ABSTRACT

LEI, HONGYAN. Performance analysis of power management in WLAN and UMTS. (Under the direction of Professor Arne A. Nilsson).

Wireless networks have enjoyed the exponential development, and wireless communication has become an essential part of modern life. Many new wireless applications demand higher speed and consume more energy. However, wireless devices are always powered by batteries, which have limited life time and constrain the use of wireless devices and the growth of wireless networks. Energy efficiency becomes an important issue in wireless networks. We study the energy efficiency in the IEEE 802.11 based WLAN (Wireless Local Area Network) and the third generation cellular system UMTS (Universal Mobile Telecommunication System), in which the basic mechanism is to put a mobile device into a low power consumption state when it is idle and wake it up periodically to transmit/receive traffic.

In WLANs, the study is focused on the MAC (Media Access Control) sublayer. Two queueing models for the power management mechanisms in an infrastructure network are proposed: the M/G/1 queue with bulk service and the D/G/1 queue. Based on the analytical and simulation results, suggestions are given about how to optimally configure the power management parameters. We also propose the enhanced power management schemes for both infrastructure and independent networks, which outperform the original schemes based on our analysis and simulation.

In UMTS, the impacts of Discontinuous Reception (DRX) mechanism and inactivity timer are studied. The simulation of the performance of power saving mechanism is carried out by inputting several typical traffic models specified by 3GPP (third generation partnership project). From the results that different traffic models demand different optimal parameters, we propose to adaptively configure the DRX cycle and inactivity timer parameters based on real-time measurements.

PERFORMANCE ANALYSIS OF POWER MANAGEMENT IN WLAN AND UMTS

by

Hongyan Lei

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial satisfaction of the
requirements for the Degree of
Doctor of Philosophy

Department of Electrical and Computer Engineering

Raleigh, NC

2005

Approved By:

Dr. Mihail Devetsikiotis	Dr. Wenye Wang
Dr. Arne A. Nilsson Chair of Advisory Committee	Dr. Alexandra Duel-Hallen

To my parents, my husband, and my brother,
for their endless love and support.

Biography

I was born in Shaaxi Province, P.R. China. I received my Bachelor degree in Computer Science at Tianjin University in July 1994, and received my Master degree in Electrical Engineering at Beijing University of Posts and Telecommunications in April 1997.

In August 2001, I joined the Electrical and Computer Engineering Department at North Carolina State University. My Ph.D research focuses on the performance analysis of power management in wireless networks. I passed final defense in August, 2005.

Acknowledgements

First, I give my sincerest gratitude to my advisor, professor Arne A. Nilsson, for his patient guidance, invaluable encouragement, wealth of knowledge and deep insight in the research. He not only helped me go through the tough study process, also taught me the optimistic attitude toward people and life. I'm honored to have the priceless experience of working with him.

I want to thank my committee members, Dr. Alexandra Duel-Hallen, Dr. Mihail Devetsikiotis, Dr. Wenye Wang, for their time, advices, and encouragements.

My fellow graduate students and friends, Jie Hui, Li Ma, Kamala Subramaniam, Xinbing Wang, and Peng Xu, had extensive discussions with me and gave me a lot of useful advices. They also made my four-year study at Raleigh enjoyable and colorful. I will cherish the friendship throughout my life.

Without supports from my family, this Ph.D. study will not be done. I'm indebt to my parents. I know how difficult it is for them to see daughter go hundreds of thousands miles away to another continent for several years. Their love, their understanding, and their support are the source of my energy, which has been holding me on the track from the elementary school until the Ph.D. study. I want to thank my brother for his support. We have so many happy memories together. These happy time gives me courage when I feel tired. Finally, I want to say "Thank you." to my husband, Xiaojun. I can not imagine how my life would be without his love that makes me love the life, his encouragement that gives me perseverance, and his humor that makes me laugh and be optimistic. This work belongs to the whole family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Wireless communication has experienced exponential growth starting from the late 1980s. Cellular networks have evolved from the first generation (analog technology), the second generation (digital technology) to the current 3G (the third generation) and beyond technology. Wireless local area network (WLAN) was proposed in the late 1990s and rapidly penetrated into the market. The inter-operation of cellular networks and WLANs will bring more prospects for the wireless communication.

With the development of wireless networks, the wireless applications demand more bandwidth and higher speed. The wireless terminal needs more energy to support the applications. However, the wireless terminals are always powered by batteries. The battery technology had no breakthrough for decades and its development is not promising in the near future [1][2][3][4][5]. The limited battery lifetime constrains the usage of wireless terminals and the advance of wireless networks. Energy efficiency has become an important issue in wireless networks.

Great effort has been made to improve the energy efficiency in wireless networks. The techniques to save energy have been developed for wireless devices (such as CPUs [6], network interfaces [7]), wireless protocols (such as CSMA (Carrier Sense Multiple Access) protocol [8], routing protocol [9], TCP (Transmission Control Protocol) [10]), and wireless applications [11].

We study the energy efficiency issue in both the IEEE 802.11 based WLAN and the 3G system UMTS (Universal Mobile Telecommunication System).

1.1 Power management in WLAN

In WLAN, we focus on the MAC (Media Access Control) sublayer. The IEEE 802.11 specifies the power management scheme to address this issue [12]. The standard defines two power states for a station. One is the Awake state, in which a station is fully powered and can transmit, receive, and listen to the network. The other is the Doze state, in which a station demands very low power, but can not transmit, receive or listen to the network. Since a station is not busy (transmit/receive) all the time, the action of idle listening wastes lots of energy. The aim of the power management scheme defined in the standard is to reduce the idle listening time by putting the stations into the Doze state as much as possible. An overview of power-saving mechanisms for WLAN is given in [13]. Stine and Veciana studied the energy conservation in a central controller of a wireless network [14]. They reduced the set of nodes in the receiving state as much as possible and put forward some mechanisms that can be implemented in the IEEE 802.11 Point Coordination Function (PCF). Many groups focus their work on the energy efficiency of Ad hoc WLANs. Bononi, Conti and Donatiello enhanced the power saving scheme in Ad hoc WLANs [15]. They added a distributed contention controller (DCC) between the standard access scheme and the physical layer to modify the access probability according to the real time contention level, which results in the decrease of collision probability and the efficiency of power usage. Jung and Vaidya modified the power management scheme of the Ad hoc configuration in the IEEE 802.11 [16]. They proposed to choose the ATIM (Ad hoc Traffic Indication Message) window size dynamically according to network conditions and permit the stations to stay awake for only a part of the beacon interval following the ATIM window. Liu suggested to use a variable length of beacon

interval and contention free schedule for the transmission of data packets in Ad hoc WLANs [17].

The purpose of our study is to find the power management schemes for the IEEE 802.11 based WLAN that achieve maximum energy efficiency without degrading the system performance. We model the power management analytically and give suggestions about how to optimally configure the power management parameters. We also propose the enhanced power management schemes.

We study how to optimally configure the parameters of the original power management scheme specified by the IEEE 802.11 protocol. For an infrastructure network, we find the optimal parameter through analysis and simulation. Most research about the parameter configuration are based on simulation. Although some intuitive results or simulation results can be applied for parameter configuration, there is no analytical models to support them. To the best of our knowledge, modeling the power management scheme from the queueing point of view is an open topic. We model the power management scheme of an infrastructure configuration in the IEEE 802.11 through two different perspectives: the M/G/1 queue with bulk service and the D/G/1 queue. We find the analytical results for the average response time of a frame and the percentage of time a station stays in the Doze state from both models, which are strongly supported by the simulation results. For an independent WLAN, we find the optimal power management parameters by simulation.

We also propose the enhanced power management schemes for the infrastructure mode WLAN and the independent WLAN respectively. In an infrastructure network, we propose to setup multiple queues in the access point (AP) buffer and apply scheduling algorithms. In an independent network, we propose the contention free transmission of the actual traffic by scheduling.

1.2 Power management in UMTS

Cellular systems use the same basic mechanism as WLAN to save energy, which puts a user equipment into the low power consumption state when it is idle and wakes it up periodically to listen to the network. Based on the same idea, different cellular systems use different parameters to achieve energy savings.

CDPD (Cellular Digital Packet Data) schedules the sleep-mode operation by the inactivity timer T203 and TEI (Temporary Equipment Identifier) timer T204 [18] [19]. The CDMA2000 supports different MAC states, which consist of Control Hold State (CHS), Active State (AS), Suspended State (SS) and Dormant State (DS). The MAC states consume different level of power. The CDMA2000 uses timers to trigger the state transitions. If the state transitions are too frequent, the signal overhead is high and the power saving is not significant. If the state transitions are too few, the delay may be too large. So and Cho studied the tradeoff of timers for sleep mode operation in CDMA2000 system [20].

UMTS specifies four Radio Resource Control (RRC) states, including CELL_DCH, CELL_FACH, CELL_PCH and URA_PCH [21]. Different states provide different communication ability and consume different level of power. Like in the CDMA2000 system, the state transitions in UMTS are triggered by timers as well. Yeh evaluated the performance of energy consumption in UMTS based on the impact and configuration of timers [22]. The tradeoff between energy consumption and delay is discussed. The inactivity timer also impact the system capacity because the system resources are occupied when the user equipment is idle within the inactivity timer value. Chuah investigated the impact of inactivity timer on UMTS capacity [23]. Lee compared the power management of UMTS and CDMA2000, especially the impact of inactivity timer [24].

While many research focus on the impact of inactivity timer that controls the state transitions, Yang looked into both inactivity timer and Discontinuous Reception

(DRX) mechanism [25]. However, the arrival traffic in Yang's study is assumed as Poisson. We know that one of the goal in 3G is to provide multimedia services, but Poisson distribution no longer suits data traffic. In our study, we use the data models suggested in 3GPP (Third Generation Partnership Project) to investigated the performance of power saving mechanism in UMTS.

We study both DRX mechanism, which is controlled by DRX cycle, and inactivity timer for energy saving in UMTS. We use the typical traffic models specified in 3GPP to evaluate the performance of power consumption. The aim of our study is to find optimal parameters of DRX cycle and inactivity timer for power saving mechanism. From simulation results, we know that different traffic types demand different optimal parameters. Therefore, we propose to dynamically set the DRX cycle and inactivity timer based on real-time measurement of buffers.

1.3 Organization

Our work is organized as following:

- In chapter 2, we give the brief review of energy efficiency techniques in wireless networks, WLAN, UMTS, and scheduling algorithms.
- In chapter 3, we describe the power management scheme in the IEEE 802.11 and introduce the performance metrics used in our study.
- In chapter 4, two queueing models for the power management in the infrastructure mode IEEE 802.11 are discussed and the analytical results for the performance metrics are obtained.
- In chapter 5, we present the simulation results, which are in good agreement with the analysis. We also propose and simulate the enhanced power management schemes for WLAN.

- In chapter 6, we introduce the power saving mechanism in UMTS and the performance metric used for evaluation.
- In chapter 7, we study the power management scheme in UMTS by simulation. We find the optimal parameters for different traffic models and propose to adaptively change the parameters for different traffic types by real time measurement.
- In chapter 8, we draw the conclusion and present the perspective.

Chapter 2

Literature Review

This chapter provides the brief review of energy efficiency techniques in wireless networks, the IEEE 802.11 based wireless LAN, UMTS, and scheduling algorithms.

2.1 Energy efficiency in wireless networks

Wireless networks have been experiencing rapid development over one decade. Now, it is more promising when wireless networks merge to 3G and beyond. However, the limited battery capacity is one of the biggest constraints for the development of wireless applications. How to achieve energy efficiency in wireless network becomes an urgent and active research area. Energy issue in wireless networks is investigated in multiple layers of the OSI (Open System Interconnection) network protocol stack. We introduce the proposals for each layer in the following [26].

Physical layer

The energy efficiency problem can be addressed in two different areas in the physical layer: battery technology and low-power hardware design.

Low-power hardware design includes the technology related to CPU, memory and disk.

The requirement for battery by wireless networks include the increase of battery

capacity and the decrease of battery weight and size. In the past 30 years, there is no breakthrough in the battery technology. It is not optimistic that significant improvement will happen in the near future. Therefore, researchers turn to higher layers of network protocol for the relief of energy issue.

MAC sublayer

The sources of energy inefficiency in the MAC layer consist of collisions and idle listening. The purpose of the energy efficiency schemes in the MAC layer is to reduce collisions and idle listening. There are three prevalent MAC protocols that addressed the energy efficiency issue.

IEEE 802.11 standard The IEEE 802.11 standard specifies two power states for a station: the Doze state and the Awake state. When a station is idle, it is put into the Doze state as much as possible to reduce idle listening and save energy. Our study focuses on how to improve energy efficiency in the MAC sublayer of the IEEE 802.11.

EC-MAC protocol The EC-MAC (Energy Conserving-Medium Access Control) protocol [27] is designed for an infrastructure network, which includes a central base station and mobile stations. The EC-MAC protocol use reservation and scheduling to save energy. In the EC-MAC protocol, the base station sends a frame synchronization message (FSM) to synchronize all the stations. The transmission frame is divided into several phases: Request/Update phase, New user phase, Scheduling message phase, Data phase (downlink) and Data phase (uplink). Because of the reservation and scheduling strategies, collisions are significantly reduced.

