國立臺灣大學資訊管理研究所碩士論文

指導教授：林永松　博士

IEEE 802.11 無線網路免競爭期間允入控制之最佳化

Optimization of Call Admission Control in IEEE 802.11 WLAN Contention-Free Periods

研究生：施翔騰　撰

中華民國九十三年七月

IEEE 802.11 無線網路免競爭期間
允入控制之最佳化

Optimization of Call Admission Control
in IEEE 802.11 WLAN Contention-Free Periods

本論文係提交國立台灣大學
資訊管理研究所作為完成碩士
學位所需條件之一部分

研究生：施翔騰　撰

中華民國九十三年七月

僅以此篇論文獻給我的母親～

林雪珠　女士

謝　誌

隨著本篇論文的完成，兩年資管所時光宛如電影片段般在腦海中瞬間飛快消逝。回想撰寫論文的過程中，同時攻讀電機所碩士學位，雙重的學業壓力時常讓自己身心俱疲，而許多人的支持與鼓勵是我堅持下去完成論文的動力來源。

首先，我要感謝恩師林永松老師，他總是循序漸進地引領著我找到問題的核心，並在關鍵處提點我思考的方向，讓我能跳出自己思考侷限的象牙塔，在最後順利完成論文；還有口試委員：莊裕澤主任、蔡益坤教授寶貴的建議與指教，讓本篇論文更臻完善。

另外感謝柏皓學長，在各方面給予我的協助與指導，讓我能有信心地繼續走下去。還有與我一起努力的同學們，瑩珍、坤威、育先、耿宏、俊達、明立、榮耀、閔元、大鈞、柏鈞，以及系上的其他同學，讓我的研究所生活充滿了無限的美好回憶。尤其是瑩珍與友群，謝謝你們的支持與鼓勵，讓我能在辛苦的論文寫作過程中，找到堅持下去的信心。還有感謝實驗室的學弟妹們，建宏、孝澤、書平、琳智、明源，平時的幫忙以及在論文口試期間的協助，讓我們可以全心準備口試。

此外，辛苦養育我的母親更是我在面對各種挫折時，最溫暖的避風港；還有妹妹適時的鼓勵，都讓我在龐大的論文壓力下，能重新調適自己的心情，再一次面對挑戰，並順利完成論文。

最後，感謝所有曾經幫助我的人，不論是在學業上、生活中，少了你們的協助，就沒有今天這篇論文的產生；多了你們的參與，也讓我的生命更加豐富精彩。要感謝的人真的太多了，不如就謝天吧！

施翔騰　謹識
于台大資訊管理研究所
民國九十三年七月二十五日

論文摘要

論文題目：IEEE 802.11無線網路免競爭期間允入控制之最佳化

作　　者：施翔騰　　　　　　　　　　　　　　　　　九十三年七月

指導教授：林永松　博士

隨著網際網路的快速發展，許多的資料傳輸相關應用已是人們日常生活中不可或缺的一部分。而無線網路所提供不受實體限制的便利性，更讓行動通訊成為當前炙手可熱的議題之一。

由於市場的快速成長，在無線區域網路 (wireless LAN)上提供即時及多媒體服務的需求與日遽增，因此相關的服務品質保證 (Quality of Service, QoS)成為是否可讓使用者得到較佳傳輸品質的重要關鍵。然而無線網路與傳統有線網路在本質上就有許多截然不同的特性；加上在 IEEE 802.11 標準中，主要是以載波偵測多重存取及碰撞避免法 (CSMA/CA：Carrier Sense Multiple Access with Collision Avoidance) 為其媒介存取控制協定 (Medium Access Control protocol)。在這樣的前提之下，需要一套更有效率的機制用以提供相關的服務品質保證。

在本篇論文中，我們提出將 IEEE 802.11 WLAN 的免競爭期間 (contention-free period) 以時槽 (time slot) 方式加以切割的概念，並配合 IEEE 802.11e 的標準，考量四種不同優先等級的資料訊框型態 (data frame type)。因此，在有限的無線頻譜 (wireless spectrum) 資源下，不同的時槽分配策略 (policy) 及允入控制 (call admission control) 機制將會導致系統有不同的收益、效能及服務品質。再者，使用者 QoS 的需求與系統收益常存在著互為消長

的關係，因此我們希望能在提供不同資料訊框差異化品質保證 (Differentiated Service, DiffServ) 的同時，最佳化整個系統的長期收益。

本研究針對 IEEE 802.11 免競爭期間時槽分配的問題，提出兩種數學模型。兩者的目的皆為在給定不同資料訊框的流量參數後，找出一個最好的時槽分配策略，藉以最佳化系統的長期收益。兩個模型的主要差別在於時間的型態，第一個模型是考慮連續的時間 (Continuous-Time) ，而第二個模型是考慮離散 (Discrete) 的形式。由於這兩個數學模型的結構及問題規模的特性，我們利用馬可夫決策過程 (Markovian Decision Process) 來解決我們的問題。

由優異的實驗數據結果顯示，我們成功地以馬可夫決策過程在我們提出的數學模型下，找到使整體系統長期收益最大化的時槽分配原則。而建構於該基礎上的解題程序比起一般的經驗法則，更可使收益達到數倍以上的成長，證明我們的方法可以提供系統營運者及網路規劃人員良好的決策。

關鍵詞：<u>**無線區域網路、服務品質保證、CSMA/CA、媒介存取控制協定、IEEE 802.11、IEEE 802.11e、時槽分配、允入控制、DiffServ、馬可夫決策過程**</u>

THESIS ABSTRACT

GRADUATE INSTITUTE OF INFORMATION MANAGEMENT
NATIONAL TAIWAN UNIVERSITY

NAME: HSIANG-TENG SHIH MONTH/YEAR: JULY, 2004
ADVISER: YEONG-SUNG LIN

Optimization of Call Admission Control in IEEE 802.11 WLAN Contention-Free Period

With the rise of the Internet, now many data transfer applications are essential to people’s daily life. In addition, wireless networks can support people’s mobility to access information regardless of where they are. Hence mobile communication has become a popular topic in today’s technology world.

Due to the rapid growth of the wireless LAN market, the need of transmitting real-time and multimedia traffic, such as voice, images, video and …, etc, over wireless LAN will gradually increase. Therefore, the relevant Quality of Service (QoS) problem has also become a critical issue. However, as there are some inherent differences between wireless networks and traditional wired networks. So with this pre-determined condition, such as using CSMA/CA in its MAC protocol under IEEE 802.11, we will need more and more effective mechanism to provide QoS assurance.

In this thesis, we bring up the concept of slotting the contention-free period in IEEE 802.11 WLAN. In addition, to be compatible with IEEE 802.11e, we consider four data frame types with different priorities. As a result, in the limited wireless spectrum resource, different slot allocation policy will generate results varying in revenue, throughput and QoS of the system. However, there is always a tradeoff between the users' QoS requirement and the system revenue. Therefore, we hope to provide differentiated service while maximizing the long-term system revenue.

We propose two mathematical models to solve the slot allocation problem in this thesis. The goal of our model is to find the best slot allocation policy to maximize the long-term system revenue under the capacity constraint. The main difference is the time type. The first model is continuous-time, while the second one is discrete-time. We apply Markovian Decision Process to deal with our problem due to the problem size and the structure of our model.

According to the good computational results, we can successfully find the best slot allocation policy that maximizes the long-term system profit by Markovian Decision Process. Compared with the heuristics that venders often use, the policy we find has great improvement in the system revenue. Therefore, our model can indeed provide good decision for system venders and network planners.

Keyword: WLAN、Quality of Service、CSMA/CA、MAC Protocol、IEEE 802.11、IEEE 802.11e、Time Slot Allocation、Call Admission Control、DiffServ、Markovian Decision Process

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Background

With the rise of the Internet, applications based on network services have grown rapidly. People use the functions of data transfer (WWW, E-mail, ftp and …, etc) to deal with daily routines. But because of the physical limitations of wired networks, more and more efforts are focusing on the development and establishment of wireless networks.

Wireless networks allow people to enjoy the convenience of accessing information freely regardless of where they are. In addition, wireless networks provide a diverse range of applications, such as data transfer, user positioning …, etc. So in the future the integration of wired and wireless networks will be an important trend (Figure 1.1), with wireless networks playing the critical role as the last hop for users to access the Internet.

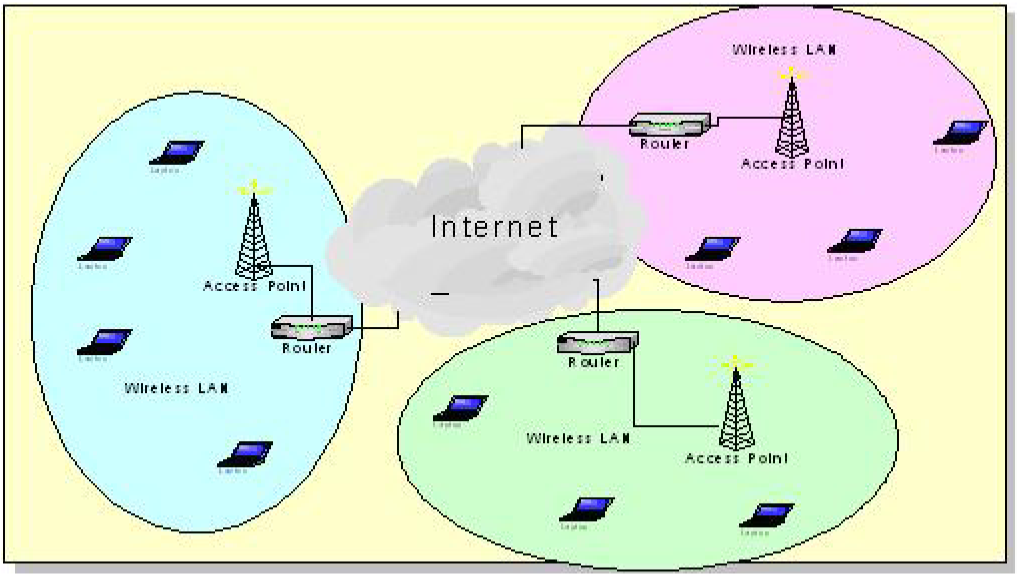


Figure 1.1 Integration of wireless networks and wired networks

From the evolution of the IEEE 802.11 standard (shown in Table 1.1), it’s clear that the transmission rate has not yet evolved to its potential ceiling.

	802.11	802.11b	802.11a
Frequency	2.4GHz	2.4GHz	5GHz
Transmission rate	2Mbps	11Mbps	54Mbps
Layer 3 transmission rate	1.2Mbps	5Mbps	32Mbps
Medium Access Control/Media Sharing	CSAM/CA	CSMA/CA	
Connectivity	Conn-less	Conn-less	Conn-less
Multicast	Y	Y	Y
QoS support	Y	Y	Y
Spread spectrum	DSSS	DSSS	Single carrier
Data encryption	40 bit RC4	40 bit RC4	40 bit RC4
Network category	Ethernet	Ethernet	Ethernet
Management	802.11 MIB	802.11 MIB	802.11 MIB
Wireless connectivity quality control	NO	NO	NO

Table 1.1 The evolution of the IEEE 802.11 standard

Therefore, transmitting real-time or multimedia traffic, such as voice, images, video …, etc, over wireless networks will become increasingly popular. However, as a result of the inherent differences between wireless networks and traditional wired networks, there are some very significant differences in the design of relevant algorithms for solving the Quality of Service (QoS) problems. Therefore, the traditional mechanisms which were used extensively in wired networks are now ineffective in today's wireless networks.

The main characteristics of wireless networks are as follows:

1. **High channel variability:** Because wireless networks transmit over a wireless medium, signals can be easily affected by external factors, such as fading, noise, mobility, and interference …, etc. As a result, there is a higher rate of loss and error in packet transmission. Also, location-dependent errors and signal problems can occur as the use of wireless devices shift from location to location. All the above factors generate fluctuations in channel availability, which are problems encountered by the wired network. Therefore, this highly inaccurate environment has rendered many previous research and data inefficient.
2. **Limited bandwidth:** Compared to the wired network, the usable bandwidth in a wireless network is relatively small. Take today's most popular standard, IEEE 802.11b, for example, it can only provide a transmission rate of 11 Mbps. Even with the faster IEEE 802.11a running at 54 Mbps, the transmission rate is still slower than the wired Ethernet which performs at 100 Mbps. Consequently, better bandwidth management has also become more important in a wireless network environment.
3. **Power constraint:** To achieve the convenience of wire-less communications, most wireless devices are battery-powered. For this

reason, power-saving is another vital issue in wireless networks. Under these conditions, every algorithm designed for use in wireless networks should be simple, and should also reduce unnecessary transmissions between mobile devices.

In short, the major problem of real-time multimedia transmission in wireless network today results from limited wireless bandwidth and ever-changing circumstances. Thus, we can’t guarantee QoS, such as bandwidth, bounded delay, delay jitter …, etc, for traffic with different service levels.

Within many wireless network standards, IEEE 802.11 Wireless LAN standard [3] proposed by the IEEE Computer Society has proved to be the most popular. Therefore, IEEE 802.11 wireless LAN is the focus of this thesis.

1.2 Motivation

As the demand for wireless communication services continues to grow, how to increase the system capacity under the limited spectral resources has become more and more important. Recently, the integrated multimedia network has been the key in the communication system, so now how to communicate under the integrated traffic becomes more important. To meet this demand, the resource reservation protocol needs to be designed such that mobile terminals can share the limited communication bandwidth in an efficient manner. This way, a wide variety of Quality of Service (QoS) requirements can be flexibly controlled. Thus, to fulfill the integrated traffic requirement within a limited bandwidth, an efficient call admission control mechanism and multiple access scheme shall be required.

There are three basic components of admission control schemes: traffic descriptors, admission criteria, and a measurement process. Figure 1.2 illustrates the relationship between these three components.

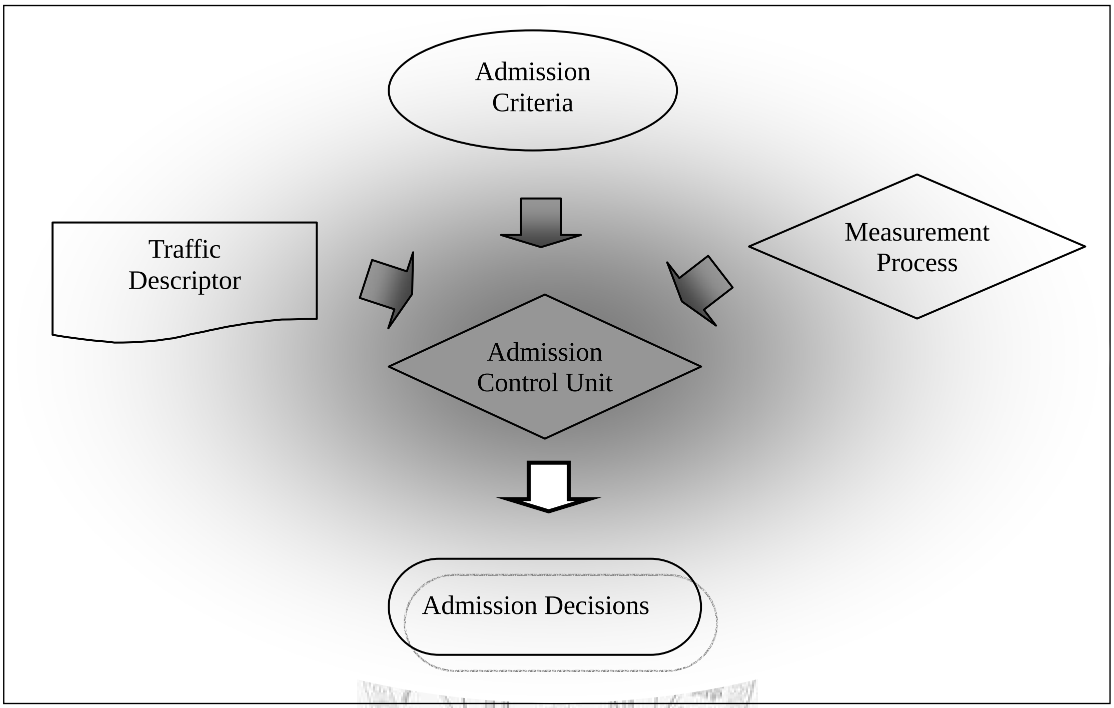


Figure 1.2 The relationship between basic components of admission control

First, we have to define different traffic parameters for each service level to obtain the traffic descriptor. We then set up some admission criteria to decide whether or not to accept a new request. Next, the admission control unit can compute a specific value by the measurement process and compare it with the admission criteria to yield a suitable admission decision. However, there is still no standard accurately defined for call admission control mechanism of IEEE 802.11 nowadays.

