access networks," *IEEE Trans. on Mobile Computing*, vol. 6, no. 4, pp. 337 –350, 2007.

[79] M. Castro, P. Dely, J. Karlsson, and A. Kassler, "Capacity Increase for Voice over IP Traffic through Packet Aggregation in Wireless Multihop Mesh Networks," in *Int. Conf. on Future generation communication and networking*, vol. 2, 2007, pp. 350 – 355.

[80] D. Hashmi, P. Kiran, and B. Lall, "Access Point priority based Capacity Enhancement scheme for VoIP over WLAN," in *Annual India Conference*, 2006, pp. 1–4.

[81] X. Ling, Y. Cheng, X. Shen, and J. Mark, "Voice capacity analysis of wlans with channel access prioritizing mechanisms," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 752 – 761, 2006.

[82] L. Yang, Z. Shengqing, and L. Xiaokang, "Enhanced MAC Protocol for Voice Communication in IEEE 802.11 WLAN," in *Int. Conf. on Digital Telecommunications*, 2007, pp. 7–12.

[83] A. D. Conceicao, J. Li, D. Florencio, and F. Kon, "Is IEEE 802.11 ready for VoIP?" in *IEEE Workshop on Multimedia Signal Processing*, 2006, pp. 108–113.

[84] Q. Wan and M. H. Du, "Queueing analysis and delay mitigation in the access point of VoWLAN," in *IEEE Int. Symposium on Communications and Information Technology*, vol. 2, 2005, pp. 1160–1163.

[85] Y. Zhang and B. H. Soong, "The effect of unreliable wireless channel on the call performance in mobile networkieee," *Trans. on Wireless Communications*, vol. 4, pp. 653 – 661, 2005.

[86] R. Jain, *The Art of Computer Systems Performance Analysis.* John Wiley and Sons, 1991.

[87] S. G. et al, "Performance optimizations for deploying voip services in mesh networks," *IEEE JSAC*, vol. 24, no. 11, pp. 2147 – 2158, 2006.

Publications

Journals

[1] Ahmed Shawish, Xiaohong Jiang, Pin-Han Ho, and Susumu Horiguchi, "Wireless Access Point Voice Capacity Analysis and Enhancement Based on Clients Spatial Distribution," , in press.

[2] Ahmed Shawish, Xiaohong Jiang, and Susumu Horiguchi, "Fair Scheduling for Delay-Sensitive VoIP Traffic," *IEICE Transactions on Communications*, in press.

[3] Ahmed Shawish, Xiaohong Jiang, and Susumu Horiguchi, "An Enhanced Bidirectional Transmission MAC Protocol for VoIP Support in WLANs," *IEICE Transactions on Communications*, under review.

[4] Ahmed Shawish, Xiaohong Jiang, Susumu Horiguchi, and Minyi Guo, "A New Medium Access Control Protocol for VoIP support in IEEE 802.11 Wireless LANs," *IEEE Transaction On Vehicular Technology*, under review.

[5] Keyvan Kashkouli Nejad, Ahmed Shawish, Xiaohong Jiang, and Susumu Horiguchi, "Hybrid Packet-Pheromone-Based Probabilistic Routing for Mobile Ad Hoc Networks," *IEICE Transactions on Communications*, under review.

[6] Keyvan Kashkouli Nejad, Ahmed Shawish, Xiaohong Jiang, and Susumu Horiguchi, "Dynamic Probabilistic Proactive Routing for Mobile Ad Hoc Networks," *IEICE transactions on communications*, under review.

Refreed Conference Papers

[7] Ahmed Shawish, Xiaohong Jiang, Pin-Han Ho, and Susumu Horiguchi, "Voice Capacity Analysis and Enhancement in Wireless LAN," in *Proc. of the IEEE WCNC*, USA, April 2008.

[8] Ahmed Shawish, Xiaohong Jiang and Susumu Horiguchi, "Capacity Evaluation of VoIP in Wireless LAN," in *Proc. of the IEEE workshop HPSRN*, Japan, March 2008.

[9] Ahmed Shawish, Xiaohong Jiang and Susumu Horiguchi, "Efficient Scheduler for the Growing VoIP Traffic," in *Proc. of the IEEE Int. Conf. on Parallel Processing*, China, August 2007.

[10] Ahmed Shawish, Xiaohong Jiang and Susumu Horiguchi, "New Scheduling Architecture for the Growing VoIP Traffic," in *Proc. of workshop HPSRN*, Japan, Mar 2007.

[11] Ahmed Shawish, Xiaohong Jiang and Susumu Horiguchi, "Efficient and Fair Scheduling for VoIP Traffic," in *Int. Symposium on Broadband Access Technologies*, Canada, October 2006.