PAMAS protocol The PAMAS (Power Aware Multi-Access) protocol [28] was designed for the Ad hoc network to save energy. The PAMAS uses a separate control channel for the channel request message RTS (Request To Send)/CTS (Clear To Send). A station can decides its power state by probing the control channel.

LLC sublayer

In the LLC sublayer, automatic repeat request (ARQ) and forward error correction (FEC) are two common techniques. However, the retransmission and overhead brought by these two techniques demands lots of energy in wireless networks.

Many research focus on this area and try to reduce the number of retransmission, such as adaptive error control with ARQ, adaptive error control with ARQ/FEC combination.

In the LLC layer, the power control combined with the coding scheme can also be used to improve channel quality and energy efficiency.

Network layer

The network layer provides the routing and congestion control functions. Typical routing metrics include shortest-path, shortest-delay, etc. The routing protocol based on these metrics overuse a small set of nodes, whose energy may deplete quicker than others, thus the network life is decreased. The routing protocols with energy efficiency goal add energy related parameters to routing metrics [29].

Transport layer

Transport layer provides end-to-end service by retransmission and congestion control. TCP is one of the most popular transport layer protocol. The purpose of the modified versions of TCP for wireless networks is to reduce the number of retransmission to save energy. The schemes for reducing retransmission include: split connection protocols, link-layer protocols and end-to-end protocols.

OS/middleware and application layers

One technique used in this layer is to slow down the CPU speed for mobile applications to save energy. The speed reduction can be compensated by pipelining and parallelism technique for computer architecture. The other technique at this layer is

to shutdown the system based on the prediction of inactivity period. The shutdown scheme brings in the increase of latency. The software architecture impact on the power consumption is also studied [30].

Application layer

The function of the application layer includes partitioning of tasks, source coding and context adaption. The research aimed to energy efficiency in this layer is related to load partitioning, proxies, databases, video processing, etc.

In conclusion, energy efficiency is a crucial topic in wireless networks. It can be addressed in every layer of the network model. Furthermore, cross-layer consideration can also bring good energy efficiency scheme. Our study focus on the MAC layer.

2.2 Review of the IEEE 802.11 based WLAN

2.2.1 Background

The IEEE 802.11 [12] working group was formed in 1987 to study WLAN and the IEEE 802.11 standard was approved in 1997. WLAN products were designed to work in the ISM (Industrial, Scientific and Medical) radio bands (900MHz, 2.4GHz and 5GHz). The original IEEE 802.11 standard has been extended to a series of specifications. The IEEE family defines the WLAN technology.

Within the IEEE 802.11 family, the original IEEE 802.11, the IEEE 802.11a, the IEEE 802.11b and the IEEE 802.11g focus on encoding schemes, while other protocols are for service enhancement and extensions. *The original 802.11* works in the 2.4GHz band and provides the data rate of 1 or 2 Mbps. It uses IR(infrared), or FHSS (Frequency Hopping Spread Spectrum) or DSSS (Direct Sequence Spread Spectrum) encoding scheme. *The 802.11a*, an extension to the 802.11, works in the 5GHz band

and supports up to 54Mbps data rate. It uses OFDM (Orthogonal Frequency Division Multiplexing) encoding scheme. *The 802.11b*, an extension to the original 802.11 and also known as the 802.11 High Rate or Wi-Fi (Wireless Fidelity), is the most widely accepted protocol of the IEEE 802.11 family now. It works in the 2.4GHz band and provides the data rate of 1, 2, 5.5 and 11Mbps. It uses DSSS encoding scheme. *The 802.11g* extends the 802.11 b. The data rate goes from 11Mbps to 54Mbps, but the wireless transmission distance is relatively short. It applies OFDM encoding scheme. *The 802.11e* standard focuses on the QoS (Quality of Service) in WLAN and *the 802.11i* is designed to enhance the security in WLAN.

2.2.2 Network structure

Two types of network structures are defined by the standard: the independent network and the infrastructure network.

The basic components in an IEEE 802.11 WLAN include the station, the BSS (Basic Service Set), the DS (Distributed System) and the ESS (Extended Service Set).

Structure of an independent network

In an independent network, a BSS includes several stations that communicate to each other directly. Fig.2.1 shows three independent BSSs. Within each BSS, the stations can communicate to each other, while no communication between stations from different BSSs.

An independent BSS is also called an Ad hoc network. It doesn't need network planning beforehand or established infrastructure. The disadvantage of an independent BSS is its limited coverage. In an Ad hoc network, if two stations locates within the radio transmission range of each other, they communicate directly. Otherwise, they can communicate if other stations of the same BSS can forward packets for them.

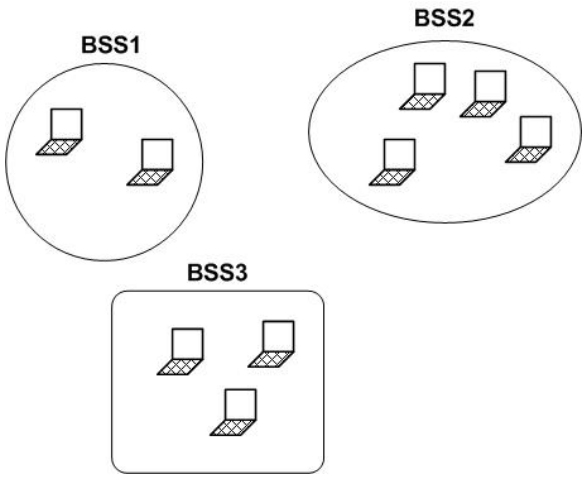

Figure 2.1: Structure of independent BSSs.

The former is the single-hop case and the latter is the multi-hop case. In this study, we focus on the single-hop network.

Structure of an infrastructure network

An infrastructure BSS includes stations and an AP. It behaves as a client-server structure, where the AP is the server and the stations are the clients. An AP is a special station which can access a DS. A DS is a backbone for connecting the BSSs within an infrastructure network or interconnecting the BSSs to other networks in order to extend the service coverage. The network extended by the BSSs and the DS is called ESS. The ESS is transparent to the LLC (Logic Link Control) sublayer in the network. The LLC still view the ESS as an independent BSS.

Fig.2.2 shows the structure of an infrastructure network. The stations communicate through the AP. If a station communicates with another station within the same BSS, it forwards the traffic to the AP first, then the AP forwards the traffic to the destination station. If a station communicates with another station from other BSSs, it forwards the traffic to the AP of its own BSS first, then the traffic is forwarded to the AP of the destination station through the help of the DS. Finally, the AP of the

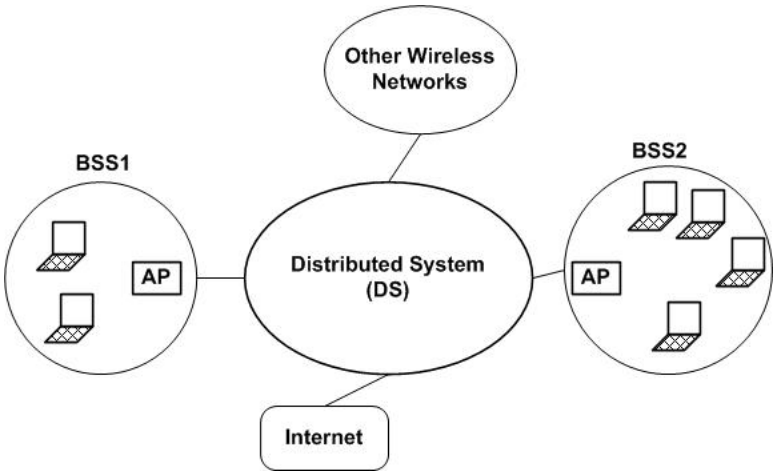

Figure 2.2: Structure of infrastructure BSSs and the network interconnection.

destination station forwards the traffic to the destination station. Also, a station can communicate with the host in the Internet or other wireless networks through the AP and the DS.

2.2.3 MAC sublayer and PHY (physical) layer

The IEEE 802.11 standard gives the specifications for the MAC sublayer and the PHY (Physical layer) in the OSI reference model. The key technologies used in the MAC sublayer and the PHY layer are introduced in the following [31].

MAC Sublayer

The basic access method of the IEEE 802.11 standard is the distributed coordination function (DCF), which is known as the carrier-sense multiple access with collision avoidance (CSMA/CA).

The protocol defines two ways for carrier-sense: the physical carrier-sense and the virtual carrier-sense. They are combined to determine the busy/idle state of the medium. The physical carrier-sense mechanism is provided by the PHY layer. It determines if the channel is clear through performing the clear channel assessment

(CCA) algorithm, which is related to the received signal strength. The virtual carrier-sense scheme is implemented by predicting the future traffic and reserving the medium before the actual data is sent. First, a source station sends out an RTS message and waits for the CTS message from the destination station or the AP. Both the RTS and the CTS messages contain the predicted duration to be reserved for data transmission. If the message exchange is successful, all other stations that hear it mark the medium status as busy and defer their transmission during the reserved period. Then the source transmits the actual data and the destination gives back an ACK (Acknowledgement) to complete the process. If the RTS/CTS message exchange fails, the source station goes back to compete for the media and the process starts over.

In order to avoid collision, a station senses the media status before its transmission. If the media is idle, the station waits for certain interframe space (IFS) time and transmits. Otherwise, the station defers its transmission, waits for some time, then tries again. This process is called backoff procedure. The time interval a station waits is called the backoff time. If more than one stations transmit at the same time, collisions happen and the stations undergo a backoff procedure. A station repeatedly experiences the backoff procedure until a successful transmission is done. The IEEE 802.11 uses the Binary Exponential backoff scheme to determine the backoff time.

The IEEE 802.11 standard also defines an optional access method for an infrastructure BSS: point coordination function (PCF). The PCF uses a pointer coordination (PC), which is located in the AP of an infrastructure BSS, to schedule the transmission sequences through polling. The access priority of the stations can be implemented by using different polling algorithms. Also, with the PCF, the collisions can be avoided.

The IEEE 802.11 standard uses the interframe space to control the priority of different functions, services, and messages in the system. The IFS is the minimum time a station has to wait after it determines the medium is idle through the carrier

sense function. The shorter IFS leads to the higher priority. Four priority levels for access to the wireless media are provided by the standard, which include SIFS (Short interframe space), PIFS (PCF interframe space), DIFS (DCF interframe space), the EIFS (Extended interframe space) with the priority order from high to low.

Energy efficiency is an important performance metric in a wireless network. The IEEE 802.11 specifies the power management scheme to address this issue in the MAC management protocol. The standard defines two power states for a station. One is the Awake state, in which a station is fully powered and can transmit, receive, and listen to the network. The other is the Doze state, in which a station consumes very low power, but it can not transmit, receive or listen to the network. Since a station is not busy (transmit/receive) all the time, the action of idle listening wastes lots of energy. The aim of the power management scheme defined in the standard is to reduce the idle listening time and put stations into the Doze state as much as possible. The standard gives the power management scheme for the infrastructure BSS and the independent BSS respectively. The purpose of our study here is to find power management schemes that achieve maximum energy efficiency without degrading the system performance.

PHY layer

The IEEE 802.11 standard defines multiple PHYs for different medium environment. The *IR PHY* is designed for indoor activities with the data rate of 1 and 2 Mbps. The *FHSS PHY* is resistant to multiple fading through the inherent frequency diversity mechanism. It works at 2.4GHz radio band and can provide the data rate of 1 or 2 Mbps. The *DSSS PHY* works well in the environment where the time delay spread is serious (such as echo). It works at the 2.4GHz radio frequency as well and can achieve the data rates of 1,2,5.5, and up to 11Mbps. The *OFDM PHY* operates at 5GHz radio band to provide higher data rate with the range of 6-54Mbps. The functions that supports the frame exchange between PHY layer and MAC sublayer

are also specified in the standard.

With the enlarging of the WLAN market, the inter-operation of the WLAN with other networks (such as UMTS) are becoming desired features and drawing lots of research attention.

2.3 Review of UMTS

2.3.1 Background

With the rapid growth of wireless communications, cellular networks undergo from the first generation analog system, the second generation digital system, to the current third generation system. The services also revolve from voice, data to multimedia, etc. The 3G technologies were proposed to meet the requirements of higher data rate, asymmetric and packet-switched traffic, which enable many IP-based applications to run in the wireless cellular networks.

The International Telecommunication Union (ITU) proposed IMT2000 (International Mobile Telecommunication) standards to satisfy the requirements. In 1998, two groups (3GPP and 3GPP2) were established to work on the standards. 3GPP mainly focuses on UMTS and GPRS/EDGE (General Packet Radio Service / Enhanced Data Rates for GSM Evolution) systems. 3GPP2 mainly focuses on CDMA2000 system.

We study the power saving mechanism in UMTS and give a brief overview of UMTS here.

2.3.2 UMTS services and applications

The first and second generation cellular networks mainly focus on providing voice communication. The goal of the third generation cellular networks is to support higher

data rate, lower delay and more variety of communications. Here, we introduce the services and applications supported by UMTS.