Furthermore, the basic communication protocol used in IEEE 802.11 [3] MAC layer is CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). It's a probabilistic medium access control method. Every wireless mobile host of the same status competes for the right to use the wireless medium with the same probability. Under this contention mode, it cannot prevent the collision phenomenon absolutely. In addition, with the increase of the mobile hosts, the loss and collision probability will rise abruptly, so that the CSMA/CA can’t guarantee transmitting packets in the bounded time to meet the need of the real-time multimedia applications.

Nevertheless, as mentioned before, with the prevalence of wireless networks, the demands of QoS for all sorts of applications will enormously increase. Hence we need more efficient mechanisms to achieve this goal.

1.3 Objective

We focus our research on infrastructure mode in wireless LAN (WLAN) since there is more randomness in the Ad hoc mode and it’s easier to get the whole information about the traffic transmission condition in the former.

There are two objectives in this thesis.

Objective 1: Derive the analytical model that can describe the wireless call admission and resource allocation problem in infrastructure mode precisely.

Although the wireless QoS issues have attracted much attention, there are still fewer researches to deal with the problem in infrastructure mode with an accurate mathematical model. In addition, we want to develop an appropriate model to solve the wireless call admission and resource allocation topic in an analytic way.

In brief, this thesis will model the problem of the wireless communication networks operational call admission control issues. We can describe the operational support and capacity management as: wireless network planning, traffic analysis of mobile data, performance optimization, network monitoring, network capacity expansion, and network servicing. Figure 1.3 shows the relationship among these issues.

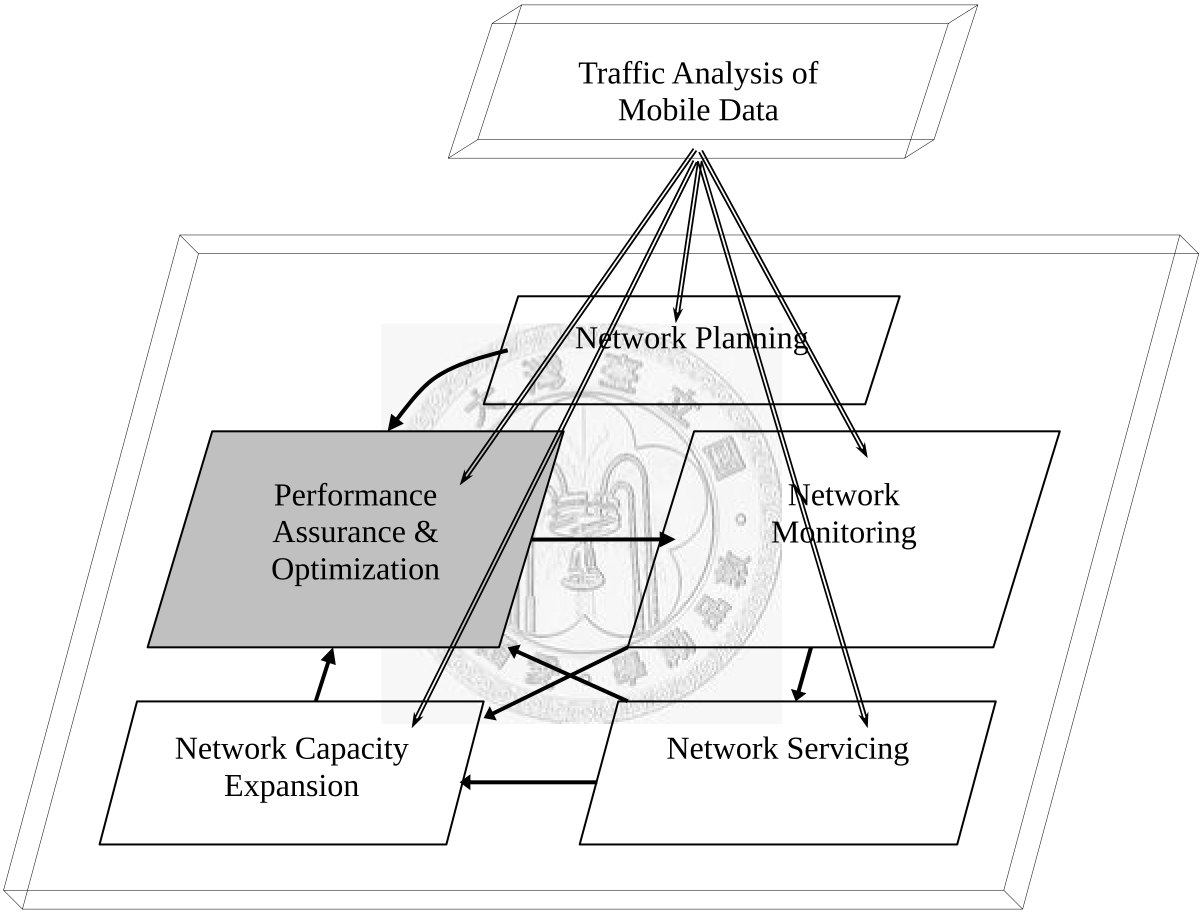


Figure 1.3 Operation support and capacity management model

As a result of much randomness in our research problem, it can't be accurately described by an average model. In order to deal with such kind of complex problem, we adopt ***Stochastic Process (or Markov Process)*** [17] to model each random element and decision process in our system.

Objective 2: Apply suitable methods to the mathematical model obtained above to compute the best call admission policy which achieve the QoS demands and maximize the system revenue at the same time.

After modeling the problem successfully, we will show how the problem can be solved in order to find the best policy. What's important is that users only focus on whether the QoS demands are achieved or not. Also, with different resource allocation policies, different revenue, throughput, and QoS of the system will be realized. Nonetheless, in the vender's point of view, revenue maximization is the main consideration, and a good criterion of call admission control would bring a satisfactory sum of revenue for the system. Therefore, in this thesis, we will take both QoS constraints and system revenue into consideration and find out the best policy to fulfill maximum system revenue considering QoS demands of different service level.

Because our problem size is too large, it's inefficient to compute all possible conditions (i.e. exhaustive search). Consequently, we adopt a dynamic programming method, ***Markovian Decision Process (M.D.P)*** [7] [18], to get the optimal call admission policy.

1.4 Proposed Approach

We will model the WLAN call admission control and resource allocation problem [9] as a stochastic process. In addition, we will employ the concepts of time slotting [13] and burst transmission [14] in our formulation.

In this thesis, we will consider four kinds of traffic: best effort, background, voice, and video. Because there are two periods, contention period (CP) and contention-free period (CFP), in 802.11 WLAN architecture, we will only consider the call admission problem in the contention-free period and divide it into lots of time slots as shown in Figure 1.4.

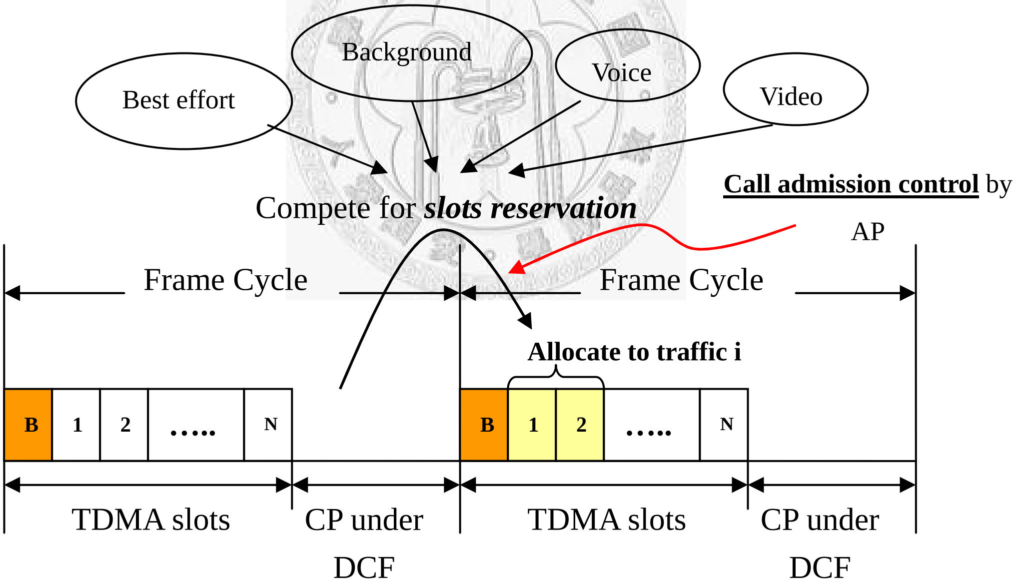


Figure 1.4 Illustration of slots allocation

We assume that the traffic arrivals can be characterized by a Poisson process, and the service time of each traffic is exponentially distributed. Under this assumption and model, we apply *Markovian Decision Process (M.D.P)* [7] [8] [18] to find the best call admission policy. Since discrete-time M.D.P and continuous-time M.D.P are different, we will use both ways to model and solve the problem in this thesis.

We list a simple example to explain our model. Assume there are two kinds of users: voice users and best effort users with the voice users having a higher arrival rate, higher service rate and higher revenue. Thus voice traffic will generate higher revenue. This is assuming that there are 8 slots in contention-free period in our system. When the system is empty, it will admit both users because the system loading is low. But when the system load is high, for example, only 2 slots are available, the system may reject best effort users because admit them will decrease long-term revenue. In this kind of situation, the system will only admit voice users due to their higher revenue rate. We want to find the best slots allocation (or call admission) policy to maximize the system revenue. It’s the goal of this thesis.

1.5 Thesis Outline

The remainder of the thesis is organized as follows. We will review the IEEE 802.11 WLAN MAC architecture, emerging IEEE 802.11e draft amendment for original specification to support QoS, and other relevant WLAN QoS issues in Chapter 2. Because we can formulate our problem in both continuous-time and discrete-time cases, we will use both ways to solve the problem. In chapter 3, we will propose both continuous-time mathematical model and discrete-time mathematical model for finding the best policy of call admission (slot allocation). Our goal is to maximize the long-term system revenue while achieving the QoS constraints. The details of mathematical formulation process will also be discussed. Chapter 4 illustrates the solution approach for this problem. Chapter 5 validates the accuracy of our model by programming implementation. Besides, we give several examples in different situation to show the meanings behind data. Finally, Chapter 6 summarizes the thesis and gives our conclusion.

Chapter 2. Literature Survey

2.1 IEEE 802.11 Wireless LAN

802.11 is a member of the IEEE 802 family, which is a series of specifications for local area network (LAN) technologies. Figure 2.1 shows the relationship between the elements of 802.11 and their place in the OSI model. IEEE 802.11 specification is focused on the functional definition of media access control (MAC) sublayer and physical (PHY) layer. The MAC is a set of rules to determine how to access the medium and send data, but the detail of transmission and reception are left to the PHY.

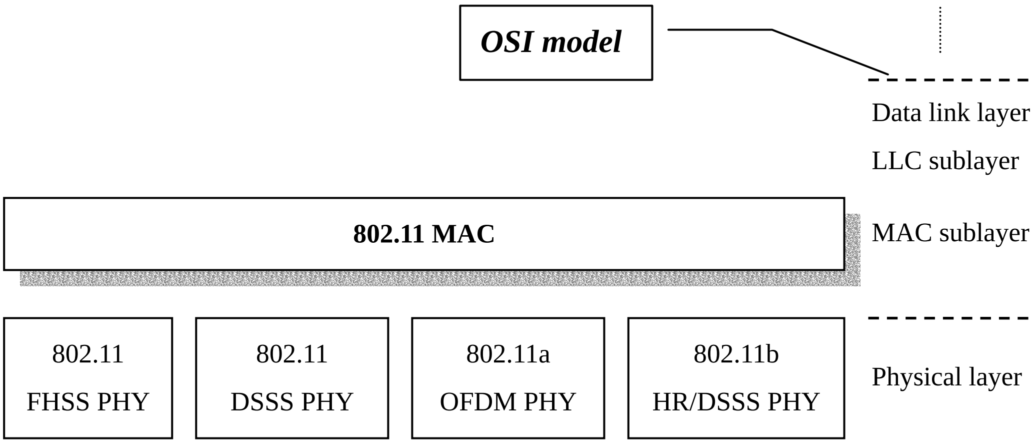


Figure 2.1 802.11 family and its relation to the OSI model

This thesis is focused on how to develop an efficient call admission control mechanism over MAC layer. Therefore, we will introduce the basic wireless network architecture and MAC operations in this section.

2.1.1 Ad Hoc Network and Infrastructure Network

In order to support all kinds of needs while setting up WLAN, there defines two optional WLAN architectures to be chosen, namely Ad hoc mode and infrastructure mode, as shown in Figure 2.2.

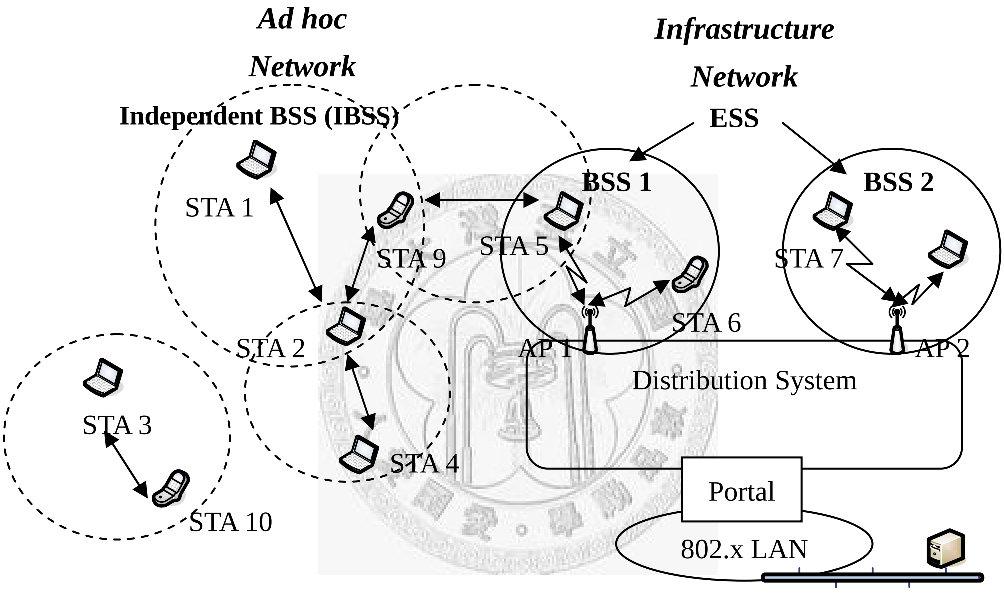


Figure 2.2 Complete IEEE 802.11 architecture

The basic building block of an 802.11 network is the basic service set (BSS), which is simply a group of stations that communicate with each other. Ad hoc mode provides STAs point-to-point communication. The receiving STA receives the data transmitted from the transmitting STA directly and this transaction is not permitted to use any intermediary STA to pass the data. As a result, they must be within direct communication range. The wireless Ad hoc mode would have more flexibility, but less extensibility. Thus the WLAN using Ad hoc mode is appropriate to temporary occasions such as temporary conferences. When the conference ends, the Ad hoc network is dissolved. Due to their short duration, small size, and focused purpose, we call it **Ad hoc** networks.

The infrastructure mode of WLAN is a distribution networking system and an access point (AP) placed in infrastructure area is responsible to control which STA has the right to access media for transmitting data. The set of an AP and the STAs which are controlled by that AP is called as (infrastructure) BSS and several BSSs are integrated as extended service set (ESS) to connect to the whole distribution networking system. Consequently, the STAs placed at different BSS would transmit data via the connectors or intermediaries (such as APs) of distribution networking system. Furthermore, the whole WLAN shall connect to wired-LAN via the Portal connector to set up the whole network. Since the popular protocol of wired-LAN is Ethernet, the AP substitutes for the Portal. This thesis would focus on infrastructure mode of WLAN.

2.1.2 Medium Access Control Protocol

The key of the 802.11 specification is the MAC. It rides on every physical layer and controls the transmission of user data into the air. Access to the wireless medium is controlled by the ***coordination functions***. 802.11 specify two mechanisms to access the wireless medium. One is ***Distributed Coordination Function (DCF)***, and the other is ***Point Coordination Function (PCF)***. The MAC architecture can be described as shown in Figure 2.3 as providing the PCF through the services of the DCF.

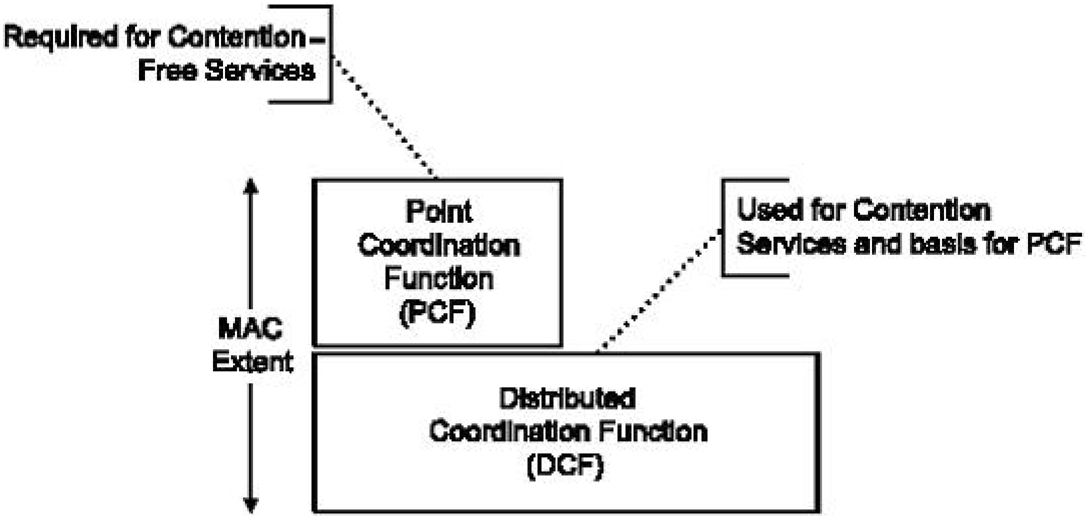


Figure 2.3 MAC architecture [3]

2.1.2.1 Distributed Coordination Function (DCF)

The fundamental access method of the IEEE 802.11 MAC is a DCF known as carrier sense multiple access with collision avoidance (CSMA/CA). The DCF shall be implemented in all STAs, for use within both IBSS and infrastructure network configurations.

Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)

There is only one single channel provided for data transmission in 802.11 environment. As a result, in order to reduce unnecessary collisions, like Ethernet, it first checks to see that the radio link is clear before transmitting. Physical and virtual carrier-sense functions are both used to determine the state of the medium. A physical carrier-sense mechanism determines the medium condition by checking whether the strength of signal beyond the threshold or not. A virtual carrier-sense mechanism is refereed to as the network allocation vector (NAV) which may be thought of as a counter. It counts down to zero at a uniform rate. When the counter is zero, the virtual carrier-sense indication is that the medium is idle.

But the wireless 802.11 MAC protocol implement collision avoidance instead of collision detection. There are a couple of reasons for this:

1. 802.11 does not have the ability to both send and receive at the same time.
2. More importantly, even if one had collision detection and sensed no collision when sending, a collision could still occur at the receiver results from hidden terminal problem and multipath fading effect.

Interframe Space (IFS)

The time interval between frames is called the IFS. 802.11 defines different waiting times upon different kind of frames and STA is allowed to transmit its frame until the corresponding waiting time is expired. Therefore, varying interframe spacings create different priority levels for different types of traffic. 802.11 uses four different interframe space. Three are used to determine medium access. They are listed in order, from the shortest to the longest. And Figure 2.4 shows some of these relationships.

a) SIFS short interframe space
b) PIFS PCF interframe space
c) DIFS DCF interframe space
d) EIFS extended interframe space

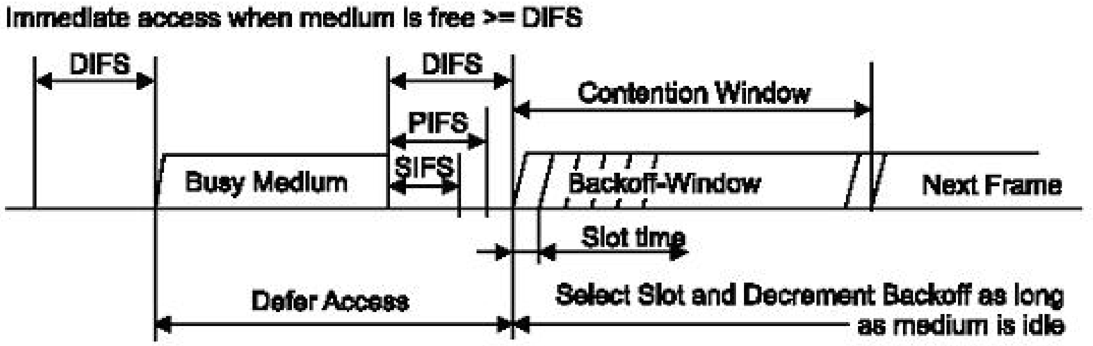


Figure 2.4 Interframe spacing relationships [3]

SIFS shall be used for a positive ACK frame, a RTS/CTS frame…, etc, which are the highest-priority transmissions, to avoid other STAs to take away the access right while the communication is not over. PIFS is used only by STAs operating under the PCF to gain priority access to the medium ate the start of the contention free period (CFP). DIFS is used by STAs operating under DCF to transmit data fames (MPDUs) and management frames (MMPDUs). EIFS is not a fixed interval and used by the DCF whenever the PHY has indicated to the MAC that a frame transmission was begun that did not result in the correct reception of a complete MAC frame with a correct frame check sequence (FCS) value.

However, the probability of collision occurred by frames with the same priority of IFS is exist. In order to solve this condition, IEEE 802.11 defines a random period of time (i.e. random backoff time) to wait after IFS expired in the DCF mechanism.

Random Backoff Time

A STA desiring to initiate transfer of data MPDUs and/or management MMPDUs shall invoke the carrier-sense mechanism to determine the state of the wireless medium. If the medium is busy, the STA shall defer a DIFS or EIFS idle time and then generate a random backoff period for an additional deferral time before transmitting. The following equation is used to calculate the backoff time.

$$Backoff\ \ Time = Random() \times aSlotTime$$

where

Random() = Pseudorandom integer drawn from a uniform distribution over the interval [0,contention window (CW)], where CW is an integer within the range of values of the PHY characteristics aCWmin and aCWmax.

aSlotTime = The value of the correspondingly named PHY characteristic.

The CW parameter shall take an initial value of aCWmin. It will double after every unsuccessful transmission until the CW reaches the value of aCWmax. As a result, the set of CW values shall be sequentially ascending integer powers of 2, minus 1, beginning with a PHT-specific aCWmin value, and continuing up to and including a PHY-specific aCWmax value as shown in Figure 2.5.

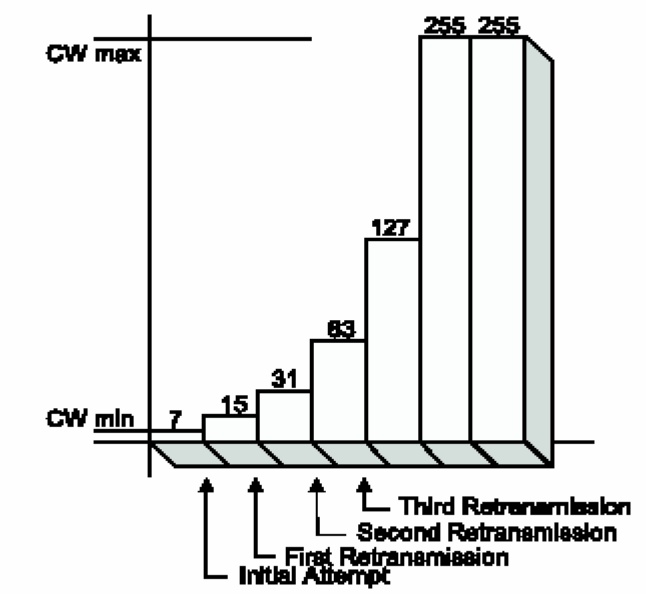


Figure 2.5 An example of exponential increase of CW [3]

DCF Access Procedure

Before attempting to transmit, each station checks whether the medium is idle. If the medium is not idle, stations defer to each other and employ an orderly exponential backoff algorithm to avoid collisions.

In summary, the rules for all transmissions using DCF are as follows:

1. If the medium has been idle for longer than the DIFS, transmission can begin immediately. Carrier sensing is performed using both a physical medium-dependent method and the virtual (NAV) method.
    a. If the previous was received without errors, the medium must be free for at least the DIFS.
    b. If the previous transmission contained errors, the medium must be free for the amount of EIFS.
2. If the medium is busy, the station must wait for the channel to become idle. If access is deferred, the station waits for the medium become idle for the DIFS and prepares for the exponential backoff procedure.
3. Extended frame sequences (RTS/CTS exchange) are required for higher-level packets that are larger than configured threshold.

2.1.2.2 Point Coordination Function (PCF)

Point coordination provides contention-free services. Special stations called point coordinators (PCs) are used to ensure that the medium is provided without contention. Point coordinators reside in access point, so the PCF is restricted to infrastructure networks.

When a STA attempts to transmit real-time data, it first associates with the PC. After association procedure, it will get a unique number, namely association ID (AID) and PC will record this AID to its polling list. Until the contention free period (CFP), PC would poll the STA in the polling list to ask if it needs to transmit any data frame. Thus STA in PCF mechanism is only permitted to transmit on condition that PC polls it. Most important of all, multiple frames can be transmitted only if the AP send multiple poll requests.

CFP Structure and Timing

The PCF controls frame transfers during a CFP. The CFP shall alternate with a CP as shown in Figure 2.6. Each CFP shall begin with a Beacon frame which contains a Delivery Traffic Indication Message (DTIM). The Beacon frame also contains a CF parameter set that records all relevant CFP information, such as CFPMaxDuration, CFPRate …, etc. The contention period must be long enough for the transfer of at least one maximum-size frame and its associated acknowledgement.

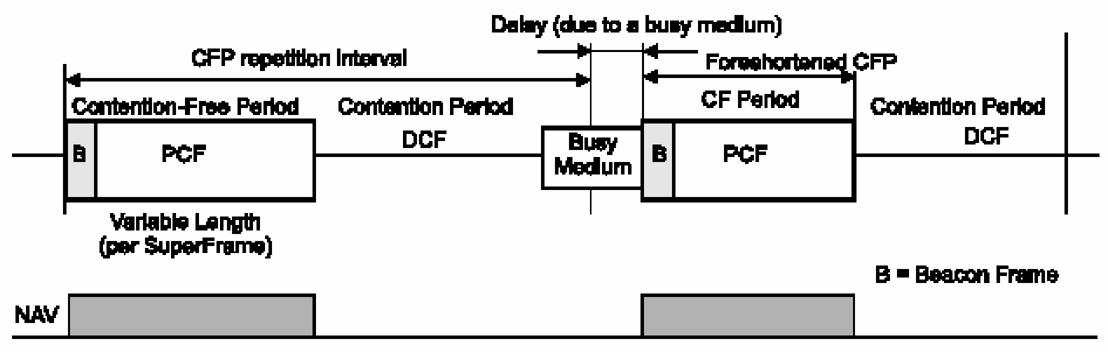


Figure 2.6 CFP/CP alternation [3]

2.2 IEEE 802.11e

To support MAC-level QoS, the IEEE 802.11 standardization committee is currently working on IEEE 802.11e [4] [15], a supplement to the original IEEE 802.11 MAC. The IEEE 802.11e MAC will support multimedia applications such as voice and video over the IEEE 802.11 WLANs.

The MAC architecture is changed as shown in Figure 2.7 as providing the PCF and hybrid coordination function (HCF) through the services of the DCF. It also defines two medium access mechanisms. Contention-based channel access is referred to as enhanced distributed channel access (EDCA), controlled channel access as HCF controlled channel access (HCCA).

With 802.11e, there may still be the two phases of operation within a superframe (i.e. CP and CFP). The EDCA is used in the CP only, while the HCCA is used in both phases. The HCF combines methods of the PCF and DCF, which is the reason it is called hybrid.

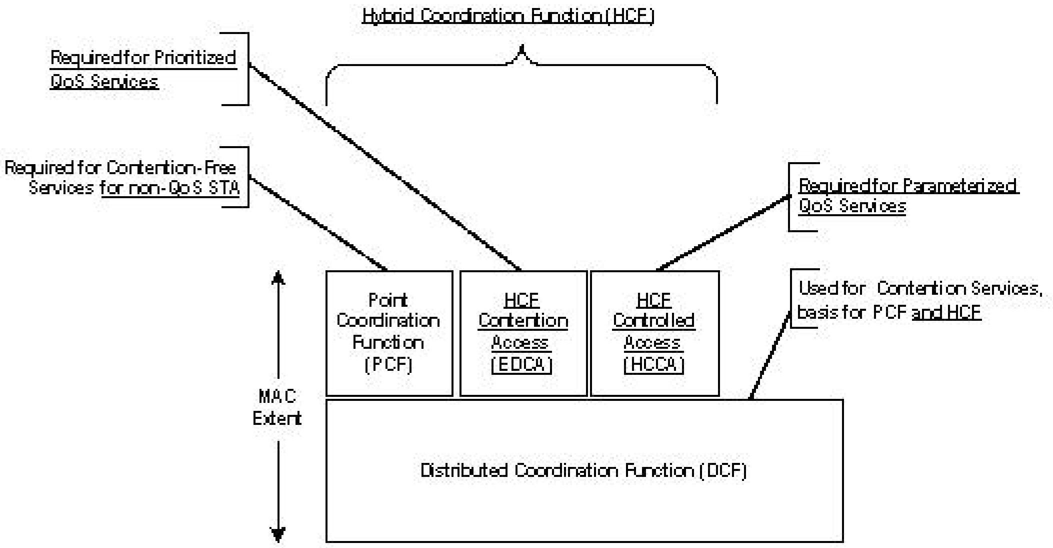


Figure 2.7 802.11e MAC architecture [4]

2.2.1 Enhanced Distributed Channel Access (EDCA)

Access Category

The QoS support in EDCA is provided by the introduction of ***access categories (ACs)*** and multiple independent backoff entities. MSDUs are delivered by parallel backoff entities within one 802.11e station, where backoff entities are prioritized using AC-specific contention parameters, called EDCA parameter set. There are four ACs; thus, four backoff entities exist in every 802.11e station. The ACs are labeled according to their target application, i.e., AC_VO (voice), AC_VI (video), AC_BE (best effort), AC_BK (background). Figure 2.8 illustrates the parallel backoff entities. The EDCA parameter set defines the priorities in medium access by setting individual interframe spaces, contention windows, transmission opportunity (TXOP) limit, and many other parameters per AC.