Person-to-Person services

- Video telephony
- Images and multimedia
- Push-to-Talk over Cellular (PoC): This is a unidirectional service. A short speech burst is transferred through packet networks, which improves bandwidth efficiency.
- Voice over IP (VoIP)
- Multiplayer Games

Content-to-Person services

- Web browsing
- Audio and video streaming
- Content download
- Multimedia broadcast multicast services

Business connectivity

This feature provides end-to-end security and considers the latency to the application caused by radio networks.

Quality of services differentiation

UMTS defines four QoS classes: conversational class, streaming class, interactive class and background class. The main parameters used to differentiate the classes include transfer delay, guaranteed bit rate, traffic handing priority and allocation/retention priority.

Location services

Location services include emergency calls, information broadcast in certain area, and location based service inquiry, etc.

2.3.3 CDMA concepts

UMTS uses WCDMA (Wideband Code Division Multiple Access) as air interface. We introduce the basic CDMA concepts in the following.

Spread spectrum concept

Spread spectrum communication is to transmit information over channels much wider than required by the original signal bandwidth. It means the energy of the information signal is spread over a much wider band channel. From Shannon's capacity equation,

$$C = W \times log(1 + SNR) \tag{2.1}$$

where, C is the system capacity, W is the channel bandwidth, and SNR is signal to noise ratio, we know that the signal can be transmitted with very low power spectral density.

Two types of spectrum spreading technologies are commonly used.

- Direct Sequence Spreading: multiply each information bit by a high-rate spreading sequence.
- Frequency Hopping: random hopping of narrow band subcarriers within a wide range of spectrum.

3G CDMA systems select Direct Sequence Spreading. It works as following. At the transmitter, a high-rate pseudo-random noise-like sequence multiplies with the information bits. At the receiver, the spread spectrum signal is correlated with the same spreading sequence. Here, each user selects different code to spread its signal and the cross-correlations between the codes are zero or very small. Only the original

signal is de-spread and its power spectral density is increased. All signals from other users which use different codes remain wideband, have almost no power at the output of the correlated receiver, and appear as interference.

Spread codes

Orthogonal codes are the foundation of CDMA technologies. Two kinds of orthogonal codes are used in CDMA systems.

Walsh codes are a set of ideally orthogonal codes generated by the Hadamard matrix [32]. PN (psedu-noise) codes are a set of pseudo-random, noise like, but deterministically generated sequence. The generator is usually a linear shift register with feedback.

In the forward link, Walsh codes are used for channelization, which separates different physical channels transmitted by a base station. Short PN codes (with period of $2^{15} - 1$) is used to distinguish different base stations, which use different offset.

In the reverse link, long PN codes (with the period of $2^{42}-1$) are used to distinguish different users. A mask used to generated the long PN code for reverse user traffic channel and access channel is assigned by the system.

Multipath radio channels and Rake receiver

Three basic mechanisms behind radio propagation are reflection, diffraction and scattering [33]. They cause multipath fading. The two effects of multipath propagation are,

- The signal energy may arrive at the receiver at different time instant.
- The signal at a specific time may consist of signals coming from different paths.

On one hand, the multipath propagation makes the received signal in the radio channel become more difficult to detect and decoding. On the other hand, it provides diversity. Rake receiver makes use of the multipath diversity. It estimates the

relative timing and strength of primary multipath components, then combines them to overcome the multipath fading.

Soft handoff

Handoff is a process in cellular network which allows mobiles to switch between cells without disconnecting ongoing calls. In TDMA-based systems, different channel operates at different frequency. A mobile breaks with the old serving cell first, then connects with the new serving cell. This is so called hard handoff. In CDMA-based systems, all base stations use the same frequency. A mobile connects with the new cell before breaking up with the old cell. This smooth handoff is called soft handoff.

Power control

In CDMA systems, mobiles operate in the same frequency and are distinguished by codes. If a mobile near the base station uses the same power as the other mobile far from the base station. It can cause very high interference to the far-away mobile and block the far-away station from accessing the network. This is so-called near-far problem.

The ideal case is to make the signals from all mobile stations in a cell arrive at the base station with the same signal level. CDMA systems use power control to achieve this goal. Power control includes forward power control and reverse power control. It is one of the keys to achieve good performance in CDMA systems.

Soft capacity

The cell capacity and coverage are interdependent in a CDMA system. The forward link capacity is mainly limited by the maximum total transmission power of a base station, the available channel elements and Walsh codes. The reverse link capacity is interference limited and the bottleneck of network capacity.

The reverse link capacity can be calculated by

$$N_{reverse} = \frac{W/R}{(E_b/N_0)_{min} \times \overline{v}(1+f)} \times L \tag{2.2}$$

where, W is the bandwidth, R is the data rate, W/R is the spreading gain,

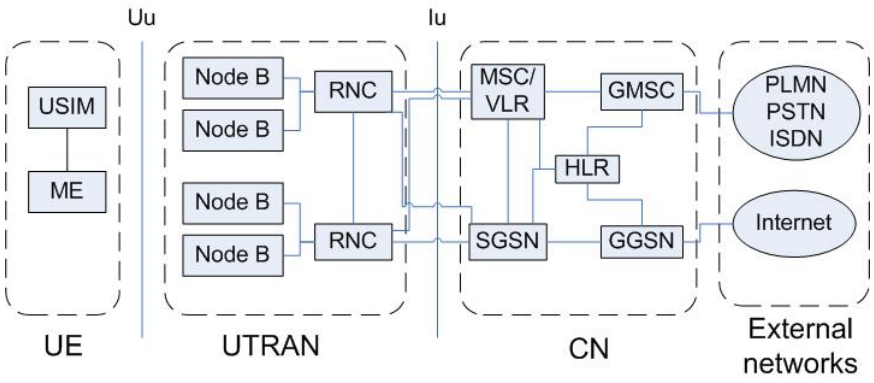

Figure 2.3: UMTS architecture and external networks.

$(E_b/N_0)_{min}$ is the minimum bit energy over noise spectral density required by the receiver. $\overline{v}$ is the voice activity factor for voice calls, f is the interference factor, which is the ratio of other-cell interference to same-cell interference, L is the loading factor, which is limited by coverage stability and handoffs.

2.3.4 UMTS architecture

UMTS architecture consists of three parts: UE (User Equipment), UTRAN (UMTS Terrestrial Radio Access Network) and CN (Core Network). Fig.2.3 shows the UMTS architecture.

Here,

- ME: Mobile Equipment
- USIM: UMTS Subscriber Identity Module
- Node B: also called base station
- RNC: Radio Network Controller, mainly for radio resource control and service access point to core network.

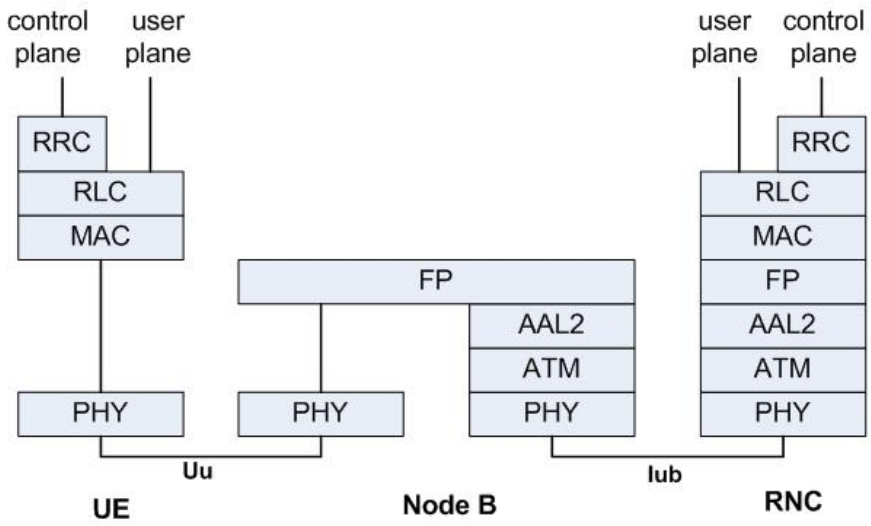

Figure 2.4: A simplified protocol diagram between UE and UTRAN.

- MSC/VLR: Mobile Service Switching Center / Visitor Location Register. They serve UE in its current location for circuit switched services.
- HLR: Home Location Register
- GMSC: Gateway MSC, the switch where UMTS and external circuit switched network are connected.
- SGSN: Serving GPRS (General Packet Radio Service) Support Node, serves the UE for packet switched services.
- GGSN: Gateway GPRS Support Node, the switch to connect UMTS and Internet.

A simplified protocol diagram between UE and UTRAN is shown in Fig.2.4. Here,

- RLC: Radio Link Cotrol
- FP: Frame Protocol

- AAL2: ATM Adaption Layer type 2
- ATM: Asynchronous transfer mode

2.3.5 Radio resource management in UMTS

Radio resource management (RRM) is one of the important issues in wireless networks. In UMTS, the goal of RRM includes:

- Guarantee Quality of Service (QoS)
- Maintain the planned coverage area
- Offer high capacity
- Manage power of mobile equipment

The algorithms of RRM include:

- Handover control: Handle the mobility of the UEs across cell boundaries.
- Power control: Keep the interference levels at minimum in the air interface.
- Admission control: Guarantee QoS. The strategy can be wideband power-based admission control, throughput-based admission control or priority-based admission control [34].
- Load control: Maintain the stability of the system.
- Packet scheduling: Guarantee QoS and efficiency use of radio resources.
- Power management scheme: Improve the energy efficiency of mobile equipments.

Our study focuses on how to improve the energy efficiency of mobile equipment power management. Since UMTS provides higher bandwidth and more services and applications than 2G systems, how to efficiently manage mobile power is becoming one of the key factors to promote the application of UMTS.

2.4 Scheduling algorithms

Scheduling has been a very active topic from the late 1980s to present driven by the rapid development of networks. Applying scheduling discipline can provide fairness and performance guarantee in networks. In wireless networks, energy efficiency can be improved by properly scheduling the transmission of the traffic.

2.4.1 Background

In order to understand scheduling algorithms better, several background concepts are introduced here.

Best-effort service and guaranteed service

There two concepts are used to indicate the different requirements of applications in networks. A best-effort service does not provide full reliability. It tries to give the best performance, but not guarantee it. A guaranteed service provides performance guarantee in networks.

Work-conserving and non-work-conserving discipline

A work-conserving scheduler is idle only when there is no packet to serve, while a non-work-conserving scheduler can be idle even if there are packets that are waiting service. Under work-conserving discipline, network resources can always be fully utilized. On the contrary, non-work-conserving discipline wastes network resources in some extent and is more complex to implement. The advantage of non-work-conserving server is that it can shape traffic to ensure its delay and delay jitter, and thus reduce the

output buffer size of servers.

2.4.2 Description of scheduling algorithms

This section introduces some commonly used scheduling algorithms.

FCFS (First Come First Served)

FCFS, also known as FIFO (First In First Out), is the simplest and widely used scheduling algorithm, which serves packets in the order of their arrivals.

Priority scheduling

Priority scheduling scheme assigns each connection a priority level. A high priority connection is always served before a lower priority connection.

Round Robin (RR)

RR is often used to schedule the best effort connections. The packets belong to each connection is buffered in one queue. It serves one packet from each queue each time. Within one cycle, it can visit each queue once.

Generalized Processor Sharing (GPS) is an ideal version of RR. It operates by serving an infinitesimally small amount of data from each queue each time. Thus, in any finite time interval, it can visit every queue at least once.

Weighted Fair Queueing (WFQ)

WFQ operates by emulating GPS. It is first really fair queueing discipline and has been adopted by router manufacturers.

Earliest Due Date (EDD)

EDD assigns each packet a deadline. The packet with the earliest due date is served first [35]. EDD can give guaranteed delay bound.

In our study, we use scheduling algorithms to enhance power management in WLAN.

Chapter 3

Power Management in WLANs

This chapter discusses the power management in the MAC layer of WLAN. The power management mechanism specified in the IEEE 802.11 is introduced first. Then, we define the performance metrics used in our study for measuring the energy efficiency of WLAN. Finally, we identify the sources of energy inefficiency in WLAN and propose general principles to save energy.

3.1 Power management mechanisms in the IEEE 802.11

Energy efficiency is one of the most important issues in wireless networks. The IEEE 802.11 standard specifies power management schemes to save energy in WLAN.

Two power states are defined for a station: the Awake state and the Doze state. The Awake state consumes far more energy than the Doze state. The basic principle of the power management of the IEEE 802.11 is to put stations into the Doze state as much as possible while satisfying the system performance requirements.

The standard defines two power management modes for a station: the active mode and the power save mode. In the active mode, a station stays in the Awake state all the time and can transmit/receive frames at any time. In the power save

mode, a station stays in the Doze state when it is idle and enters the Awake state to transmit/receive traffic periodically. The system provides the mechanism for a source (either a station in an independent network or an AP in an infrastructure network) to determine the power management mode of its destination station. If the destination station is in the power save mode, the source needs to buffer the frames temporarily and transmit them later at the time designated by the standard. Otherwise, the source forwards the frames to the destination station.