The EDCA access procedure is similar to the DCF access procedure. The differences between them are that within each STA, every AC has its specific parameters, such as AIFS[AC], CWmin[AC], CWmax[AC], TXOPlimit[AC] …, etc, in EDCA as shown in Figure 2.9, but there is only one priority (backoff entity) in legacy 802.11 DCF.

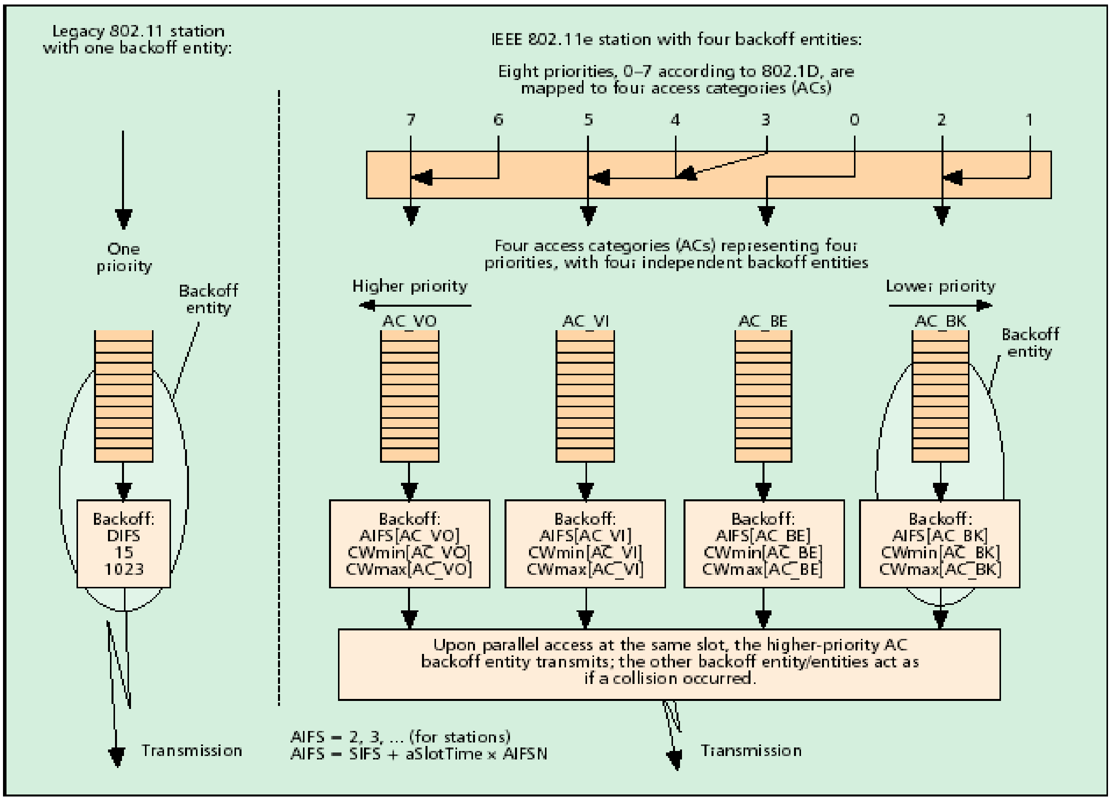


Figure 2.8 Legacy 802.11 station and 802.11e station with four ACs within one station [15]

Transmission Opportunity

When a STA obtains medium access in contention period, it will be assigned an EDCA-TXOP which is an interval of time defined by its starting time and duration during which a backoff entity has the right to deliver MSDUs separated by SIFS ***(burst transmission)***.

As described above, four backoff entities with different EDCA parameter sets reside inside an 802.11e STA. Collisions between contending channel access functions within an 802.11e STA are resolved within the STA such that the data frames from higher-priority AC receives the TXOP and the data frames from the lower-priority colliding AC(s) behave as if there were an external collision on the wireless medium ***(virtual collision)***. However, this collision behavior does not include setting retry bits in the MAC headers of MPDUs at the heads of lower-priority ACs.

It may still occur that the transmission of the backoff entity with higher-priority collides with another transmission initiated by other station.

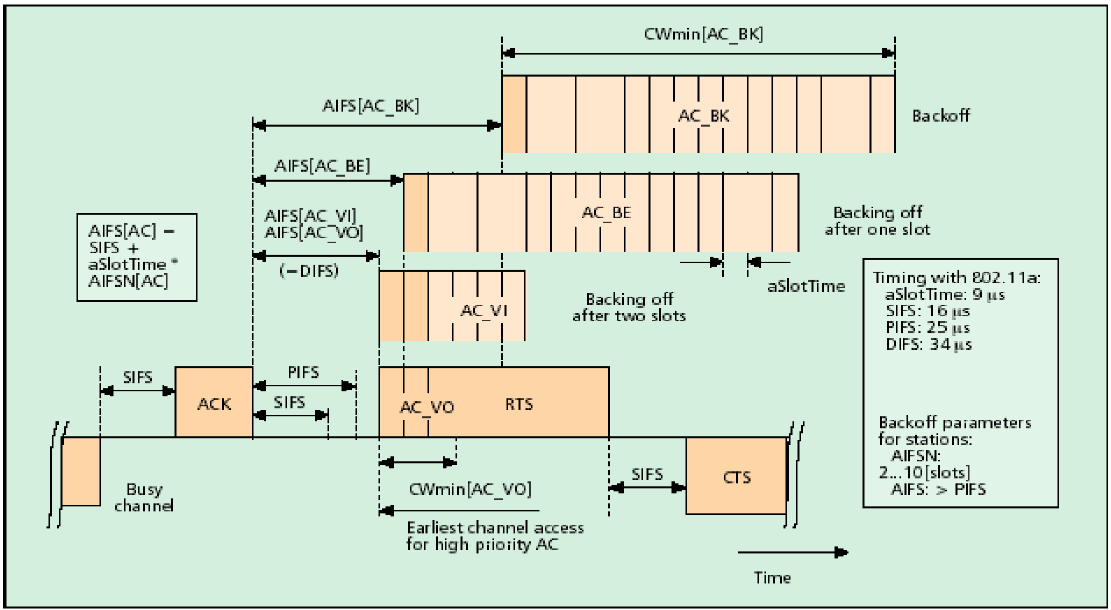


Figure 2.9 In EDCA, multiple backoff entities contend for medium access with different priorities in parallel [15]

2.2.2 HCF Controlled Channel Access (HCCA)

The HCF controlled channel access mechanism uses a QoS-aware centralized coordinator, called a hybrid coordinator (HC), and operates under rules that are different from the point coordinator (PC) of the PCF.

HCF Frame Exchange Timing

HCCA extends the EDCA access rules by allowing the highest priority medium access to the HC during both the CFP and CF as shown in Figure 2.10.

During CP, each TXOP of an 802.11e STA begins either when the medium is determined to be available under the EDCA rules, that is, after AIFS[AC] plus the random backoff time, or when a backoff entity receives a polling frame, the QoS CF-Poll, from the HC, so as to receive limited-duration controlled access phase (CAP) for contention-free transfer of QoS data. The QoS CF-Poll from the HC can be transmitted after a PIFS idle period, without any backoff, by the HC.

During CFP, 802.11e backoff entities will not attempt to access the medium without being explicitly polled, hence, only the HC can allocate TXOPs by transmitting QoS CF-Poll frames.

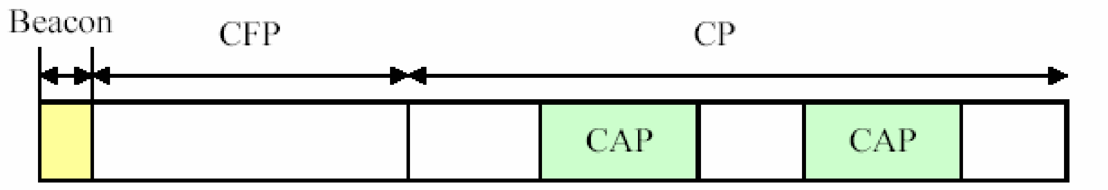


Figure 2.10 The superframe interval defined in HCF

Transmission Opportunity

When a STA receives QoS CF-Poll from the HC, it will be assigned a HCCA-TXOP, or a polled TXOP. During a polled TXOP, a polled STA can transmit multiple frames ***(burst transmission)*** that the STA selects to transmit according to its scheduling algorithm, with a SIFS time gap between two consecutive frames as long as the entire frame exchange duration is not over the allocated maximum TXOPlimit.

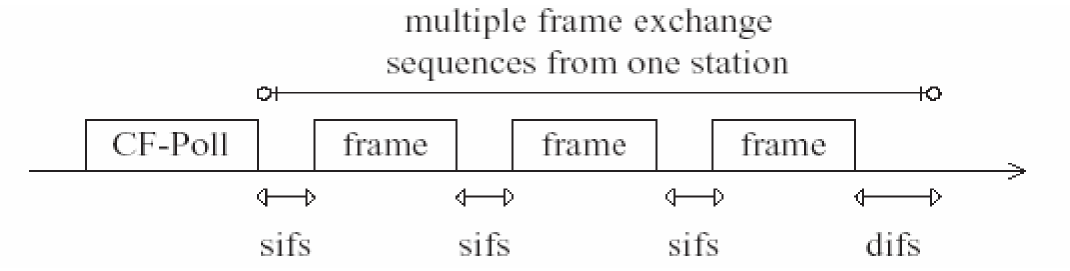


Figure 2.11 Transmission opportunity

2.3 Related Works

The current Internet architecture with its best effort service model is inadequate for new classes of applications that need QoS assurances. The basic way to provide QoS is to provide differential service. There are two approaches to achieve this target.

(i) *Reservation-based:* In this model, network resource are explicitly identified and reserved along the path from the sender to the receiver **(end-to-end)**. Network nodes classify incoming packets and use the reservations to provide **per-flow** QoS. The Integrated Service (IntServ) model [1] is based on this approach.

(ii) *Reservation-less:* In this model, resources are not explicitly reserved. Instead, traffic is differentiated into a set of classes, and network nodes provide priority-based treatment based on these classes **(per-class QoS)**. The Differentiated Service (DiffServ) model [2] is based on this approach. It is realized by mapping the Differentiated Service Code Point (DSCP) contained in the IP packet header to a particular treatment or **per-hop behavior (PHB)**, at each network node along its path.

As a result of the complex implementation and scaling problem of Integrated Service (IntServ) [12], we choose Differentiated Service (DiffServ) as the basis of our QoS solution. And some of the DiffServ techniques based on centralized or distributed situation in 802.11 WLAN are discussed as follows.

2.3.1 Distributed QoS Mechanisms in 802.11 WLAN

Differential Contention Window

In [20], different CWmin and CWmax are assigned to STAs with different priority. The experimental results show that it’s efficient to provide service differentiation. [5] brings up the concept of backoff increase function which is also based on scaling the contention window according to the priority of the traffic.

Differential Interframe Spaces

[5] assigns different interframe spacings to different users. If a packet in a STA waits a shorter IFS time, then it has higher probability to be prior transmitted than other packets that wait a longer IFS time.

However, a fixed and larger IFS distance would suffer from degradation of throughput whenever there are only low priority classes inside the serving area, and a smaller IFS distance would cause the confusion of service differentiation.

Differential Frame size

[5] uses different maximum frame lengths for different users. But the longer frame would have a higher probability of suffering from interference. In addition, this mechanism will not be helpful in real-time data transmission, because it does not give

any constraints on delays.

All above concepts have been included in the IEEE 802.11e, and this thesis will follow the definition of the 802.11e specifications. In order to make the proposed mechanism compatible with the EDCA, we will use the relevant values defined in the specifications as the basis of our research.

2.3.2 Centralized QoS Mechanisms in 802.11 WLAN

Distributed TDM

This mechanism uses polling-like regular PCF, but combined with this technique, we can set up Time Division Multiplexing (TDM)-like time slot periods, and specify which station gets which time slot.

Distributed fair scheduling

It's not always desirable to completely sacrifice the performance of low priority traffic in order to give very good service to high priority traffic. [19] proposed DFS which applies the ideas behind fair queueing in the wireless domain. Each flow gets bandwidth proportional to some weight that has been assigned to it. By using different weight, DFS can achieve differentiation between flows. The simulation results in [11] show that DFS can give a relative differentiation and consequently avoids starvation of low priority traffic.

Different Polling Scheme

In [21], several polling schemes are introduced. Round-Robin scheme finds the lowest address of STAs, and then, after checking whether there is any data for this address in its queue, determines the type of frames to transmit. It can provide fairness, but may result in lower wireless link utilization.

First-In-First-Out (FIFO) scheme poll the STAs according to the order of frames in the PC's queue. It's easily implemented. Furthermore, it saves PC the additional cost and time-consumed, while preserves the frame order in queue. However, it may cause unfairness.

Priority scheme poll the STAs in accordance with their priority. It may provide simple support of priority transmission and suitable to make arrangement for real-time transmission. Nevertheless, it may result in starvation to lower priority traffic.

For the following reasons:

1. There are fewer researches addressing the QoS issues under centralized situation.
2. It's easier to get the sufficient information about the traffic transmission status in the centralized situation.
3. Scheduling and polling mechanisms can only make decisions based on the current traffic status rather than long-term considerations.

We will propose a call admission control (resource allocation) mechanism under centralized situation to achieve the QoS constraints while taking long-term revenue into consideration.

Chapter 3. Problem Formulation

3.1 Problem Description

In the present WLAN environment, more and more efforts are focused on transmitting a variety of traffic, especially the multimedia and real-time service. Due to the scarce wireless spectrum resources and different network characteristics compared with the wired network, it’s more difficult to guarantee the QoS demands in WLAN.

In this thesis, the problem to be solved is that, how to fine the best call admission control policy to allocate the wireless medium resources to a variety of traffic so as to achieve the QoS demand and maximize the system revenue. Two models are defined. One is continuous-time case, and the other one is discrete-time case.

3.2 Derive Needed Information

In this section, we will derive the complete useful parameters to formulate our problem.

3.2.1 Parameter Sets

Transmission Opportunity Limit for All Kinds of Traffic

We adopt the concept of bursting transmission [14] to reduce the wasted wireless resource in competing for the medium access right, so it's not limited to transmit one MSDU after winning the access right. As a result, we have to find the maximum transmission time (the same as TXOP limit) for each traffic as the coefficients of our constraint. In order to be compatible with EDCA, we follow the definition in EDCA specifications. We classify the whole traffic into four types: background, best effort, voice, and video. Their access priority is: background < best effort < voice < video.

Priority	**User priority (UP – Same as 802.1D User Priority)**	**802.1D Designation**	**Access Category (AC)**	**Designation (Informative)**	
lowest	1	BK	AC_BK	Background	1. Background Traffic
↓	2	-	AC_BK	Background	
	0	BE	AC_BE	Best Effort	2. Best Effort Traffic
	3	EE	AC_VI	Video	3. Video Traffic
	4	CL	AC_VI	Video	
	5	VI	AC_VI	Video	
highest	6	VO	AC_VO	Voice	4. Voice Traffic
	7	NC	AC_VO	Voice	

Table 3.1 User Priority (UP) to Access Category (AC) mappings [4]

According to the 802.11e draft as shown in Table 3.2, we could find the TXOP limit for each class of traffic are 0, 0, 6.016, and 3.008 (ms) for background, best effort, voice, and video respectively. And a TXOP limit value of 0 indicates that a single MAC Service Data Unit (MSDU) or MAC management protocol data unit

(MMPDU), in addition to a possible RTS/CTS exchange or CTS to itself, may be transmitted at any rate for each TXOP.