The IEEE 802.11 standard specifies the power management mechanisms according to the different network structures. We introduce the power management schemes in an infrastructure network and an independent network respectively in the following.

3.1.1 Power management in an infrastructure BSS

In an infrastructure network, the AP acts as the server. When two stations want to communicate, the source station sends its frames to the AP first, then the AP forwards the frames to the destination. When the destination station in an infrastructure BSS is in the Doze state, the AP temporarily buffers the frames first, then find an appropriate time to forward them. The IEEE 802.11 standard uses the beacons to schedule the transmission time for the buffered frames. A beacon is a system frame used to synchronize all the stations within a BSS. In an infrastructure BSS, the AP periodically transmits the beacons and all the stations synchronize their clock with the AP. Every beacon piggybacks a traffic indication map (TIM) message that can identify the stations that have traffic buffered in the AP. The stations stay in the Doze state when idle and listen for the beacons periodically. The time duration a station periodically wakes up to listen for beacons is called listen interval. A listen interval is equal to one or more beacon intervals.

If a station is indicated that it has traffic pending in the AP by the TIM, it stays awake until the buffered traffic is received. Otherwise, it goes to the doze state

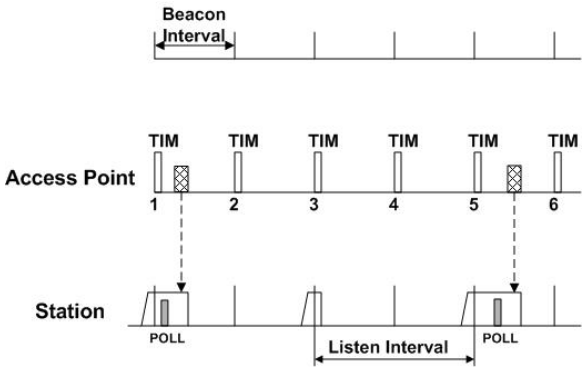

Figure 3.1: Infrastructure BSS power management operation.

immediately.

Fig.3.1 shows the operation of the power management mechanism in an infrastructure BSS. The TIMs are announced at the beginning of each beacon interval by the AP. The listen interval of the station is two beacon intervals. In other words, the station enters the Awake state to receive the TIMs every two beacon intervals. The first TIM indicates that there is pending traffic in the AP destined for the station. The station stays awake and sends a polling frame to retrieve the traffic buffered in the AP. If no collision happens, the station receives the traffic from the AP and goes to the Doze state until the beginning of the next listen interval. If a collision happens, the station backs off for some time and polls the AP again. The station does not listen to the second TIM according to its power management scheme. From the third TIM, the station knows that there is no buffered traffic for it in the AP. So it goes to the Doze state immediately.

In an infrastructure BSS, the stations in the active mode can always transmit/receive frames, they do not need to interpret the TIMs. Each station that in the power save mode can achieve its own power management requirement by setting the length of its listen interval.

In our study, we suggest how to configure the listen interval of a station in power save mode to improve energy efficiency.

3.1.2 Power management in an independent BSS

The power management mechanism in an independent BSS works in the similar way as that in an infrastructure BSS.

A station has two power states: the Awake state and the Doze state. It can works in one of the two power management modes: the active mode and the power save mode. The stations in an independent BSS communicate with each other directly and can estimate the power management mode of other stations through the mechanism provided by the system. If the destination station is in the active mode, it can receive traffic at any time. Otherwise, the source station has to temporarily buffer the outgoing traffic and transmit it at the time specified by the standard.

The frames destined to the stations in the power save mode are announced during a time period when all the stations stay awake. Then, the stations that have traffic to transmit/receive stay awake and others enter the Doze state.

An independent BSS also uses beacons to synchronize the time of all the stations in it. The time instant the beacon is scheduled to transmit is the target beacon transmission time (TBTT). Prior to the TBTT, all stations in the power save mode wake up to listen for beacons. Following the TBTT, all the stations stay awake for some time, during which the Ad hoc traffic indication messages (ATIM) are announced by the source stations to indicate the buffered traffic of the destination station. If the announcement is sucessful, the destination station sent back an acknowledgement for the directed ATIM. No acknowledge for the multicast ATIMs. A station in the power save mode listens for the announcement to decide if it needs to stay awake in the entire beacon interval or not. The time in which the ATIMs are announced is defined as the ATIM window. During the ATIM window, only beacons and ATIM/ATIM ACK can be transmitted. From the end of the ATIM window to the end of the beacon interval, all stations that have traffic to transmit/receive stay awake to exchange the actual traffic. The media access mechanism used is the normal DCF.

Fig.3.2 illustrates the operation of the power management in an independent BSS.

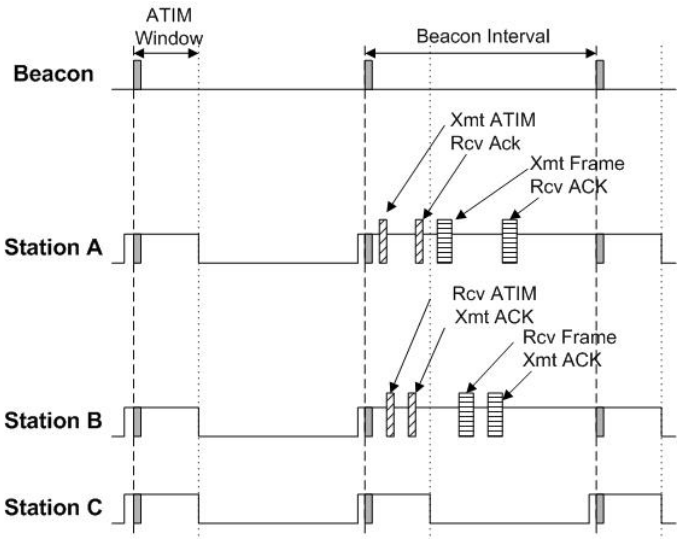

Figure 3.2: Independent BSS power management operation.

All stations (A, B, C) wake up prior to the TBTT and stay in the Awake state during the ATIM window. In the first ATIM window, there is no ATIM frames exchanged since none of them has traffic to transmit/receive. All three stations go to the Doze state after the ATIM window until the next TBTT. In the second ATIM window, the station A has messages destined to the station B and transmits an ATIM frame to announce it. The station B listens to announcements, receives the ATIM frame from A, then sends an acknowledgement back to A. A gets the acknowledgement successfully. Hence, both A and B stay awake during the entire beacon interval to exchange the traffic since the station A has traffic to transmit and the station B has traffic to receive. The station C still has no traffic to transmit/receive in the second beacon interval. It goes to the Doze state after the ATIM window to save energy.

An appropriate ATIM window size can optimize the energy efficiency. If the ATIM window is too small, not all the buffered traffic can be announced within the ATIM window. The buffered traffic is accumulated and the system latency increases. If the ATIM window is too big, the time that all stations stay awake increase and the time interval for the transmission of the actual traffic becomes smaller. This may results

in the announced traffic can not finish the transmission in the current beacon interval and have to be postponed to the next beacon interval.

We will study the impact of the ATIM window size on the energy efficiency and system latency, then give suggestions for the selection of the ATIM window size.

3.1.3 Packet delay with power management

Delay is an important metric to evaluate performance of a network. We will use frame response delay as one of our performance metrics.

With power management, in an infrastructure network, the delay of a frame includes the delay from a source station to the AP, the delay in the AP, and the delay from the AP to the destination station. In an independent network, the delay of a frame includes the delay in the buffer of a source station and the delay from the source station to the destination station.

In this section, we introduce the calculation of delay between a station and an AP. The transmission delay between two station is the same as the transmission delay of a station to the AP.

Many research have evaluated the performance of WLAN [36][37][38], especially saturation throughput and delay analysis.

We call the transmission from a station to an AP as upload and the transmission from an AP to a station as download. Fig.3.3 shows the transmission of upload and download. We denote the average packet delay of upload and download as D_{up} and D_{down} respectively. We refer to Bianchi's result [36] to calculate D_{up} and D_{down}. Bianchi studied saturation throughput of IEEE 802.11. The average packet delay under saturation condition is easily obtained as $D = \frac{P}{R_s}$, where D denotes the average packet delay, P denotes the packet length, and R_s denotes the saturation throughput. The saturation throughput is (Because of page limitation, please refer [36] for the

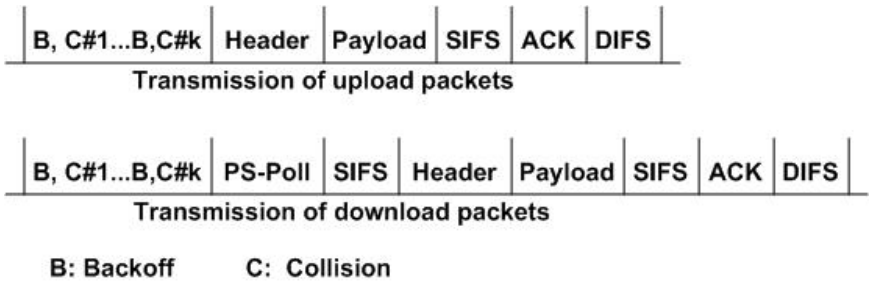

Figure 3.3: Transmission of upload and download packets.

derivation),

$$R_s = \frac{P_s P_{tr} E[P]}{(1-P_{tr})\sigma + P_{tr} P_s T_s + P_{tr}(1-P_s) T_c} \tag{3.1}$$

where P_s is the probability that a transmission occurring on the channel is successful, P_{tr} is the probability that there is at least one transmission in the considered slot time and σ is the duration of an empty slot time. T_s is the average time the channel is sensed busy. T_c is the average time the channel is sensed busy by each station during a collision. In our study, We use constant packet length in our study, E[P]=P. P_s, P_{tr} and σ remain the same as in the paper [36]. T_s and T_c needs to be modified for our study.

The number of packets that go up and down is the same, so the modified T_s is,

$$\begin{aligned} T_{s_{up}} &= H + P + SIFS + \delta + ACK + DIFS + \delta \\ T_{s_{down}} &= Poll + SIFS + \delta + T_{s_{up}} \\ T_s &= (T_{s_{up}} + T_{s_{down}})/2 \end{aligned} \tag{3.2}$$

Where $H = Header = PHY_{header} + MAC_{header}$, $P = Payload$, δ is the propagation delay.

We only consider the collisions between two packets as an approximation. The collision can happen between two upload packets, two download packets and a upload

and a download packets, each case has 1/3 probability. We obtain the modified T_c,

$$\begin{aligned} T_{c_{uu}} &= T_{c_{up}} = H + P + DIFS + \delta \\ T_{c_{dd}} &= Poll + DIFS + \delta \\ T_c &= (T_{c_{uu}} + T_{c_{ud}} + T_{c_{dd}})/3 \end{aligned} \tag{3.3}$$

Thus, we find the packet delay between stations and an AP under saturation condition.

$$\begin{aligned} D_{up} &= P/R_s \\ D_{down} &= P/R_s + Poll + SIFS + \delta \end{aligned} \tag{3.4}$$

We will analyze the performance of power management based on the packet delay under saturation condition. This means we will obtain the worse-case performance. Our analytical model will focus on the delay in the AP.

3.2 Performance metrics of power management in WLAN

To evaluate the performance of a network, the general metrics include latency, throughput, fairness, and bandwidth utilization. For wireless networks, energy efficiency is also an important aspect to be considered because of the limited capacity of the batteries in the mobile terminals. We use the percentage of time a station stays in the Doze state (PTD) and the response time of a frame (FRT) as our performance metrics.

3.2.1 PTD (Percentage of Time a station stays in the Doze state)

The purpose of this study is to improve the energy efficiency of WLAN. In the IEEE 802.11, a station stays either in the Awake state or in the Doze state. The

Doze state consumes far less energy than the Awake state. Thus, putting a station in the Doze state when it is idle can save energy. We use the percentage of time that a station stays in the Doze state, which is denoted as PTD, to measure the energy efficiency of a power management scheme. The larger the PTD is, the more energy efficiency a power management scheme can achieve. The energy efficiency of a wireless network can be indicated from the average PTD of its stations.

3.2.2 FRT (Frame Response Time)

In the power management of the IEEE 802.11, a station is put into the Doze state when it is idle and wakes up periodically to check if there is buffered traffic for it. The station may not be able to respond to some traffic requests immediately, which brings extra delay into the system. We want to save energy without degrading the system performance. Therefore, we choose FRT as another performance metric for our study. When the system load is fixed, the smaller response time gives the larger throughput. In an infrastructure network, the FRT is defined as the time interval between the instant that a frame arrives at the AP and the instant that the frame is received by the destination station. In an independent network, the FRT is defined as the time interval from the instant the source station attempts to transmit it the first time to the instant the destination station receives the frame.

The goal of this study is to find the power management mechanisms that maximize the PTD in WLAN and satisfy the requirement of the FRT at the same time.

3.3 Sources of energy inefficiency

In WLAN, energy inefficiency is mainly caused by idle listening and collisions.