A maximum MSDU is nearly 2304 bytes, and under 802.11b, the transmission rate is 11Mbps. Therefore, the TXOP limit for background and best effort traffic is calculated as follows:

2304 / (11*10^6/8) ~ 1.67563*10 ^ (-3) s = 1.6756 ms

In addition, there are lots of header fields and preambles with the MSDU, so we estimate the maximum TXOP limit for both background and best effort traffic is 2ms.

AC	**CWmin**	**CWmax**	**AIFSN**	**TXOP Limit DS-CCK/PBCC PHY**	**TXOP Limit OFDM/CCK-OFDM PHY**
AC_BK	aCWmin	aCWmax	7	0	0
AC_BE	aCWmin	aCWmax	3	0	0
AC_VI	(aCWmin+1)/2 – 1	aCWmin	2	6.016ms	3.008ms
AC_VO	(aCWmin+1)/4 – 1	(aCWmin+1) /2 - 1	2	3.008ms	1.504ms

Table 3.2 Default EDCA parameter set [4]

Consequently, we choose 2, 2, 3, and 6 (ms) as the TXOP limit of each traffic for our problem formulation.

Maximum Contention-Free Period

The IEEE document does not establish any guidelines to calculate the duration of contention-free period. As such, its value is assigned at the discretion of the HC.

However, even if the WLAN is prepared to use polling based access, the DCF (and EDCA) must still be used for management tasks like association, dissociation, reassociation, etc. If the contention-free parts are too long, this will be translated into higher handover latencies

As a result, we use the value in [6]. It assumes that handover latencies (not taking into account scanning and context transfer components) would not be greater than 10/20 ms, so it has made the multiframe length equal to 30 ms, while only 1/3 of that (i.e. 10 ms = aCFPMaxDuration) would be dedicated to polling based access.

Moreover, we divide the CFP into lots of slots. In accordance with the 802.11e draft, one time unit (TU) is 1024 μs, which is nearly 1 ms. As a result, the total CFP is 10 slots in this thesis.

3.2.2 Stochastic Process

Any realistic model of a real-world phenomenon must take into account the possibility of randomness. That is, more often than not, the quantities we are interested in will not be predictable in advance but, rather, will exhibit an inherent variation that should be taken into account by the model. This is usually accomplished by allowing the model to be probabilistic in nature. Such a model is, naturally enough, referred to as a stochastic model [17].

Stochastic processes concern sequences of events governed by probabilistic laws (Karlin & Taylor, "A First Course in Stochastic Processes", Academic Press, 1975). In our problem, sequences of events take time, so we can think on random events along the time. In order to define a stochastic process completely, we need to specify the time index and state space first. A little more formal definition is as follows:

"A stochastic process $\mathbf{X} = \{ X(t), t \in T \}$ is a collection of random variables. That is, for each t in the index set T, X(t) is a random variable. We often interpret t as time and call X(t) the state of the process at time t"

The index set T can be countable set and we have a ***discrete-time*** stochastic process, or non-countable continuous set and we have a ***continuous-time*** stochastic process.

After getting the needed parameters, we can use the stochastic process to model the call admission control and resource allocation problem, and then adopt M.D.P to compute the best policy that optimizes the long-term system revenue. Figure 3.1 shows the framework of our problem briefly.

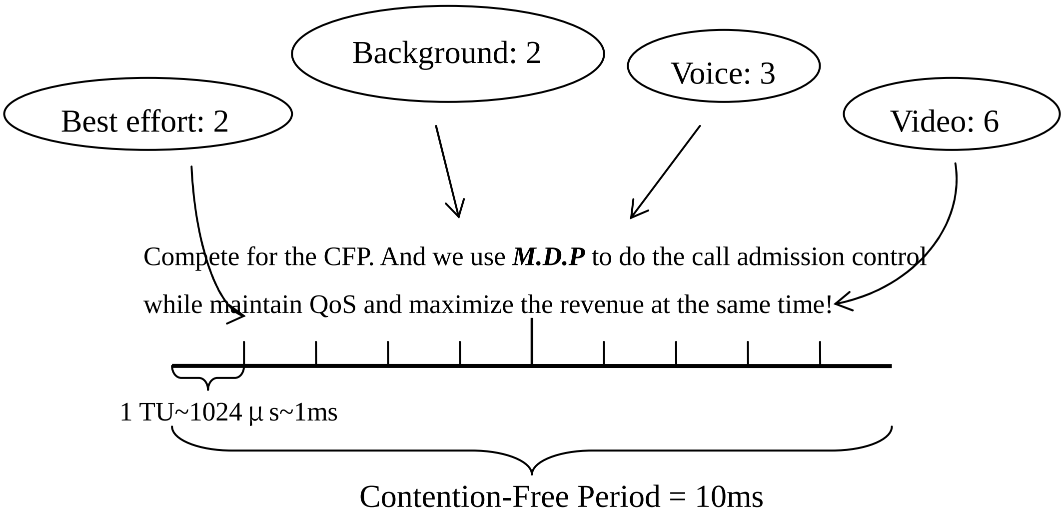


Figure 3.1 Brief illustration of thesis problem

3.3 Problem Formulation: Continuous-Time Case

Assumptions:

1. Four classes of traffic type: background, best effort, voice, and video. And each with different slot reservation quantity.
2. Poisson arrivals to request the CFP reservation.
3. Exponential service time (i.e. slots holding time is exponentially distributed).
4. Each traffic is independent.
5. Total CFP: 10 slots.
6. For best effort and background traffic, 2 slots are dedicated. For voice and video traffic, 3 and 6 slots are dedicated respectively.

We assume {X(t), t>0} is a continuous time Markov Chain

X(t) = (a, b, c, d) which is the admit traffic combination in the system at time t

States:

(a, b, c, d):

a means the number of background traffic which needs 2 slots for one traffic

b means the number of best effort traffic which needs 2 slots for one traffic

c means the number of voice traffic which needs 3 slots for one traffic

d means the number of video traffic which needs 6 slots for one traffic

For example, (2, 1, 1, 0) means there are 2 background traffic, 1 best effort traffic, 1 voice traffic, and 0 video traffic. Since total CFP is 10 slots, 1 slot is free for use (2*2+1*2+1*3+0*6=9, 10-9=1).

Depend on the resource constraint 2a +2b +3c +6d <=10, we could compute the whole feasible states showed below.

1	(0,0,0,0)	24	(1,1,0,1)
2	(0,0,0,1)	25	(1,1,1,0)
3	(0,0,1,0)	26	(1,1,2,0)
4	(0,0,1,1)	27	(1,2,0,0)
5	(0,0,2,0)	28	(1,2,1,0)
6	(0,0,3,0)	29	(1,3,0,0)
7	(0,1,0,0)	30	(1,4,0,0)
8	(0,1,0,1)	31	(2,0,0,0)
9	(0,1,1,0)	32	(2,0,0,1)
10	(0,1,2,0)	33	(2,0,1,0)
11	(0,2,0,0)	34	(2,0,2,0)
12	(0,2,0,1)	35	(2,1,0,0)
13	(0,2,1,0)	36	(2,1,1,0)
14	(0,2,2,0)	37	(2,2,0,0)
15	(0,3,0,0)	38	(2,3,0,0)
16	(0,3,1,0)	39	(3,0,0,0)
17	(0,4,0,0)	40	(3,0,1,0)
18	(0,5,0,0)	41	(3,1,0,0)
19	(1,0,0,0)	42	(3,2,0,0)
20	(1,0,0,1)	43	(4,0,0,0)
21	(1,0,1,0)	44	(4,1,0,0)
22	(1,0,2,0)	45	(5,0,0,0)
23	(1,1,0,0)		

Table 3.3　The sequence of the states

Table 3.4 is a verbal description of WLAN call admission control problem we considered.

Given:

- The arrival rate of each type of traffic
- The service rate of each type of traffic
- System revenue according to the different decision

To determine:

- The best slot allocation policy

Objective:

To maximize the system revenue

Subject to:

- Slot capacity constraint

Table 3.4 WLAN call admission control problem description (continuous-time case)

Notation:

Given Parameters	
Notation	**Description**
λ_i	The arrival rate of class i traffic, i = a, b, c, and d for background, best effort, voice, and video respectively
μ_i	The service rate of class i traffic, i = a, b, c, and d for background, best effort, voice, and video respectively
R_i	System revenue of serving one class i traffic, i = a, b, c, and d for background, best effort, voice, and video respectively
a_{ij}^{k}	The appropriate transition rate from state i to state j given decision k

Table 3.5 Notations of given parameters

Decision Matrix:

With each state we have 16 decisions in our model. Table 3.6 is the description of decisions and actions.

Decision *k*	**Action**
1	Admit all classes of traffic
2	Only not admit background traffic
3	Only not admit best effort traffic
4	Only not admit voice traffic
5	Only not admit video traffic
6	Admit voice and video traffic
7	Admit background and video traffic
8	Admit background and best effort traffic
9	Admit best effort and voice traffic
10	Admit background and voice traffic
11	Admit best effort and video traffic
12	Only admit background traffic
13	Only admit best effort traffic
14	Only admit voice traffic
15	Only admit video traffic
16	Do not admit any traffic

Table 3.6 Decision and action matrix

Figure 3.2 is the 4-dimensional Markov chain model of our problem, and Figure 3.3 is the details of a parallelogram in Figure 3.2.

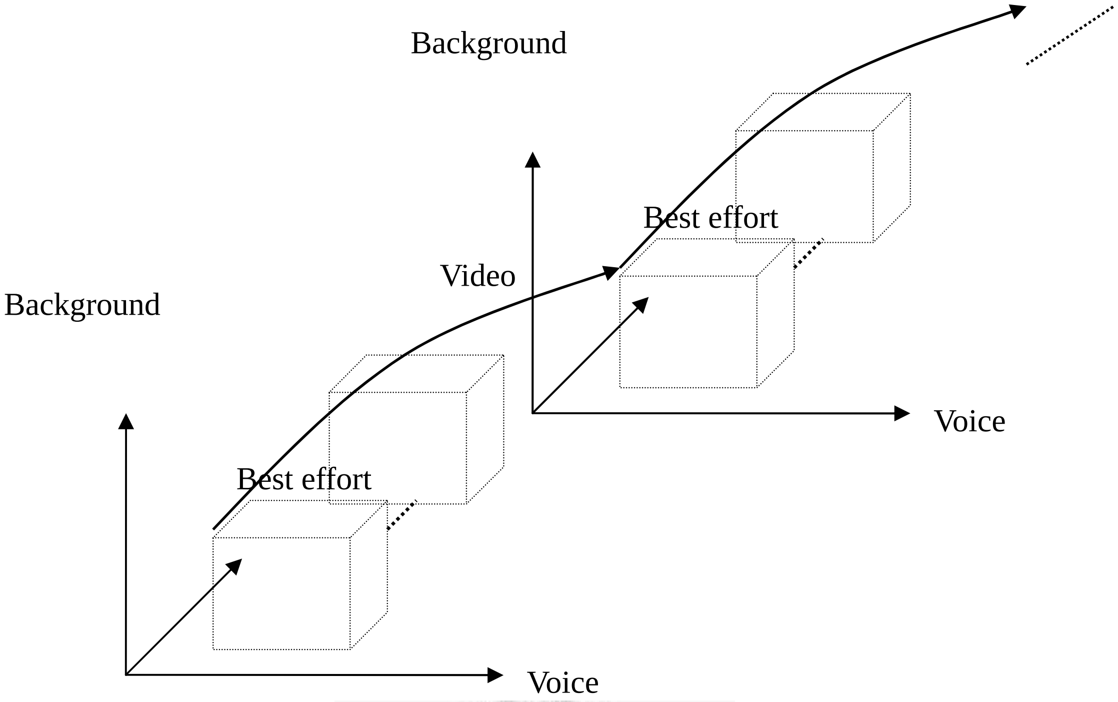


Figure 3.2 4-dimensional Markov chain model (continuous-time case)

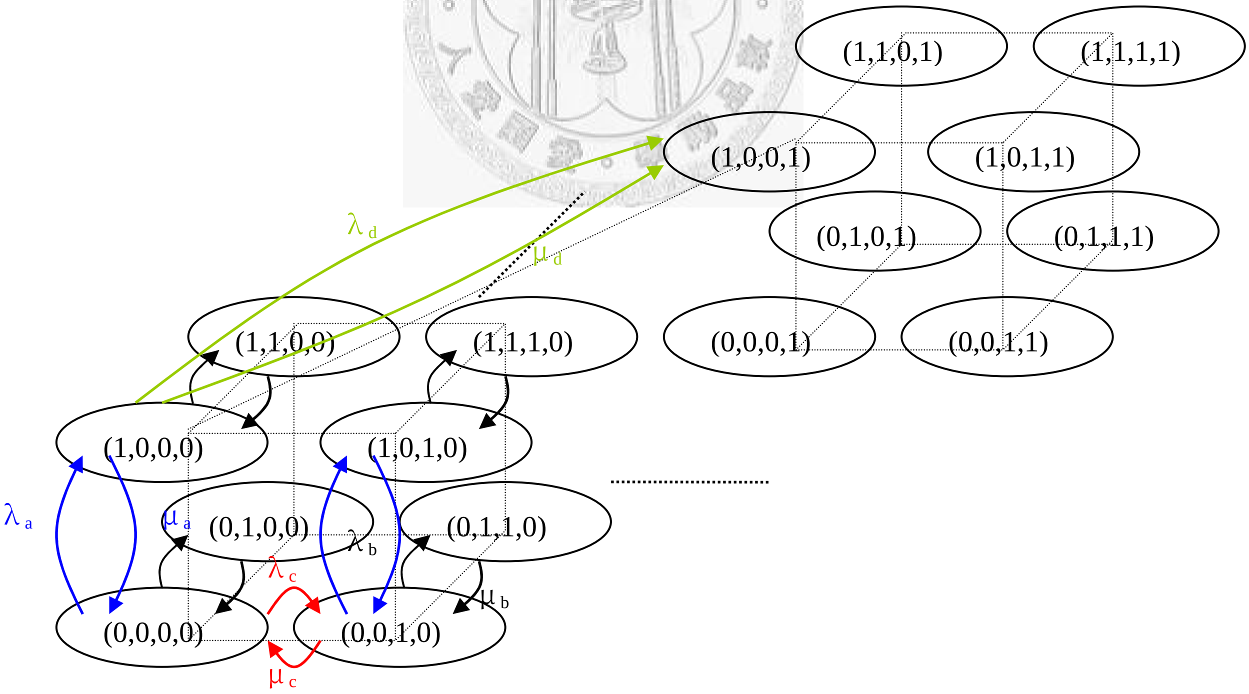


Figure 3.3 Details of a parallelogram in Figure 3.1

Because traffic arrivals follow a Poisson process, in accordance with the rare event property which Prob(1 event occurs in small interval h) = $\lambda h + o(h)$ and Prob(2 events occur in small interval h) = o(h), there is always one event occurred at one time.

3.4 Problem Formulation: Discrete-Time Case

The main difference between discrete-time case and continuous-time case is the mathematical model. In continuous-time case, we consider the arrival rate and service rate. We assume they are exponential distributed. Thus at one time only one event will occur. That is, at one point there will only be one traffic arrival or departure. But in discrete-time case this constraint does not exist. In other words, there may be multiple events happened in each time interval. There will be one, two or more traffic arrival and departure. Thus our model becomes more complicated. It’s no longer a birth-death process, and every two states have transitions. And we now consider transition *probability* matrix instead of transition *rate* matrix.