Idle listening means that a station always stays in the Awake state even when it is idle. Since the Awake state demands far more energy than the Doze state, energy

is wasted when the station is idle and still listens to the network. Thus, reducing the time of the idle listening of a station can save energy. The power management mechanisms that can schedule the time interval of the Awake state and the Doze state of a station to reduce its idle listening and save energy as much as possible is what we are interested in.

Collisions stem from the basic media access method of the IEEE 802.11 standard, DCF, which has the inherent possibility that collisions may happen when two or more stations compete for the media. When collisions happen, the transmitted traffic can not arrive at the destination successfully and the transmission procedure has to be attempted again. This leads to energy wasting. The power management schemes that can reduce the number of collisions or eliminate collisions will improve energy efficiency of WLAN.

We will analyze and improve the power management mechanisms in WLAN based on the identification of the sources of energy inefficiency.

Chapter 4

Analytical Models for Power Management in an Infrastructure BSS

In this chapter, we study the energy efficiency in the infrastructure mode IEEE 802.11 analytically. First, we give the assumptions and notations used in the analytical models. Then, we propose the M/G/1 queue with bulk service model and the D/G/1 queue model for the power management mechanism in an infrastructure BSS. Based on the analytical results from the proposed models, we obtain the average FRT and the lower and upper bounds of the average PTD. Both performance metrics are controlled by the listen interval. Next, we propose to select the largest listen interval under response time requirement to improve energy efficiency in an infrastructure WLAN. Finally, we compare the two models and give a numerical example.

4.1 Assumptions and notations for the models

In an infrastructure BSS, when two stations communicate with each other, the source station sends a frame to the AP first, then the AP forwards the frame to its destination station. Thus, the AP can be modeled as a server and the transmission of frames from the AP to the stations can be modeled as the server serves the traffic.

The buffered traffic in the AP can be viewed as the queue to the server. In the power management scheme, a station goes to the Doze state when idle and wakes up periodically to listen to the TIM messages. The traffic destined for the stations in the power save mode is temporarily buffered in the AP and is transmitted later at designated time.

In order to model the system, we make the following assumptions:

- There are m stations in an infrastructure network and all the stations have the same priority.
- The traffic is generated evenly from all the stations in the BSS.
- The total traffic arrives at the AP with a Poisson process with parameter λ.
- All the traffic buffered in the AP goes to the stations in the BSS and is equally destined for all the stations.
- The packet[1] length is constant and the traffic is transmitted with a single rate, so the transmission time of a packet is constant.
- All the stations are in the power save mode of WLAN.

Here are some notations we are going to use for the model. We denote the beacon interval in the system by B; the listen interval of a station by T, where $T = kB,\ k = 1, 2, 3, \ldots$; the service time of a packet over the media by S. The service time of a packet includes the transmission time of the poll message from the station to the AP and the transmission time of the actual traffic from the AP to the station. The propagation time can be ignored here since the radio communication range is short in a WLAN. We calculate the service time S by the packet delay between stations

[1]For convenience, the term "packet" has the same meaning as the term "frame" in the following.

and an AP under saturation condition, which is derived in chapter 3,

$$\begin{aligned} D_{up} &= P/R_s \\ D_{down} &= P/R_s + Poll + SIFS + \sigma \end{aligned} \tag{4.1}$$

Here, the P in D_{up} is the length of the poll frame, the P in the D_{down} is the length of message frame. We denote the maximum number of packets that can be transmitted by the AP during a beacon interval by L, where $L = \lfloor \frac{B}{E[S]} \rfloor$.

4.2 M/G/1 queue with bulk service model

Here, we present the M/G/1 queue with bulk service model. We propose our analytical model according to the length of a listen interval. We call the model a single-beacon model when $T = B$, and call the model a multiple-beacon model when $T = kB, where\ k = 2, 3, \ldots$. We propose the single-beacon model first, then extend it to the multiple-beacon model. Also, we observe that the single-beacon model is a special case of the multiple-beacon model.

4.2.1 Single-beacon model ($T = B$)

We model the system as an M/G/1 queue with bulk service [39] [40], where the arrival to the AP is Poisson process with parameter λ and the service time of a packet is S. Also, the service is only available at the beginning of each listen interval, that is, the interval between service instants is constant with T. In any service instants, the maximum number of packets that can be served is L.

We obtain an embedded Markov chain by looking at the number of packets in the server at the beginning of each listen interval. Then, based on the solution of the Markov chain, we can calculate the average number of packets in the system.

Embedded Markov chain

Let X_n be the number of packets in the server at the beginning of the nth listen interval, A_n and D_n be the number of arrivals and departures during the nth listen interval respectively. Fig.4.1 depicts the relation of X_n, X_{n-1}, A_{n-1} and D_{n-1}.

The maximum number of packets that can be served within a listen interval is L. If X_{n-1} is less than or equal to L, X_n is exactly equal to A_{n-1}. Otherwise, X_n is the summation of A_{n-1} and X_{n-1} reduced by L.

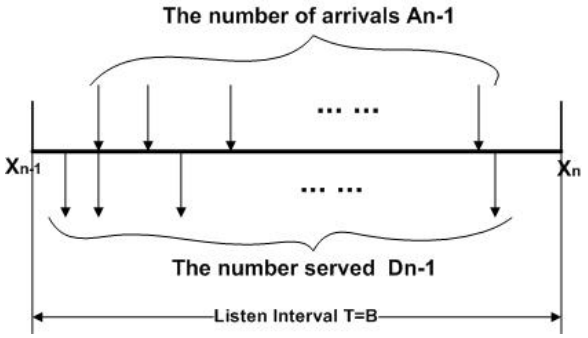

Figure 4.1: The relation of arrivals, departures during a listen interval and the number of packets at the beginning of a listen interval.

Therefore,

$$\begin{aligned} X_n &= X_{n-1} - D_{n-1} + A_{n-1} = (X_{n-1} - L)^+ + A_{n-1} \\ &= \begin{cases} X_{n-1} + A_{n-1} - L & X_{n-1} \geq L \\ A_{n-1} & X_{n-1} < L \end{cases} \end{aligned} \tag{4.2}$$

where, D_{n-1} is the number of packets actually served during the $(n-1)th$ listen interval.

Here, L is a constant. X_n only depends on X_{n-1} and A_{n-1}. The variable A_{n-1} is an independent, identically distributed random variable because the arrival process is Poisson. X_n only depends on X_{n-1}. Therefore, $\{X_n, n = 0, 1, 2, 3...\}$ is a Markov chain. Its state space is $\{0, 1, 2, 3...\}$.

The transition probability $p_{ij} = P\{X_n = j | X_{n-1} = i\}$ of this Markov chain can be represented by the distribution of the number of arrivals in a listen interval. The

probability that there are k arrivals in a listen interval is

$$a_k = P(A_n = k) = e^{-\lambda T}\frac{(\lambda T)^k}{k!} \qquad k = 0, 1, 2, \ldots \tag{4.3}$$

Based on (4.2) (4.3), we find the transition probability p_{ij}.

When $i \geq L$,

$$\begin{aligned} p_{ij} &= P\{X_n = j | X_{n-1} = i\} \\ &= P\{X_{n-1} + A_n - L = j \mid X_{n-1} = i\} \\ &= P\{A_n = j + L - i\} \\ &= \begin{cases} a_{j+L-i} & j \geq i - L \\ 0 & j < i - L \end{cases} \end{aligned} \tag{4.4}$$

When $i < L$,

$$\begin{aligned} p_{ij} &= P\{X_n = j \mid X_{n-1} = i\} \\ &= P\{A_n = j \mid X_{n-1} - i\} \\ &= P\{A_n = j\} = a_j \end{aligned} \tag{4.5}$$

Therefore, the transition matrix of this Markov chain is:

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{L-1} & a_L & a_{L+1} & \cdots \\ a_0 & a_1 & a_2 & \cdots & a_{L-1} & a_L & a_{L+1} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ a_0 & a_1 & a_2 & \cdots & a_{L-1} & a_L & a_{L+1} & \cdots \\ a_0 & a_1 & a_2 & \cdots & a_{L-1} & a_L & a_{L+1} & \cdots \\ 0 & a_0 & a_1 & \cdots & a_{L-2} & a_{L-1} & a_L & \cdots \\ 0 & 0 & a_0 & \cdots & a_{L-3} & a_{L-2} & a_{L-1} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Where, the first $(L + 1)$ rows are the same. It means $p_{0j} = p_{1j} = \ldots = p_{Lj} = a_j$, *where* $j = 0, 1, 2, \ldots$.

Solution of Markov chain

By solving the Markov chain [41], we can get the steady state probability $(\pi_0, \pi_1, ...\pi_{L-1})$, where $\pi_i = P\{X = i\}, i = 0, 1, 2, \ldots$,(L-1), and X is the number of packets in the server at the beginning of a listen interval when the system is in the steady state.

We apply the steady state equation of a Markov chain, $\boldsymbol{\pi} = \boldsymbol{\pi P}$, to the transition matrix and obtain,

$$\pi_n = a_n \sum_{k=0}^{L-1} \pi_k + \sum_{k=L}^{L+n} \pi_k a_{n+L-k} \tag{4.6}$$

We find the Z-transform of the steady state probability $\pi(z) = \sum_{n=0}^{\infty} \pi_n z^n$.

$$\begin{aligned}
\sum_{n=0}^{\infty} \pi_n z^n &= \sum_{n=0}^{\infty} a_n \sum_{k=0}^{L-1} \pi_k z^n + \sum_{n=0}^{\infty} \sum_{k=L}^{L+n} \pi_k a_{n+L-k} z^n \\
\pi(z) &= A(z) \sum_{k=0}^{L-1} \pi_k + \sum_{k=L}^{\infty} \sum_{n=k-L}^{\infty} \pi_k a_{n+L-k} z^{n+L-k} \\
\pi(z) &= A(z) \sum_{k=0}^{L-1} \pi_k + \frac{1}{z^L} \sum_{k=L}^{\infty} \pi_k z^k \sum_{n=k-L}^{\infty} a_{n+L-k} z^{n+L-k} \\
\pi(z) &= A(z) \sum_{k=0}^{L-1} \pi_k + \frac{1}{z^L} [\pi(z) - \sum_{k=0}^{L-1} \pi_k z^k] A(z) \\
z^L \pi(z) &= z^L A(z) \sum_{k=0}^{L-1} \pi_k + \pi(z) A(z) - \sum_{k=0}^{L-1} \pi_k z^k A(z) \\
\pi(z) &= \frac{A(z)[z^L \sum_{k=0}^{L-1} \pi_k - \sum_{k=0}^{L-1} \pi_k z^k]}{z^L - A(z)} \\
\pi(z) &= \frac{\sum_{k=0}^{L-1} \pi_k (z^L - z^k)}{z^L A(z)^{-1} - 1}
\end{aligned} \tag{4.7}$$

Here, the z-transform of the arrival process is:

$$\begin{aligned}
A(z) &= \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} e^{-\lambda T} \frac{(\lambda T)^n}{n!} z^n \\
&= e^{-\lambda T} \sum_{n=0}^{\infty} \frac{(\lambda T z)^n}{n!} = e^{-\lambda T} e^{\lambda T z} = e^{-\lambda T (1-z)}
\end{aligned} \tag{4.8}$$

By analytic property [42], we can find $\pi_k,\ k = 0, 1, \ldots, L-1$ by solving the following two equations.

$$\begin{cases} \sum_{k=0}^{L-1} \pi_k (z^L - z^k) = 0 \\ z^L - e^{-\lambda T(1-z)} = 0 \end{cases} \tag{4.9}$$

Where, the second equation comes from the equation $\frac{z^L}{e^{-\lambda T(1-z)}} - 1 = 0$.

Here, we need to solve the second nonlinear equation first, then plug the roots of z to the first equation to form an equation matrix and find out $(\pi_0, \pi_1, \ldots, \pi_{L-1})$. The process is shown as following.

(a). Before solving the second equation, we prove that for the equation $z^L - e^{-\lambda T(1-z)} = 0$, there are exactly L roots in the range of $|z| \leq 1$.

Proof. Write the equation $z^L - e^{-\lambda T(1-z)} = 0$ as $e^{-\lambda T}(e^{\lambda T} z^L - e^{\lambda T z}) = 0$

Let $a = e^{\lambda T}$, we have, $az^L - a^z = 0$

Split this function into two parts, $f(z) + g(z)$, where $f(z) = az^L, \quad g(z) = -a^z$.

Let ξ be any real number, $0 < \xi < 1$.

Define the closed contour C as a circle about the origin, with the radius $(1 + \xi)$. So, on the contour C, $|z| = 1 + \xi$.

Then, $|f(z)| = |az^L| = a(1+\xi)^L, \quad |g(z)| = |-a^z| = a^{1+\xi}$

since, $log_a[a(1+\xi)^L] = 1 + Llog_a(1+\xi) = 1 + \frac{L}{\lambda T} ln(1+\xi)$,

$log_a[a^{1+\xi}] = 1 + \xi$,

Also, $\lim_{\xi \to \infty} \frac{\frac{L}{\lambda T} ln(1+\xi)}{\xi} = \lim_{\xi \to 0} \frac{\frac{L}{\lambda T} \frac{1}{1+\xi}}{1} = \frac{L}{\lambda T} > 1$

So, we have, $a(1+\xi)^L > 1, \quad a^{1+\xi} > 1$,

That is, $\quad log_a|f(z)| > log_a|g(z)|$

Finally, we get, $|f(z)| > |g(z)|$ on the closed contour C.