Assumptions:

1. Four classes of traffic type: background, best effort, voice, and video. And each with different slot reservation quantity.
2. Poisson arrivals to request the CFP reservation.
3. Exponential service time (i.e. slots holding time is exponentially distributed).
4. Slot holding time follows a **Geometric distribution**. We give the probability p for traffic to calculate the leaving probability.
5. Each traffic is independent.
6. Total CFP: 10 slots.
7. For best effort and background traffic, 2 slots are dedicated. For voice and video traffic, 3 and 6 slots are dedicated respectively.

We assume $\{X_n, n = 0, 1, 2\ldots\}$ is a discrete time Markov Chain

$X_n = (a, b, c, d)$ which is the admit traffic combination in the system

States:

(a, b, c, d):

a means the number of background traffic which needs 2 slots for one traffic

b means the number of best effort traffic which needs 2 slots for one traffic

c means the number of voice traffic which needs 3 slots for one traffic

d means the number of video traffic which needs 6 slots for one traffic

Because the resource constraint $2a + 2b + 3c + 6d \le 10$ is not changed in discrete-time case, the whole feasible states are the same as Table 3.3. Table 3.7 is a verbal description of WLAN call admission control problem we considered.

Given:

- The arrival rate of each type of traffic
- The service **probability** of each type of traffic
- System revenue according to the different decision

To determine:

- The best slot allocation policy

Objective:

To maximize the system revenue

Subject to:

- Slot capacity constraint

Table 3.7 WLAN call admission control problem description (discrete-time case)

Notation:

Given Parameters	
Notation	**Description**
λ_i	The arrival rate of class i traffic, i = a, b, c, and d for background, best effort, voice, and video respectively
μ_i	The service **probability** of class i traffic, i = a, b, c, and d for background, best effort, voice, and video respectively
R_i	System revenue of serving one class i traffic, i = a, b, c, and d for background, best effort, voice, and video respectively
P_{ij}^k	The appropriate transition **probability** from state i to state j given decision k

Table 3.8 Notations of given parameters

With each state we also have 16 decisions in our model. Table 3.6 is the description of decisions and actions.

Figure 3.4 is the 4-dimensional Markov chain model of our problem, and Figure 3.5 is the details of a quadrangle in Figure 3.4.

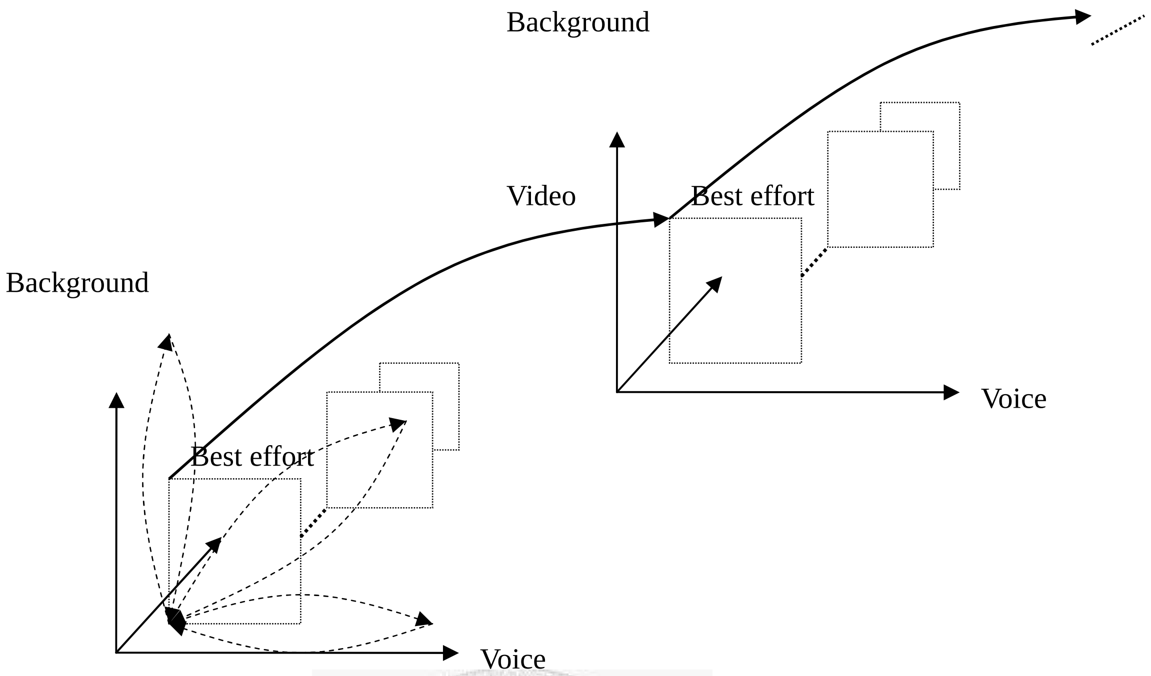


Figure 3.4 4-dimensional Markov chain model (discrete-time case)

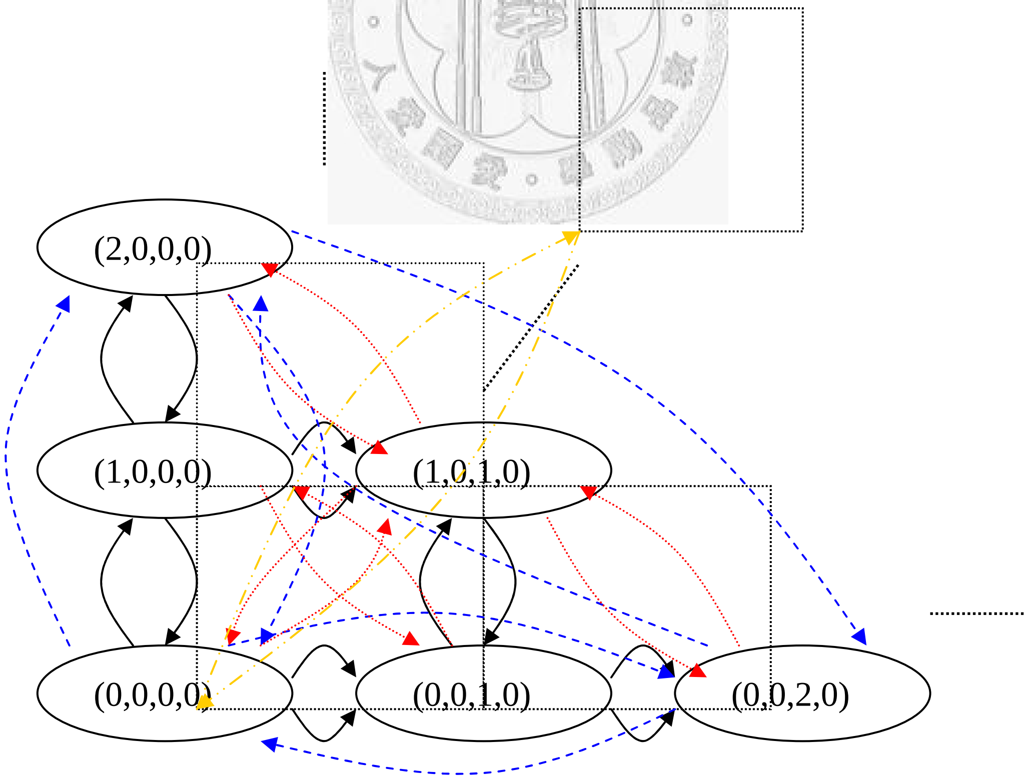


Figure 3.5 Details of quadrangle in Figure 3.4

Chapter 4. Solution Approach

4.1 Markovian Decision Process

Markovian Decision Process [7] [8] [18] is an application of dynamic programming. It is used to solve a stochastic decision process that can be described by a finite number of states. The transition probabilities between the states are described by a Markov chain. The reward structure of the process can also be described by a matrix whose individual elements represent the revenue (or cost) resulting from moving from one state to another. Both the transition and revenue matrices depend on the decision alternatives available to the decision maker. The objective of the problem is to determine the optimal policy that maximizes the expected revenue of the process over a finite or infinite number of states [18].

We have two models in our problem. One is discrete-time [16], and the other is continuous- time. We will discuss two different approaches as follows.

4.1.1 Policy Iteration Method: Discrete Time Case

State Transition ***Probability*** Matrix:

State	0	1	..	N
0	P_{00}	P_{01}	..	P_{0N}
1	P_{10}	P_{11}	..	P_{1N}
:	:	:		
N	P_{N0}	P_{N1}	..	P_{NN}

Revenue Matrix:

State	0	1	..	N
0	R_{00}	R_{01}	..	R_{0N}
1	R_{10}	R_{11}	..	R_{1N}
:	:	:		
N	R_{N0}	R_{N1}	..	R_{NN}

If there are N states in our system, the transition matrices above are identified for the transition probability (P) and the transition revenue (R) matrix. Suppose that the process is allowed to make transitions for a long time and we are interested in the earnings of the process. The limiting state probabilities π_i are independent of the starting state, and the gain g of system is: $g = \sum_{i=1}^{N} \pi_i q_i$, where q_i is the expected immediate return in state i defined by $\sum_{i=j}^{N} p_{ij} r_{ij}$ from the transition matrix [18].

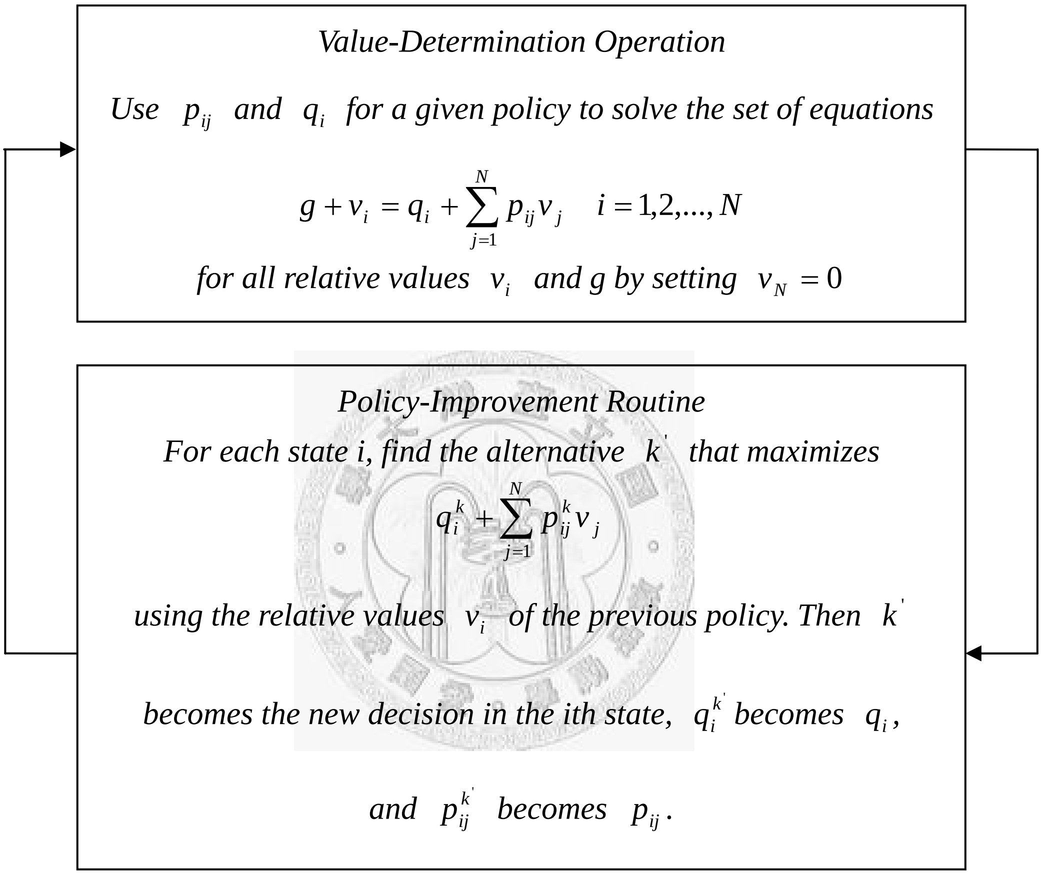
For each state, we will have K decisions to choose. What we want to find is the decision of each state that will maximize the total system revenue. The decision of all states can be described by a $1 \times N$ matrix called the policy, $D = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$.

If this process were to be allowed to operate for n transitions, we could define $v_i\,(n)$ as the total expected reward that the system will earn in n moves if it starts from state i under the given policy, and the $v_i\,(n)$ can be formulated from the result of Operational Research (OR), $v_i(n) = q_i + \sum_{j=1}^{N} p_{ij} v_j (n-1) \quad i = 1,2,\ldots,N \quad n = 1,2,3,\ldots$ .If the problem has large n, $v_i\,(n) = ng + v_i$, where v_i is the asymptotic intercepts of $v_i\,(n)$ [18].

From the exist proof of OR [18], the equation will be $g + v_i = q_i + \sum_{j=1}^{N} p_{ij} v_j$ for each state i. So there will leave N+1 unknown, N v_i and one g, with N equalities. We can solve the problem by Policy-Improvement Routine (PIR) described below.

PIR is defined as follows: For each state i, find the alternative k that maximizes the test quantity $q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j$, using the relative values determined under the old policy. This alternative k now becomes d_i, the decision in the *i*th state. A new policy matrix D will be determined when this procedure has been performed for every state. And the optimal policy has been reached (g is maximized) when the policies on two successive iterations are identical.

The Iteration Cycle:



Value-Determination Operation

Use p_{ij} *and* q_i *for a given policy to solve the set of equations*

$$g + v_i = q_i + \sum_{j=1}^{N} p_{ij} v_j \quad i = 1,2,\ldots,N$$

for all relative values v_i *and g by setting* $v_N = 0$

Policy-Improvement Routine

For each state i, find the alternative $k^{'}$ *that maximizes*

$$q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j$$

using the relative values v_i *of the previous policy. Then* $k^{'}$ *becomes the new decision in the ith state,* $q_i^{k^{'}}$ *becomes* q_i*, and* $p_{ij}^{k^{'}}$ *becomes* p_{ij}*.*

Table 4.1 Policy iteration cycle of the discrete time model [8]

4.1.2 Policy Iteration Method: Continuous Time Case

State Transition ***Rate*** Matrix:

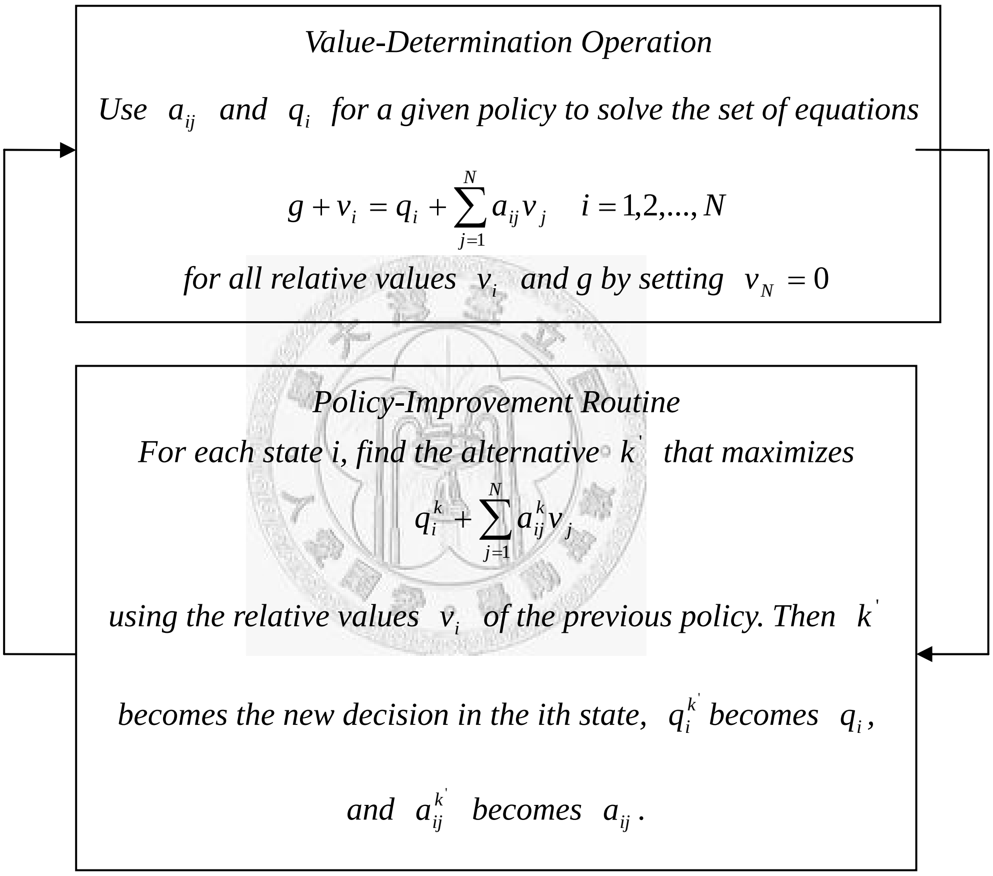
State	0	1	..	N
0	$-v_0$	a_{01}	..	a_{0N}
1	a_{10}	$-v_1$	..	a_{1N}
:	:	:		
N	a_{N0}	a_{N1}	..	$-v_N$

Revenue Matrix:

State	0	1	..	N
0	R_{00}	R_{01}	..	R_{0N}
1	R_{10}	R_{11}	..	R_{1N}
:	:	:		
N	R_{N0}	R_{N1}	..	R_{NN}

When the transitions between states are at random time intervals, the M.D.P must view it as a continuous time case problem. Reflection shows that the significant parameters of the process must be **transition rates** rather than transition probabilities. And we will use transition rate matrix A instead of transition probability matrix P. The transition rate has a new property of $a_{jj} = -\sum_{i \neq j} a_{ji}$ .The gain g of system is still: $g = \sum_{i=1}^{N} \pi_i q_i$, the q_i is the earning rate in state i of the system, and redefined by $r_{ii} + \sum_{j=1, i \neq j}^{N} a_{ij} r_{ij}$ from the transition rate matrix A and revenue matrix R [18].

The Iteration Cycle:



Value-Determination Operation

Use a_{ij} *and* q_i *for a given policy to solve the set of equations*

$$g + v_i = q_i + \sum_{j=1}^{N} a_{ij} v_j \quad i = 1,2,\ldots,N$$

for all relative values v_i *and g by setting* $v_N = 0$

Policy-Improvement Routine

For each state i, find the alternative $k^{'}$ *that maximizes*

$$q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j$$

using the relative values v_i *of the previous policy. Then* $k^{'}$ *becomes the new decision in the ith state,* $q_i^{k^{'}}$ *becomes* q_i, *and* $a_{ij}^{k^{'}}$ *becomes* a_{ij}.

Table 4.2 Policy iteration cycle of the continuous time model [8]

From the proof of the OR, the continuous-time case PIR is defined as follows: For each state i, find the alternative k that maximizes the test quantity $q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j$, using the relative values determined under the old policy. This alternative k now becomes d_i, the decision in the *i*th state. A new policy matrix D will be determined when this procedure has been performed for every state [18]. The same as discrete-time case, the optimal policy has been reached (g is maximized) when the policies on two successive iterations are identical.

In summary, the policy iteration method has the following properties:

1. The solution of the sequential decision process is reduced to solving sets of linear simultaneous equations and subsequent comparisons.
2. Each succeeding policy found in the iteration cycle has a higher gain than the previous one.
3. The iteration cycle will terminate on the policy that has largest gain stainable within the realm of the problem; it will usually find this policy in a small number of iterations.

4.2 Linear Algebra [10]

It's difficult to solve sets of linear simultaneous equations. As a result, we convert the linear system to a matrix form and then we can compute the solution of the corresponding sets of linear simultaneous equations in a systematic way.

4.2.1 Reduced row echelon form

Definition:

An m × n matrix is said to be in *reduced row echelon form* [10] when it satisfies the following properties:

(a) All rows consisting entirely of zeros, if any, are ate the bottom of the matrix.

(b) Reading from left to right, the first nonzero entry in each row that does not consist entirely of zeros is a 1, called the leading entry of its row.

(c) If rows i and i ＋ 1 are two successive rows that do not consist entirely of zeros, then the leading entry of row i ＋ 1 is to the right of the leading entry of row i.

(d) If a column contains a leading entry of some row, then all other entries in that column are zero.

(Note that a matrix in reduced row echelon form might not have any rows that consist entirely of zeros.)

Procedure:

Step 1.

Find the first (counting from left to right) column in matrix not all of whose entries are zero. This column is called the ***pivotal column***.

$$\begin{bmatrix} 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 2 & 2 & -5 & 2 & 4 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

pivotal column

Step 2.

Identify the first (counting from top to bottom) nonzero entry in the pivotal column. This element is called the ***pivot***.

$$\begin{bmatrix} 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 2 & 2 & -5 & 2 & 4 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

pivot

Step 3.

Interchange, if necessary, the first row with the row where the pivot occurs so that the pivot is now in the first row.

$$\begin{bmatrix} 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 2 & 2 & -5 & 2 & 4 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} 2 & 2 & -5 & 2 & 4 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

Step 4.

Multiply the first row by the reciprocal of the pivot. Thus the entry in the first row and pivotal column (where the pivot was located) is now a 1.

$$\begin{array}{l} \rightarrow \\ \\ \\ \\ \end{array}\begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

Step 5.

Add multiples of the first row to all other rows to make all entries in the pivotal column, except the entry where the pivot was located, equal to zero.

$$\begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & -2 & -1 & 7 & 3 \end{bmatrix}$$

Step 6.

Identify the (m − 1) × n submatrix obtained by neglecting the first row. Repeat steps 1 through 5 on the submatrix.

$$\begin{matrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \end{matrix}$$

$$\begin{bmatrix} 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & -2 & -1 & 7 & 3 \end{bmatrix}$$

Step 7.

Repeat steps 1 through 6 on the whole matrix until all rows are shaded.

$$\begin{matrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Step 8.

Add multiples of each row of the matrix having a leading 1 to zero out all entries above the leading 1. And we could get the reduced row echelon form of the matrix.

$$\begin{bmatrix} 1 & 0 & 0 & 9 & \frac{19}{2} \\ 0 & 1 & 0 & -\frac{17}{4} & -\frac{5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

4.2.2 Gauss-Jordan reduction method

The **Gauss-Jordan reduction** procedure [10] for solving the linear system **Ax = b** is as follows.

Step 1. Form the augmented matrix $\left[A \quad | \quad b\right]$.

Step 2. Transform the augmented matrix to reduced row echelon form by using elementary row operations.

Step 3. The linear system that corresponds to the matrix in reduced row echelon form that has been obtained in step 2 has exactly the same solutions as the given linear system. For each nonzero row of the matrix in reduced row echelon form, solve the corresponding equation for the unknown that corresponds to the leading entry of the row. The rows consisting entirely of zeros can be ignored, since the corresponding equation will be satisfied for any values of the unknowns.

Chapter 5. Computational Results

We will use several examples to demonstrate our problem in different situations and show the computational results of our model. Both non-preemptive and preemptive cases are considered.

Non-Preemptive

Example 1:

Given parameters:

Traffic type **Parameters**	Background (BK)	Best effort (BE)	Voice (VO)	Video (VI)
Arrival rate (λ_i)	$\lambda_a = 1$	$\lambda_b = 1$	$\lambda_c = 3$	$\lambda_d = 9$
Service rate (μ_i)	$\mu_a = 2$	$\mu_b = 2$	$\mu_c = 4$	$\mu_d = 10$
Revenue (R_i)	$R_a = 1$	$R_b = 2$	$R_c = 3$	$R_d = 10$

Table 5.1 Given parameters of example 1

Used policy:

No slot preservation
Non-preemptive

Table 5.2 The original policy of example 1

Results:

4 iterations to achieve the optimal policy
9 decisions are changed
Original profit: 51.5224
Maximum profit: 52.5113
Added revenue: 0.9889

Table 5.3 The results of example 1

The changed states:

State	Original Decision	New Decision
(0, 0, 0, 0)	All	Voice & Video
(0, 0, 0, 1)	Only no Video	Voice
(0, 0, 1, 0)	All	Video
(0, 1, 0, 0)	All	Video
(0, 2, 0, 0)	All	Video
(1, 0, 0, 0)	All	Best effort & Video
(1, 0, 0, 1)	Background & Best effort	Best effort
(1, 1, 0, 0)	All	Video
(2, 0, 0, 0)	All	Video

Table 5.4 The changed states of example 1

Result discussion:

In this example, we assume video traffic has a higher arrival rate, service rate, and revenue. Thus, the system will admit video traffic more often than other types of traffic to get the optimal long-term revenue. In some states the system will reject background and best effort traffic requests because they will stay in system longer and generate less revenue than the other two traffic. The computational result is consistent with our expectation. In some states, even there are still some slots for all traffic but the system only admit video traffic due to its long-term high revenue.

In this example only 7 decisions are changed, which means the original policy is good for this situation. As a result, the difference between maximum profit and original profit is not much. We use the M.D.P to compute the best slots allocation policy that maximizes the system revenue successfully.

We can compare our result with other static slot allocation policies to show that our policy is the best one. At first, we divide the four type traffic into two subsets, which are low-priority and high-priority respectively. And policy (4, 6) means that the system reserves 4 slots for low-priority traffic and reserves 6 slots for high-priority traffic.

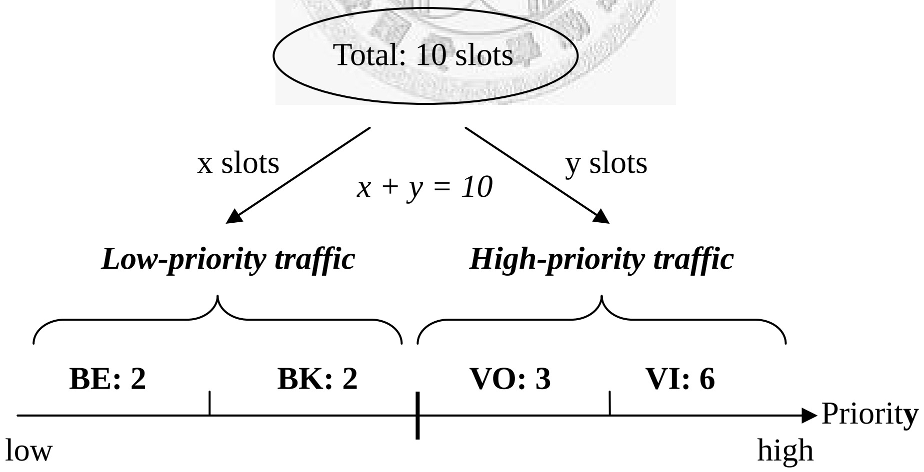


Figure 5.1 Priority subsets for slot reservation policy

Table 5.5 lists some other policies and their corresponding revenue. We can find that our policy is the best one that maximizes long-term system revenue.

Policy	**Revenue**
(10, 0)	32.1458
(8, 2)	30.8569
(6, 4)	33.2698
(4, 6)	35.6982
(2, 8)	49.3687
(0, 10)	50.3697

Table 5.5 Other policies and their corresponding revenue of example 1

Example 2:

Given parameters:

Traffic type **Parameters**	Background (BK)	Best effort (BE)	Voice (VO)	Video (VI)
Arrival rate (λ_i)	$\lambda_a = 10$	$\lambda_b = 1$	$\lambda_c = 3$	$\lambda_d = 2$
Service rate (μ_i)	$\mu_a = 11$	$\mu_b = 2$	$\mu_c = 4$	$\mu_d = 3$
Revenue (R_i)	$R_a = 10$	$R_b = 2$	$R_c = 3$	$R_d = 2$

Table 5.6 Given parameters of example 2

Used policy:

No slot preservation
Non-preemptive

Table 5.7 The original policy of example 2

Results:

5 iterations to achieve the optimal policy
27 decisions are changed
Original profit: 78.0558
Maximum profit: 88.7502
Added revenue: 10.6944

Table 5.8 The results of example 2

The changed states:

State	**Original Decision**	**New Decision**	**State**	**Original Decision**	**New Decision**
0	1	5	**20**	5	10
1	5	10	**21**	8	12
2	1	12	**22**	1	10
4	5	12	**24**	5	12
6	1	10	**26**	5	12
7	8	12	**28**	8	12
8	5	10	**30**	1	5
9	8	12	**32**	5	12
10	1	12	**34**	5	12
12	5	12	**36**	8	12
14	5	12	**38**	5	10
16	8	12	**40**	8	12
18	1	5	**42**	8	12
19	8	12			

Table 5.9 The changed states of example 2

Result discussion:

In this example, we give background traffic a higher arrival rate, service rate and revenue. Thus the system should admit background traffic as possible to get the optimum long-term revenue. Our results show that in many states the system will only admit background traffic although it still has free slots available to other traffic. There are 27 decisions changed, which means the original policy is not good for this example. And our maximum profit is much better than original profit. Table 5.10 lists the other policies and their corresponding revenue.

Policy	**Revenue**
(10, 0)	86.6974
(8, 2)	76.2689
(6, 4)	77.6935
(4, 6)	76.3642
(2, 8)	72.3684
(0, 10)	70.9836

Table 5.10 Other policies and their corresponding revenue of example 2

Preemptive

In this section, we use the preemptive policy (i.e. low-priority traffic must stop transmission immediately to allow high-priority traffic to use the slots if all slots are occupied) to demonstrate the same data as previous examples to discuss the difference between these two policies.

Example 3:

Given parameters:

Traffic type **Parameters**	Background (BK)	Best effort (BE)	Voice (VO)	Video (VI)
Arrival rate (λ_i)	$\lambda_a = 1$	$\lambda_b = 1$	$\lambda_c = 3$	$\lambda_d = 9$
Service rate (μ_i)	$\mu_a = 2$	$\mu_b = 2$	$\mu_c = 4$	$\mu_d = 10$
Revenue (R_i)	$R_a = 1$	$R_b = 2$	$R_c = 3$	$R_d = 10$

Table 5.11 Given parameters of example 3

Used policy:

No slot preservation
Preemptive (i.e. low-priority traffic must stop transmission immediately to allow high-priority traffic to use the slots if all slots are occupied)

Table 5.12 The original policy of example 3

Results:

5 iterations to achieve the optimal policy
40 decisions are changed
Original profit: 35.5224
Maximum profit: 53.6542
Added revenue: 18.1318

Table 5.13 The results of example 3

The changed states:

State	Original Decision	New Decision	State	Original Decision	New Decision
1	5	1	**25**	16	11
3	16	1	**26**	5	2
4	5	1	**27**	16	2
5	16	1	**28**	8	2
7	8	1	**29**	16	2
8	5	1	**30**	1	2
9	8	4	**31**	16	1
11	16	1	**32**	5	2
12	5	1	**33**	16	4
13	16	4	**34**	5	2
14	5	1	**35**	16	2
15	16	1	**36**	8	2
16	8	1	**37**	16	2
17	16	1	**38**	5	2
19	8	1	**39**	16	1
20	5	1	**40**	8	2
21	8	4	**41**	16	2
22	1	2	**42**	8	2
23	16	2	**43**	16	2
24	5	2	**44**	16	1

Table 5.14 The changed states of example 3

Result discussion:

In this example, we give video traffic a higher arrival rate, service rate, and revenue. In addition, under preemptive policy, video traffic has the highest priority to use the slots, even though all slots are allocated. Comparing video traffic and voice traffic, as a result of higher revenue, arrival rate and service rate of video traffic, the system may reject voice traffic more often than video traffic in heavy-loaded system.