By Rouché's theorem [43] [39], there are exactly L roots in the range $|z| \leq 1$. $\square$

(b). Solve the nonlinear equation $z^L - e^{-\lambda T(1-z)} = 0$.

Many research have been done regarding this problem. Here, we use Matlab to get the solution.

Our method consists of two steps:

Step 1. Expand the item $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ and take the first $(L+1)$ items. Solve the polynomial using the Matlab command "roots".

Step 2. Use numerical method to increase the accuracy of the roots. Use the Matlab command "$fsolve$" to adjust the value of z.

$$z = fsolve(@function, x_0)$$

Where, the function is $f(x) = x \wedge L - exp(-\lambda * T + \lambda * T * x)$, and the x_0 is the start value.

Use the approximate value of the variable z got in the step 1 as the start value x_0 in the step 2.

(c). Find the steady state solution.

Plug in the solution of the second equation $(z_0, z_1, z_2, ..., z_{L-1})$ to the first equation $\sum_{k=0}^{L-1} \pi_k (z^L - z^k) = 0$. We obtain an equation matrix consisting of L equations.

For the special case $z_0 = 1$, we compute the equation as following:

$$\begin{aligned}
\pi(z)|_{z=1} &= \lim_{z \to 1} \pi(z) \\
&= \lim_{z \to 1} \frac{A(z)[z^L \sum_{k=0}^{L-1} \pi_k - \sum_{k=0}^{L-1} \pi_k z^k]}{z^L - A(z)} \\
&= \lim_{z \to 1} \frac{A'(z)[z^L \sum_{k=0}^{L-1} \pi_k - \sum_{k=0}^{L-1} \pi_k z^k] + A(z)[Lz^{l-1} \sum_{k=0}^{L-1} \pi_k - \sum_{k=0}^{L-1} \pi_k k z^{k-1}]}{Lz^{L-1} - A'(z)} \\
&= \frac{A(1)[L \sum_{k=0}^{L-1} \pi_k - \sum_{k=0}^{L-1} k\pi_k]}{L - A'(1)} = 1
\end{aligned} \tag{4.10}$$

Here,

$$A(1) = A(z)|_{z=1} = e^{-\lambda T(1-z)}|_{z=1} = 1$$

also,

$$A'(1) = A'(z)_{z=1} = e^{-\lambda T(1-z)}(-\lambda T)(-1)|_{z=1} = \lambda T \tag{4.11}$$

so, for the special case $z_0 = 1$, we get the equation

$$\sum_{k=0}^{L-1} \pi_k - \sum_{k=0}^{L-1} k\pi_k = \frac{L - A'(1)}{A(1)} = L - A'(1) = L - \lambda T \tag{4.12}$$

Then, we solve the equation matrix which consists of L equations related to $\pi_k (k = 0, 1, 2, \ldots)$ and obtain the steady state probability $(\pi_0, \pi_1, ...\pi_{L-1})$.

It is noticed that, in terms of the steady state probability of the number of packets in the server at the beginning of a listen interval, this model is equivalent to the $M/D/L$ system. However, in terms of the response time of a packet in the system, they are different because the $M/D/L$ model serves the packets in parallel while the model $M/D/1$ with bulk service model serves the packets in serial.

Average number of packets

We denote the number of packets left behind by a departure by Y. $E[Y]$ is the average number of packets in the system because of the Poisson arrival process and the PASTA (Poison Arrivals See Time Average) property [44]. $E[Y]$ can be calculated based on $E[X]$.

To find $E[X]$, we define $\Delta_{L,s} = \begin{cases} L & s \geq L \\ s & s \leq L \end{cases}$. Then, by plugging $\Delta_{L,s}$ into (4.2), we can rewrite (4.2) as

$$X_n = X_{n-1} - \Delta_{L,X_{n-1}} + A_{n-1} \tag{4.13}$$

We square on both sides of the equation (4.13) and then take expectation. Let $n \to \infty$, define $\widetilde{X} = \lim_{n\to\infty} X_n$. Where, $\widetilde{X}$ stands for the number of packets at the beginning of a listen interval in the system steady state.

We obtain,

$$\begin{aligned} E[\widetilde{X}^2] = E[\widetilde{X}^2] + E[\Delta^2_{L,\widetilde{X}}] + E[\widetilde{A}^2] - 2E[\widetilde{X}\Delta_{L,\widetilde{X}}] \\ + 2E[\widetilde{X}]E[\widetilde{A}] - 2E[\Delta_{L,\widetilde{X}}]E[\widetilde{A}] \end{aligned} \tag{4.14}$$

Where, $\widetilde{X}$ and $\widetilde{A}$ are independent, also, $\Delta_{L,\widetilde{X}}$ and $\widetilde{A}$ are independent.

We also know,

$$E[\Delta^2_{L,\widetilde{X}}] = \sum_{i=0}^{L-1} i^2\pi_i + \sum_{i=L}^{\infty} L^2\pi_i = \sum_{i=0}^{L-1} i^2\pi_i + L^2(1 - \sum_{i=0}^{L-1}\pi_i)$$

$$E[\widetilde{X}\Delta_{L,\widetilde{X}}] = \sum_{i=0}^{L-1} i^2\pi_i + \sum_{i=L}^{\infty} iL\pi_i = LE[\widetilde{X}] + \sum_{i=0}^{L-1} i^2\pi_i - L\sum_{i=0}^{L-1} i\pi_i$$

$$E[A^2] = \lambda T + (\lambda T)^2 \qquad E[\Delta_{L,\widetilde{X}}] = E[\widetilde{A}] = \lambda T$$

Therefore, we obtain the average number of packets in the server at the beginning of a listen interval.

$$E[X] = \frac{\lambda T - (\lambda T)^2 + L^2 - \sum_{i=0}^{L-1}(i-L)^2\pi_i}{2(L - \lambda T)} \tag{4.15}$$

Then, we calculate $E[Y]$ by conditioning on the number of departures in a listen interval. X is the number of packets at the beginning of a listen interval. So the number of packets left behind by the ith departure is $(X - i)$ plus the number of arrivals during the serving time of the first i departed packets. The average number of packets left behind by the ith departure is:

$$\sum_{j=0}^{\infty}(X - i + j)e^{-\lambda iE[S]}\frac{(\lambda iE[S])^j}{j!} = X - i + \lambda iE(S)$$

where $E(S)$ is the average service time of a packet.

In a listen interval, there are at most L departures. When $X \geq L$, there are exactly L departures; When $X = i$, where $i < L$, there are i departures. The average number of packets left behind by a departure in a listen interval for each possible

value of X is:

$When\ X \geq L,$

$$\begin{aligned} & E[Y|X \geq L] \\ =&\frac{1}{L}[\ (X-1+\lambda E[S]) + (X-2+2\lambda E[S]) + \ldots + (X-L+L\lambda E[S])] \\ =&X + \frac{L+1}{2}(\lambda E[S]-1) \end{aligned}$$

$When\ X = L-1,$

$$\begin{aligned} & E[Y|X = L-1] \\ =&\frac{1}{L-1}[\ (X-1+\lambda E[S]) + (X-2+2\lambda E[S]) + \ \ldots\ + (X-(L-1)+(L-1)\lambda E[S])\ ] \\ =&X + \frac{L}{2}(\lambda E[S]-1) \end{aligned}$$

and so on.

Also, the average number of departures in a listen interval is:

$$\begin{aligned} N &= \sum_{i=1}^{L-1} i\pi_i + L\sum_{i=L}^{\infty}\pi_i = \sum_{i=1}^{L-1} i\pi_i + L(1-\sum_{i=0}^{L-1}\pi_i) \\ &= L - \sum_{i=0}^{L-1}(L-i)\pi_i \end{aligned} \tag{4.16}$$

Thus, the fraction of departures for different X in the system is:

$$\begin{aligned} &\frac{LP\{X \geq L\}}{N} = \frac{L(1-\sum_{i=0}^{L-1}\pi_i)}{N},\quad when\ X \geq L, \\ &\frac{(L-1)P\{X = L-1\}}{N} = \frac{(L-1)\pi_{L-1}}{N},\quad when\ X = L-1, \end{aligned}$$

and so on.

Therefore, we can find $E[Y]$ through conditioning on the number of departures in a listen interval.

$$
\begin{aligned}
E[Y] &= E[Y|X \geq L]\frac{LP(X \geq L)}{N} + E[Y|X = L-1]\frac{(L-1)\pi_{L-1}}{N} \\
&\quad + \ldots + E[Y|X = 2]\frac{2\pi_2}{N} + E[Y|X = 1]\frac{\pi_1}{N} \\
&= \frac{1}{N}\Big\{L(E[X] - \sum_{i=0}^{L-1} i\pi_i) + \sum_{i=1}^{L-1} i^2\pi_i + \frac{\lambda E[S]-1}{2} \\
&* [L(L+1)(1 - \sum_{i=0}^{L-1}\pi_i) + (L-1)L\pi_{L-1} + \ldots + 2\pi_1]\Big\} \\
&= \frac{1}{N}\Big\{\frac{\lambda E[S]+1}{2}\sum_{i=0}^{L-1} i^2\pi_i + \frac{\lambda E[S]-1-2L}{2}\sum_{i=0}^{L-1} i\pi_i \\
&- \frac{\lambda E[S]-1}{2}L(L+1)\sum_{i=0}^{L-1}\pi_i + \frac{\lambda E[S]-1}{2}L(L+1) + LE[X]\Big\}
\end{aligned}
\tag{4.17}
$$

4.2.2 Multiple-beacon model ($T = KB$)

We look at the system when a listen interval equals multiple beacon intervals, i.e., $T = kB$, where $k > 1$. We assume the listen intervals of the stations are configured such that at the beginning of each beacon interval, the number of stations that wake up is approximately the same. In other words, if the number of the stations in the system is m, the number of the stations that wake up at the beginning of each beacon interval is m/k.

Referring to the single-beacon analytical model, where all the stations wake up for each beacon, we only need to do several modifications to model the system when the listen interval equals multiple beacon intervals. The system can still be modeled as an $M/G/1$ queue with bulk service, while the arrival rate becomes λ/k and the interval between service instants is $T = kB$.

In Fig.4.2, we let X'_n denote the number of the packets eligible for transmission at the beginning of the nth beacon interval. The eligibility packets are the packets whose destination stations wake up for this beacon. Let A'_{n-1} stands for all the packets that arrive with the rate λ/k during the $(n-1)th$ listen interval kB. D'_{n-1} is still the actual number of packets served in the $(n-1)th$ beacon interval. Also, L is the

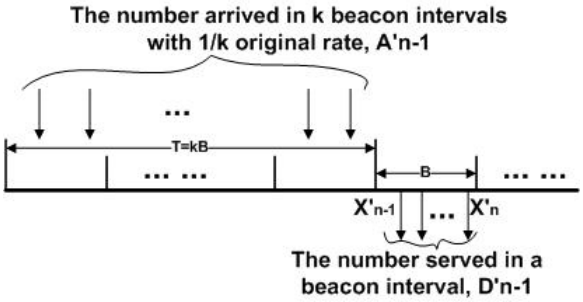

Figure 4.2: The relation of arrivals during a listen interval, departures during a beacon interval and the number of packets eligible for transmission at the beginning of a beacon interval.

maximum number of packets that can be served in a beacon interval. So, we obtain the same Markov chain as that from (4.2), but with different physical meaning.

Following the method for the single-beacon model, we can compute $E[X']$, the average number of packets eligible for transmission at the beginning of a beacon interval.

The difference between the single-beacon model and the multiple-beacon model is that all the packets are eligible for transmission at the beginning of each beacon interval in the single-beacon model, but only part of the packets are in the multiple-beacon model. In the multiple-beacon model, the average number of packets buffered and not eligible for transmission at the beginning of a beacon interval is:

$$\frac{\lambda}{k}B + \frac{\lambda}{k}2B + \ldots + \frac{\lambda}{k}(k-1)B = \frac{k-1}{2}\lambda B \quad k = 2, 3, \ldots \tag{4.18}$$

Therefore, the average number of packets in the system of the multiple-beacon model is the summation of $E[Y]$, the average number of packets in the system of the single-beacon model, and $\frac{k-1}{2}\lambda B$. We denote the average number of packets in the system in the multiple-beacon model by $E[Y_m]$.

$$E[Y_m] = E[Y] + \frac{k-1}{2}\lambda B, \quad k = 1, 2, 3, \ldots \tag{4.19}$$

In fact, the single-beacon model is a special case of the multiple-beacon model

when $k = 1$. In the single-beacon model, all stations wake up for every beacon and the number of packets not eligible for transmission is $\frac{k-1}{2}\lambda B|_{k=1} = 0$. In the following discussion, we use the multiple-beacon model as the general model for any length of listen interval.