Furthermore, although best effort traffic has a lower priority than background traffic, but the former has slightly larger revenue. Therefore, the system will admit best effort traffic more often than background traffic in order to get the optimal long-term revenue.

In contrast to the non-preemptive policy, the system will get more profit due to its flexibility in resource allocation. The computational result is consistent with our expectation. Table 5.15 lists the other policies and their corresponding revenue.

Policy	**Revenue**
(10, 0)	35.2365
(8, 2)	33.6829
(6, 4)	34.6853
(4, 6)	45.3697
(2, 8)	40.3953
(0, 10)	46.8587

Table 5.15 Other policies and their corresponding revenue of example 3

Example 4:

Given parameters:

Traffic type **Parameters**	Background (BK)	Best effort (BE)	Voice (VO)	Video (VI)
Arrival rate (λ_i)	$\lambda_a = 10$	$\lambda_b = 1$	$\lambda_c = 3$	$\lambda_d = 2$
Service rate (μ_i)	$\mu_a = 11$	$\mu_b = 2$	$\mu_c = 4$	$\mu_d = 3$
Revenue (R_i)	$R_a = 10$	$R_b = 2$	$R_c = 3$	$R_d = 2$

Table 5.16 Given parameters of example 4

Used policy:

No slot preservation
Preemptive (i.e. low-priority traffic must stop transmission immediately to allow high-priority traffic to use the slots if all slots are occupied)

Table 5.17 The original policy of example 4

Results:

5 iterations to achieve the optimal policy
38 decisions are changed
Original profit: 78.0558
Maximum profit: 89.5339
Added revenue: 11.4781

Table 5.18 The results of example 4

The changed states:

State	**Original Decision**	**New Decision**	**State**	**Original Decision**	**New Decision**
0	1	5	**22**	1	5
1	5	4	**23**	16	4
2	1	8	**24**	5	8
3	16	1	**25**	16	8
4	5	8	**27**	16	8
5	16	1	**28**	8	5
6	1	5	**29**	16	10
7	8	4	**30**	1	5
8	5	8	**31**	16	4
10	1	5	**32**	5	8
11	16	4	**33**	16	8
12	5	8	**34**	5	8
13	16	8	**35**	16	8
15	16	8	**37**	16	12
16	8	5	**38**	5	8
17	16	5	**39**	16	8
18	1	5	**41**	16	8
19	8	4	**43**	16	12
20	5	8	**44**	16	8

Table 5.19 The changed states of example 4

Result discussion:

In this example, we give background traffic the highest arrival rate, service rate, and revenue. The system should admit background traffic in order to get the optimal long-term revenue. Moreover, voice and video traffic use more slots than background and best effort traffic while leading to slightly difference in their revenue. However, under preemptive policy, low-priority traffic must stop transmission to allow high-priority traffic to use the slots if all slots are occupied. As a result, the system may only admit best effort and background traffic so the optimal long-term revenue may be achieved.

The same as example 3, the system will get more profit compared with non-preemptive policy. The computational result is also consistent with our expectation.

Policy	**Revenue**
(10, 0)	88.3654
(8, 2)	85.3614
(6, 4)	85.3975
(4, 6)	80.3619
(2, 8)	78.3691
(0, 10)	74.3697

Table 5.20 Other policies and their corresponding revenue of example 4

Example 5:

Given parameters:

Traffic type **Parameters**	Background (BK)	Best effort (BE)	Voice (VO)	Video (VI)
Arrival rate (λ_i)	$\lambda_a = 1$	$\lambda_b = 1$	$\lambda_c = 3$	$\lambda_d = 2$
Service rate (μ_i)	$\mu_a = 2$	$\mu_b = 2$	$\mu_c = 4$	$\mu_d = 3$
Revenue (R_i)	$R_a = 1$	$R_b = 10$	$R_c = 3$	$R_d = 3$

Table 5.21 Given parameters of example 5

Used policy:

No slot preservation
Preemptive (i.e. low-priority traffic must stop transmission immediately to allow high-priority traffic to use the slots if all slots are occupied)

Table 5.22 The original policy of example 5

Results:

4 iterations to achieve the optimal policy
40 decisions are changed
Original profit: 16.0914
Maximum profit: 16.9557
Added revenue: 0.8643

Table 5.23 The results of example 5

The changed states:

State	**Original Decision**	**New Decision**	**State**	**Original Decision**	**New Decision**
1	5	1	**25**	16	13
4	5	8	**26**	5	9
5	16	1	**27**	16	13
7	8	11	**28**	8	13
9	8	3	**29**	16	13
10	1	5	**30**	1	2
11	16	11	**31**	16	1
12	5	9	**32**	5	2
13	16	13	**33**	16	4
14	5	9	**34**	5	9

15	16	13	**35**	16	13
16	8	13	**36**	8	13
17	16	13	**37**	16	13
18	1	2	**38**	5	2
19	8	9	**39**	16	1
20	5	9	**40**	8	13
21	8	13	**41**	16	13
22	1	9	**42**	8	2
23	16	11	**43**	16	13
24	5	9	**44**	16	1

Table 5.24 The changed states of example 5

Result discussion:

In this example, we give the higher revenue to the best effort traffic, while it has the same arrival and service rate as background traffic, but lower than voice and video traffic. Under this condition, the system will reject background traffic rather than other traffic in spite of the sufficient slots to allocate for background traffic. This is due to the higher long-term revenue of other traffic. The computational result is consistent with our estimate.

In addition, we give voice traffic a higher arrival rate and service rate than video traffic, even when they have the same revenue. As a result, some states will admit both best effort and voice traffic instead of background and video traffic to maximize the long-term profit.

Policy	**Revenue**
(10, 0)	15.3987
(8, 2)	13.3628
(6, 4)	14.6823
(4, 6)	13.3978
(2, 8)	8.3975
(0, 10)	7.6839

Table 5.25 Other policies and their corresponding revenue of example 5

Example 6:

Given parameters:

Traffic type **Parameters**	Background (BK)	Best effort (BE)	Voice (VO)	Video (VI)
Arrival rate (λ_i)	$\lambda_a = 1$	$\lambda_b = 9$	$\lambda_c = 3$	$\lambda_d = 8$
Service rate (μ_i)	$\mu_a = 2$	$\mu_b = 10$	$\mu_c = 4$	$\mu_d = 4$
Revenue (R_i)	$R_a = 1$	$R_b = 5$	$R_c = 2$	$R_d = 15$

Table 5.26 Given parameters of example 6

Used policy:

No slot preservation
Preemptive (i.e. low-priority traffic must stop transmission immediately to allow high-priority traffic to use the slots if all slots are occupied)

Table 5.27 The original policy of example 6

Results:

5 iterations to achieve the optimal policy
45 decisions are changed
Original profit: 56.2881
Maximum profit: 72.4224
Added revenue: 16.1343

Table 5.28 The results of example 6

The changed states:

State	**Original Decision**	**New Decision**	**State**	**Original Decision**	**New Decision**
0	1	11	**23**	16	11
1	5	13	**24**	5	11
2	1	11	**25**	16	11
3	16	1	**26**	5	11
4	5	11	**27**	16	11
5	16	1	**28**	8	11
6	1	11	**29**	16	11
7	8	11	**30**	1	2
8	5	11	**31**	16	1
9	8	11	**32**	5	11
10	1	11	**33**	16	4
11	16	11	**34**	5	2
12	5	13	**35**	16	11
13	16	15	**36**	8	2
14	5	11	**37**	16	2
15	16	13	**38**	5	2
16	8	11	**39**	16	1

17	16	11	**40**	8	11
18	1	11	**41**	16	11
19	8	13	**42**	8	2
20	5	11	**43**	16	2
21	8	11	**44**	16	1
22	1	11			

Table 5.29 The changed states of example 6

Result discussion:

In this example, we give best effort traffic the highest arrival rate and service rate, but it only has a value of 5 for its corresponding revenue. It is a little larger than the value for background and voice traffic, but far below those from the value for video traffic. However, the arrival rate and service rate of the video traffic are smaller than the values of best effort traffic. Thus, theoretically the system will admit best effort and video traffic with nearly same probability as possible to get the optimal long-term revenue. The computational result is consistent with our expectation. Most states only admit best effort and video traffic even there are still enough slots to be allocated to other types of traffic. And some states only reject background traffic as a result of the lowest arrival rate, service rate and revenue. By this slots allocation mechanism, the system revenue could be optimized. Table 5.30 lists the other policies and their corresponding revenue.

Policy	**Revenue**
(10, 0)	58.6374
(8, 2)	55.3974
(6, 4)	55.9637
(4, 6)	66.9875
(2, 8)	60.9869
(0, 10)	59.6387

Table 5.30 Other policies and their corresponding revenue of example 6

In brief, we use the M.D.P to compute the best slot allocation policy that maximizes the long-term system revenue successfully.

Chapter 6. Summary and Future Research

6.1 Summary

In this thesis, we bring up an idea to slot the 802.11 WLAN contention-free period. Via this idea, the WLAN QoS demands can be considered as a slot allocation problem (or call admission control problem). In addition, the overall system could eliminate the unnecessary interference with slotting synchronization.

We introduce the background of the IEEE 802.11 architecture and several researches focused on providing QoS for real-time traffic and multimedia …, etc. It also includes an emerging standard, IEEE 802.11e. Furthermore, this thesis uses stochastic process to model the WLAN slot allocation problem (or call admission control problem). We model the problem in two ways: continuous-time and discrete-time. Our goal is to find the best policy for time slot allocation to maximize the system long-term revenue. Markovian Decision Process (M.D.P) is applied to solve our problem. There are two methods when applying Markovian Decision Process, and as a result of our large problem size, we adopt the policy iteration method to solve the problem.

Finally, we list six examples for testing our model. It’s easy to find that it’s not a good policy to admit any traffic as long as there are still enough slots in the system. From the long-term point of view, rejecting the current low-profit traffic may lead to more profitable traffic in the future. As a result, from this thesis, we provide a systematic slot allocation (or call admission) mechanism for WLAN vendors to fulfill the QoS demands of users and maximize the long-term system revenue at the same time.

6.2 Future Research

There are still many researches that can be done on the QoS issues for WLAN network, especially on the measurement and control of the delay, throughput, and the end-to-end QoS. Furthermore, in order to be compatible with EDCA contention-based mechanism, this thesis follows the 802.11e specification and only considers four types of traffic. However, there may be other kinds of traffic to be considered in the real world. And the total contention-free period should be decided dynamically according to the system status from the HC. Therefore, our model can be extended to more types of traffic or even more slots to be allocated. But with the growth of the problem size, Markovian Decision Process will not be suitable for solving the problem. Hence we should find the more efficient dynamic programming method to solve the problem.

Besides, we may provide other decisions to be chosen for each state to stretch our model. It will increase the flexibility of our model to be a better fit for applications under real conditions.

Finally, we can take the contention period into consideration with our call admission control (slot allocation) mechanism in contention-free period. How to combine the two parts gracefully to support the QoS demands over WLAN is an important but formidable task today.

Considering these extensions will make our model more complicated. But in order to improve the accuracy of the model and to promote its practicality in the real world, all of the above mentioned are irreplaceable in their value.

Reference

[1] IETF "Integrated Service" Working Group. http://www2.ietf.org/html.charters/intserv-charter.html

[2] IETF "Differentiated Service" Working Group. http://www2.ietf.org/html.charters/diffserv-charter.html

[3] The Institute of Electrical and Electronics Engineers, Inc, "***Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications,***" ANSI/IEEE Std 802.11, 1999 Edition, 20 Aug. 1999.

[4] The Institute of Electrical and Electronics Engineers, Inc, "***Draft Amendment to IEEE Std 802.11, 1999 Edition, Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service,***" IEEE P802.11e/D5.1, Oct. 2003.

[5] I. Aad, and C. Castelluccia, "***Differentiation Mechanisms for IEEE 802.11,***" IEEE INFOCOM 2001, vol. 1, pp.209-218, Apr. 2001.

[6] A. Grilo, and M. Nunes, "***Performance Evaluation of IEEE 802.11e,***" The 13th IEEE International Symposium on Indoor and Mobile Radio Communications, vol.1, pp.511-517, Sep. 2002.

[7] Hillier, and Lieberman, "***Introduction to Operations Research,***" fourth edition, Holden-Day, Inc, 1986.

[8] R.A. Howard, "***Dynamic Programming and Markov Process,***" sixth printing, Wiley, 1970.

[9] H.K. Hung, "***Operation of GPRS Time Slot Allocation,***" Jun. 2001.

[10] B. Kolman, "***Introductory Linear Algebra with Applications,***" sixth edition,

Prentice-Hall, Inc, 1997.

[11] A. Lindgren, A. Almquist, and O. Schelen, “***Quality of Service Schemes for IEEE 802.11 Wireless LANs – An Evaluation,***” 26th Annual IEEE Conference on Local Computer Networks, pp.348-351, Nov. 2001.

[12] I. Mahadevan, and K.M. Sivalingam, “***Quality of Service Architectures for Wireless Networks: IntServ and DiffServ Models,***” Workshop on Mobile Computing at the International Symposium on Parallel Architectures, Algorithms, and Networks, pp.420-425, Perth, Australia, Jun. 1999.

[13] S. Mangold, “***IEEE 802.11e: Coexistence of Overlapping Basic Service Sets,***” Proc. of the Mobile Venue’02, pp.131-135, Athens, Greece, May 2002.

[14] S. Mangold, S. Choi, P. May, and G. Hiertz, “***IEEE 802.11e – Fair Resource Sharing between Overlapping Basic Service Sets,***” The 13th IEEE International Symposium on Indoor and Mobile Radio Communications, vol. 1, pp.166-171, Sep. 2002.

[15] S. Mangold, S. Choi, G. R. Hiertz, O. Klein, and B. Walke, “***Analysis of IEEE 802.11e for QoS Support in Wireless LANs,***” IEEE Wireless Communications, vol. 10, pp.40-50, Dec. 2003.

[16] M.L. Puterman, “***Markov Decision Process: Discrete Stochastic Dynamic Programming,***” Wiley, New York, 1994.

[17] S.M. Ross, “***Introduction to Probability Models,***” eighth edition, Academic-Press, 2003.

[18] H.A. Taha, “***Operations Research: an Introduction,***” fifth edition, Macmillan, 1992.

[19] N.H. Vaidya, P. Bahl, and S. Gupta, “***Distributed Fair Scheduling in a Wireless LAN,***” Sixth Annual International Conference on Mobile Computing and Networking, Boston, 2000.

[20] A. Veres, A.T. Campbell, M. Barry, and L.H. Sun, “***Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control,***” IEEE Journal on Selected Areas in Communications, vol. 19, pp.2081-2093, Oct. 2001.

[21] J.Y. Yeh, and C.H. Chen, “***Support of Multimedia Service with the IEEE 802.11 MAC Protocol,***” Jul. 2001.

簡歷

姓名：施翔騰

出生地：台灣省台北縣

出生日：中華民國六十九年十月二日

學歷：八十七年九月至九十一年六月

國立台灣大學工商管理學系 工業管理組

九十一年九月至九十三年七月

國立台灣大學資訊管理學研究所