4.2.3 Average FRT and average PTD

We calculate the performance metrics: the average FRT and the bounds of the PTD here.

FRT

Based on $E[Y_m]$, we obtain the average FRT by Little's formula [44].

$$E[FRT] = \frac{1}{\lambda}E[Y_m] = \frac{1}{\lambda}E[Y] + \frac{k-1}{2}B, \quad k = 1, 2, 3, \ldots \tag{4.20}$$

PTD

The average PTD depends on the order of the packets retrieved, which is controlled by the DCF [12]. Fig.4.3 shows an example of the retrieving order of packets. Three stations (station 1, station 2 and station 3) are retrieving packets from the AP. Assume each station has three packets buffered in the AP. According to the retrieving order in Fig.4.3, the station 3 finishes retrieving all its packets at the time instant B and enters the Doze state. The station 2 and station 1 enters the Doze state at the time instant C and D respectively. Because the retrieving sequence can change from beacon interval to beacon interval, we can not obtain a fixed value as the average PTD.

Instead, we develop a lower bound and a upper bound for the average PTD. The lower bound is the average idle time of the AP. In Fig.4.4, all stations are in the Doze state during the time interval DE, which is the idle time of the AP. We take the average idle time of the AP as the lower bound of the average PTD.

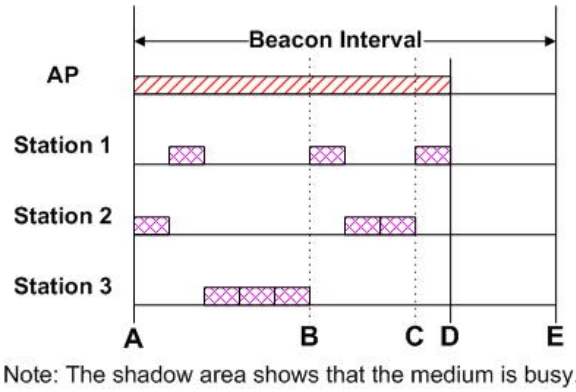

Figure 4.3: Stations retrieve packets from the AP under DCF.

$$E[PTD]_{lower\ bound} = \frac{kB - \frac{\lambda}{k}kBE[S]}{kB} = 1 - \frac{\lambda E[S]}{k} \tag{4.21}$$

where $k = 1, 2, 3, \ldots$.

The upper bound is developed by assuming the buffered traffic is transmitted exclusively station by station. It means that one station reserves the medium and continuously retrieves the traffic until all its buffered packets are received. Then, the station enters the Doze state. The other station repeats the process. From Fig.4.3 and Fig.4.4, we observe the time interval B'D is larger than BD and the time interval C'D is larger than CD. Therefore, the upper bound of the average PTD is,

$$\begin{aligned} E[PTD]_{upper\ bound} &= \frac{kB - \frac{\lambda}{k}kBE[S] + \frac{\frac{\lambda BE[S]}{m/k}\frac{\frac{m}{k}(\frac{m}{k}-1)}{2}}{m/k}}{kB} \\ &= 1 - \frac{\lambda E[S]}{2k} - \frac{\lambda E[S]}{2m} \end{aligned} \tag{4.22}$$

Where, $k = 1, 2, \ldots$; m is the number of stations in the system.

From Fig.4.3 and Fig.4.4, we also notice that, during the time period A to D, different algorithms can be used to schedule the transmission sequence from the AP to the stations. A good algorithm can reduce the total waiting time of the stations and save energy. We will propose an enhanced power management scheme later and supports our upper bound derivation.

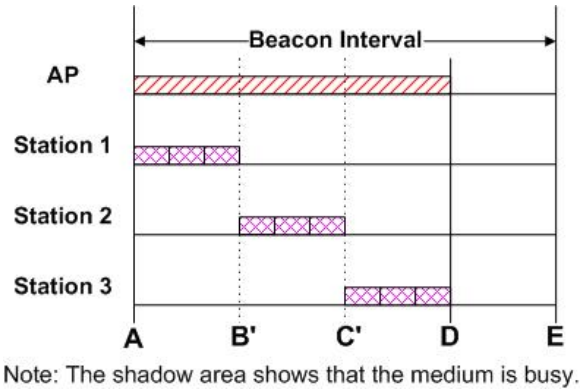

Figure 4.4: Stations retrieve packets from the AP station by station.

4.3 D/G/1 queue model

We approach the system from another perspective and model it as a D/G/1 model. The same, we start from the single-beacon model and extended it to the multiple-beacon model. Also, we observe that the single-beacon model is a special case of the multiple-beacon model.

4.3.1 Single-beacon model ($T = B$)

We model the system as a $D/G/1$ queue.

We view all the traffic arrived within one listen interval as a batch. Our model is developed based on the behavior of the batches.

The earliest time to serve the batch is at the beginning of the next listen interval. So the beginning of the next listen interval can be treated as the arrival instant of a batch. Therefore, batches arrive with the constant interval B.

The service time of a batch is the summation of the service time of all the packets arrived within a listen interval. This is where the "G" in the model comes from.

The "1" in the model stands for that only one batch is permitted to transmit each time.

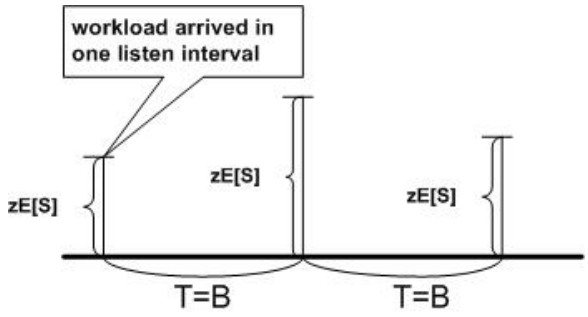

Figure 4.5: D/G/1 Queue Model with single-beacon.

Fig.4.5 illustrates the $D/G/1$ queue model. The batches arrive at the constant interval B. A batch consists of all the traffic arrived within one listen interval. We denote the number of packets arrived in one listen interval by z, which is Poisson distributed with parameter λB. The service time of a batch is $zE[S]$. Also, one batch is transmitted each time.

4.3.2 Multiple-beacon model ($T = kB$)

We assume the listen intervals of the stations are configured such that at the beginning of each beacon interval, the number of stations that wake up is approximately the same. For example, there are 10 stations in a BSS and the listen interval of the stations is 2 beacon intervals. At the beginning of each beacon interval, 5 stations wake up.

We still model the system as a $D/G/1$ queue. A batch arrives at the beginning of each beacon interval. In the single-beacon model, the packets in a batch arrive in a beacon interval with rate λ; while in the multiple-beacon model, the packets in a batch arrive in k beacon intervals with rate $\frac{\lambda}{k}$. The number of packets arrived in a listen interval is Poisson distributed with parameter $\frac{\lambda}{k} * kB = \lambda B, \quad k = 1, 2, 3, \ldots$, which is independent of the length of the listen interval. Therefore, we obtain the same D/G/1 model.

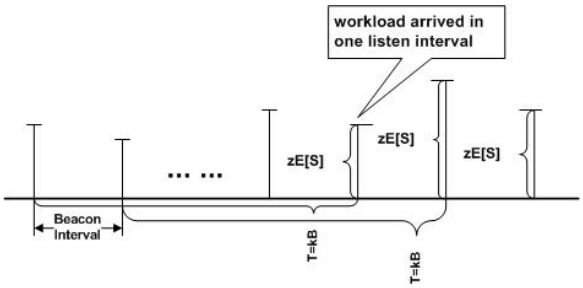

Figure 4.6: D/G/1 Queue Model with multiple-beacon.

The difference between the single-beacon model and the multiple-beacon model is the time interval between the arrival instant of a packet and the time instant that the packet is eligible for retrieving. In the single-beacon model, this waiting time is less than or equal to one beacon interval; in the multiple-beacon interval, this waiting time ranges from 0 to k beacon intervals. The single-beacon model is a special case of the multiple-beacon interval.

Fig.4.6 shows the multiple-beacon model. The batches with size z arrive with the constant interval B. Because the traffic (from all the stations) arrive at the AP with rate λ and $\frac{1}{k}$ stations wake up at the beginning of each beacon interval, the traffic that are eligible for transmission (the destination station wake up) arrive at the AP with the rate $\frac{\lambda}{k}$. The packets arrive with the rate $\frac{\lambda}{k}$ in k beacon intervals forms one batch and are eligible for transmission in the beginning of the next listen interval (the $(k+1)th$ beacon interval).

4.3.3 Average FRT and average PTD

We calculate the average FRT and derive the bounds of the average PTD in the following.

FRT

We divide the total waiting time of a packet into three phases.

W_1: The time between the arrival of a packet in the original system and the arrival of the batch in the $D/G/1$ queue model. In other words, the time interval between the arrival instant of the packet and the beginning of the next listen interval in the original system.

W_2: The waiting time experienced by a batch in the $D/G/1$ queue. Since a packet is in a batch, it also experiences this part of waiting time.

W_3: The waiting time caused by other packets in the same batch.

We calculate the average waiting time of a packet in each phase.

Calculation of the average waiting time W_1

Since the packet arrival is Poisson process, the arrival instant of each packet is uniformly distributed in the beacon interval $(0, kB)$. Thus, the expected waiting time of a packet is $\frac{kB}{2}$.

Calculation of the average waiting time $W2$

The average waiting time of a $D/G/1$ queue with no vacation was deduced by Servi in [45]. It is,

$$-\frac{N(N-1)-G''(1)}{2(N-G'(1))}+\sum_{r=1}^{N-1}\frac{1}{1-z_r}$$

where, N is the quantized length of the interarrival time of the customers to the queue; $G(z)$ is the z-transform of the service time; $z_1, z_2, \ \ldots, z_{N-1}$ are the unique roots of $z^N - G(z) = 0$ that are on or within the unit circle but not equal to 1.

Here, we quantize the beacon interval by the service time of a packet E[S]. So, the interarrival time of the batches in our model is L.

In the quantized time frame, the service time of a packet is $E[S]/E[S] = 1$. The service time of a batch is the summation of the service time of all the packets in the batch. The number of packets in a batch is the number of packets arrived in a listen interval, which is Poisson distributed with parameter λB. We calculate the z-transform of the batch service time.

$$G(z) = \sum_{i=0}^{\infty} z^i e^{-\lambda B} \frac{(\lambda B)^i}{i!} = e^{-\lambda B(1-z)}$$

$$G'(1) = G'(z)|_{z=1} = \lambda B$$

$$G''(1) = G''(z)|_{z=1} = (\lambda B)^2$$

The roots of $z^L - G(z) = 0$ inside the unit circle can be computed by the same approach used for solving the Markov chain in the previous section.

Therefore, based on the known variables, the average waiting time W_2 is,

$$E[W_2] = [-\frac{L(L-1) - G''(1)}{2(L - G'(1))} + \sum_{r=1}^{L-1} \frac{1}{1 - z_r}] * E[S] \tag{4.23}$$

Calculation of the average waiting time W_3

We've denoted the number of arrivals within a listen interval by A'. Because of the Poisson arrival process, the distribution of A' is:

$$P(A' = i) = e^{-\lambda B} \frac{(\lambda B)^i}{i!} \quad i = 1, 2, \ldots.$$

A batch includes all the packets arrived in a listen interval. If the batch size is i, the average waiting time caused by other packets within the batch is:

$$\frac{0 + 1 + 2 + \ \ldots \ + (i-1)}{i} E[S] = \frac{i-1}{2} E[S], \quad i = 1, 2, \ldots.$$

The percentage of the packets whose average waiting time equals to $\frac{i-1}{2} E[S]$ is:

$$\frac{iP(A' = i)}{\sum_{i=1}^{\infty} iP(A' = i)}$$

So, the average waiting time W_3 is:

$$\begin{aligned} E[W_3] &= \sum_{i=1}^{\infty} \Big(\frac{(i-1)}{2} E[S] * \frac{iP(A'=i)}{\sum_{i=1}^{\infty} iP(A'=i)} \Big) \\ &= \sum_{i=1}^{\infty} \Big(\frac{i-1}{2} E[S] * \frac{iP(A'=i)}{\lambda B} \Big) \\ &= \frac{E[S]}{2\lambda B} \sum_{i=1}^{\infty} (i^2 - i) P(A'=i) \\ &= \frac{(\lambda B)^2 + \lambda B - \lambda B}{2\lambda B} E[S] \\ &= \frac{\lambda B}{2} E[S] \end{aligned}$$

Therefore,

$$\begin{aligned} E[FRT] &= E[W_1] + E[W_2] + E[W_3] + E[S] \\ &= \frac{kB}{2} + E[W_2] + E[W_3] + E[S] \qquad (4.24) \\ &where, k = 1, 2, 3, \ldots . \end{aligned}$$

PTD

For the D/G/1 queue model, we use the same lower and upper bounds of the average PTD developed for the M/G/1 queue with bulk service model.

4.4 Suggestion for the selection of listen interval

In (4.20) and (4.24) , E[FRT] increases linearly with the increase of the variable k. In (4.21) and (4.22), the lower and upper bounds of the PTD increases with the increase of k as well. In other words, when the listen interval increases, the energy efficiency gets better, but the response time becomes poorer. Therefore, we suggest selecting the largest listen interval under the response time requirement.

In fact, it is quite intuitive that the largest listen interval with the satisfaction of the response time should be selected. However, to the best of our knowledge, no

analytical model supports this intuitive result. We propose two analytical models for the original power management scheme in the IEEE 802.11 and claim the result based on our analysis. Furthermore, given an infrastructure BSS, we can calculate the average FRT and the range of the average PTD analytically.

4.5 Comparison of two models

We compare the two models in the following. We call the M/G/1 with bulk service model as the A model and call the D/G/1 queue model as the B model.

4.5.1 Intuitiveness and computational complexity

In terms of the intuitiveness, the A model is easier to understand than the B model. In terms of computational complexity, the B model is simpler.

4.5.2 Accuracy of modeling

Both models have advantage and disadvantage for some special case of the original system.

Beacon shifting

In the IEEE 802.11 protocol, if the remaining time in the current listen interval is not big enough for one packet to finish its transmission, the AP still transmits the packet and lets its beacon announcement wait. Once the media becomes idle, AP makes its beacon announcement immediately. The next beacon announcement will be made as scheduled. We call this case as "Beacon shifting".

Fig.4.7 illustrates a case where the D/G/1 model works better, where A stands for the operation of the $M/G/1$ with bulk service model, and B stands for the operation of the $D/G/1$ model. A postpones the transmission of the packet to the beginning of the next beacon interval; while B finishes the transmission first and shifts the beacon

announcement to the end of the transmission. Obviously, B models the real system more accurately.

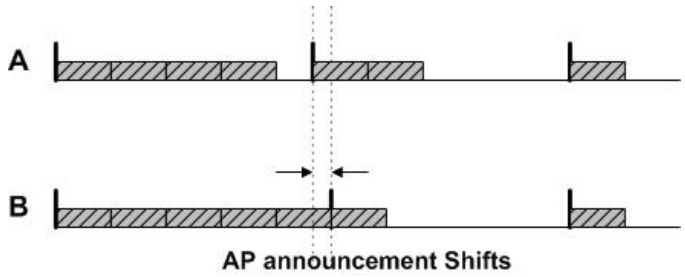

Figure 4.7: Beacon shifting.

Service order of the packets

If the announced packets can not be completely transmitted in the current beacon interval, the remaining packets will be announced again together with newly arrived packets. Under the IEEE 802.11 DCF mechanism, the remaining packets and the newly arrived packets have the same priority to compete for the medium to transmit. In this point, our A model is consistent with the original system. In our B model, the remaining packets from previous beacon interval are transmitted before the newly arrived packets because the model serve the traffic batch by batch. Thus, in terms of the transmission of the remaining packets from previous beacon interval, the A model is closer to the original system.

Although each model cannot perfectly match the original system, the analysis results of both models are very close. Both models are good for the original system. The computational example given below shows this claim.

4.5.3 A computational example

We give a numerical example to illustrate our analysis. We apply the parameters used in the simulation in order to compare the analytical results and the simulation results.

The beacon interval is $B = 100\ msec$ and the system load is 50%. The interarrival time of packets is 6 $msec$. The service time of a packet is 3 $msec$.

M/G/1 queue with bulk service

First, we compute the 33 roots of $z^L - G(z) = 0$ on and within the unit circle and the steady state probability $(\pi_0, \pi_1, \ldots, \pi_{32})$. Then, we obtain $E[X] = 16.6703,\ N = 16.67,\ E[Y] = 13.0508(msec)$. Therefore, $E[FRT] = \frac{E[Y]}{\lambda} + \frac{k-1}{2}B = 78.3045 + 50(k-1) = 50k + 28.3045(msec),\quad k = 1, 2, 3\ldots$

D/G/1 queue

We compute $E[W_1] = \frac{kB}{2} = 50k(msec)$, $E[W_3] = \frac{\lambda B}{2}E[S] = 25(msec)$. We obtain $L = \lfloor \frac{B}{E[S]} \rfloor = 33$, $G'(1) = \lambda B = \frac{100}{6}$, $G''(1) = (\lambda B)^2 = (\frac{100}{6})^2$. Also, there are 32 roots of $z^L - G(z) = 0$ inside the unit circle and $\sum_{r=1}^{L-1} \frac{1}{1-z_r} = 23.8249$. So, $E[W_2] = 0.0053$. By (4.24), the average FRT is $50k + 28.0053(msec),\quad k = 1, 2, 3\ldots$.

The analytical results from two models are very close.

Chapter 5

Simulation of Power Management in WLANs

This chapter studies the energy efficiency in WLAN by simulation. The simulation is an important method to study the performance of systems and to verify analytical models, which can provide general guideline for system design and configuration.

First, we introduce the simulation environment and the basic configuration. Then, we simulate the power management schemes and find the optimal power management parameters for energy efficiency. Finally, we propose the enhanced power management schemes for an infrastructure network and an independent network respectively.

5.1 Simulation environment and configuration

The simulation is carried out in the Arena platform [46]. On one hand, the Arena platform satisfies our simulation requirement. Because our study only focuses on the power management scheme in WLAN, the simulation uses the compact model of the system. We do not need to simulate all the detailed operations of a WLAN, the performance of other parts in the IEEE 802.11 protocol can be assumed to be the perfect inputs/outputs of the model in the simulation. On the other hand, the Arena platform not only is easy to use, but also can provide powerful optimization and analysis tools. Therefore, we choose the Arena platform to run the simulation.

We simulate a single BSS with 10 stations and assume that the traffic is evenly distributed among them. For an infrastructure BSS, there is also an AP. The arrival is modeled as a Poisson process and the packet length of 6000 bits is used in the simulation.

The system parameters used in the simulation are configured based on the IEEE 802.11 recommendation [12]. For the backoff procedure parameters, the Backoff Slot Time is $50\mu s$ and the contention window is between 32 and 256. For the basic multi-access parameters, the SIFS is $28\mu s$ and the DIFS is $128\mu s$. The max propagation time is $1\mu s$. The length of the PHY (physical layer) header is set to 128 bits. The length of a beacon interval is set to 100 milliseconds, which is a typical value good for most applications.

We analyze the output when the system reaches steady state and let the system run enough replications to make sure the half width of the 95% confidence interval of a variable is within 5% of its average value.

5.2 Optimization of the power management parameters

The power management of the IEEE 802.11 puts the stations into the Doze state when they are idle to save energy. The duration of the time that a station stays in the Doze state is scheduled by the power management parameters. In an infrastructure network, the power management parameter is the listen interval of a station, which equals to one or several beacon intervals. In an independent network, the power management parameter is the ATIM window size, which is a fraction of a beacon interval. We want to optimally configure the listen interval or the ATIM window size to achieve good energy efficiency. Here, we simulate the operation of the power management in a WLAN to seek the optimal power management parameters that

can maximize the average PTD of stations with the satisfaction of the average FRT.

We present our simulation results of the infrastructure and the independent network respectively in the following.

5.2.1 Simulation results of an infrastructure BSS

In an infrastructure BSS, the AP announces the TIM message at the beginning of each beacon interval and the stations, which are scheduled to wake up, enter the Awake state to listen to the TIM. If there is traffic buffered for a station, it stays awake until it finishes retrieving the traffic from the AP. Otherwise, it enters the Doze state immediately. In order to retrieve the traffic buffered in the AP, a station polls the AP with the DCF mechanism first. If the poll is successful, the AP responds with the buffered traffic. The buffer of the AP maintains a FIFO queue for the pending traffic.

The system load 50% are simulated. The listen interval that equals to 1∼10 beacon intervals are selected as the control variables.

Fig.5.1 shows the influence of the listen interval over the average FRT. We plot the simulation result together with the analytical results. The average FRT increase linearly with the increase of the listen interval. We observe that the lines of the simulation result, the analytical result by the D/G/1 queue model and the analytical result by the M/G/1 model are very close.

Fig.5.2 shows the impact of the listen interval over the average PTD. The dotted curve is our simulation result, which locates between the analytical lower bound and the analytical upper bound. Also, the trend of simulation result is similar with that of the analytical bounds. They all increase with the increase of the listen interval, which indicates that the larger listen interval leads to good energy efficiency.

Therefore, a larger listen interval gives better energy efficiency and poorer response time. The simulation results also suggest to select the largest listen interval under

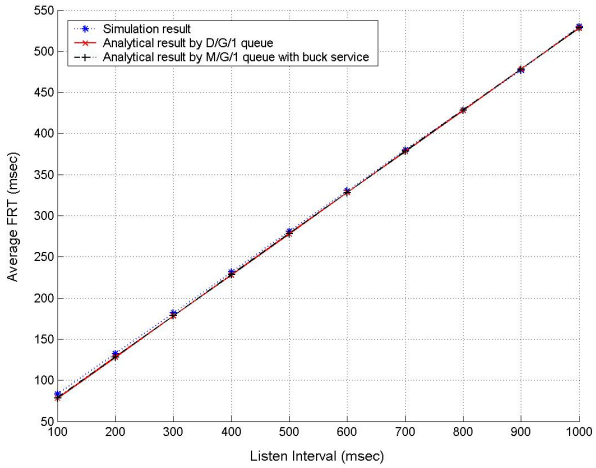

Figure 5.1: FRT increases with listen intervals.

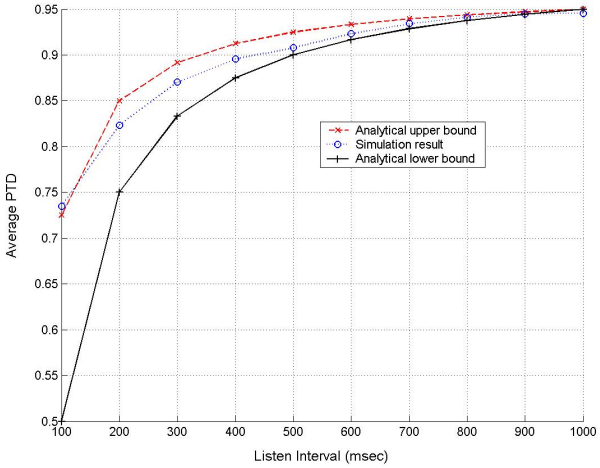

Figure 5.2: PTD increases with listen intervals.

the requirement of the response time. The simulation results support our analytical results.

In addition, the transition between the Awake state and the Doze state experiences delay and consumes energy. A larger listen interval that leads to the less number of the transitions can contribute to the energy saving objective.

5.2.2 Simulation results of an independent BSS

In the power management of an independent BSS, all stations enter the Awake state at the beginning of each beacon interval and stay awake for a period of time called ATIM window. During the ATIM window, the stations that have buffered traffic announce the ATIMs and wait for the acknowledgement from the destination stations. Following the ATIM window, the stations that successfully announced or acknowledged the ATIMs stay awake and exchange the actual traffic. Other stations go to the Doze state in the remaining beacon interval. Both the ATIMs and the actual traffic are transmitted with the DCF mechanism.

The ATIM window size is the power management parameter of an independent BSS in the IEEE 802.11 standard. Different ATIM window sizes can be used to schedule the traffic transmission and the period of time a station stays in the Doze state. Here, we use the ratio of the beacon interval over the ATIM window size as the control variable. Since the ATIM window is a part of the beacon interval, a smaller ratio value indicates a larger ATIM window size. The system load 50%, 37.5%, and 25% are simulated.

Fig.5.3 shows the change of the average FRT with the ratio of beacon interval over ATIM window size. Generally, when the ratio increases, the response time decreases first, then increases, which indicates that it is a concave curve and that there is a minimum value for the response time. The FRT keeps low and stable when the ratio is between 8 and 14. Also, the FRT of the heavy system load is larger than that of

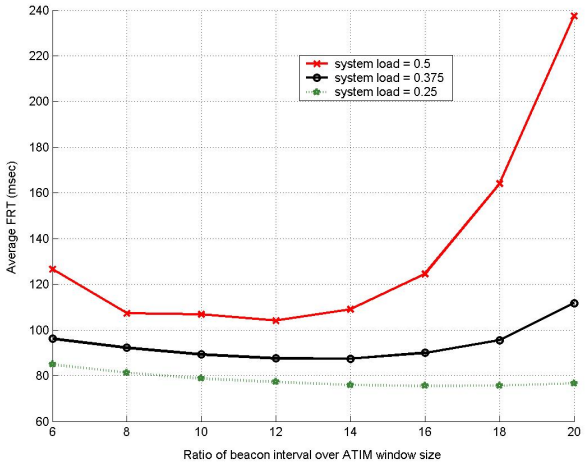

Figure 5.3: Concave curves of FRT with different ratios of beacon interval over ATIM window size.

the light system load.

Fig.5.4 illustrates that the PTD increases slowly with the increase of the ratio. The curves approximately stay stable when the ratio value ranges from 10 to 14. Also, the PTD of the heavy system load is smaller than that of the light system load.

Based on the results of the FRT and the PTD, we find that the system achieves good performance metrics when the ratio value is between 10 and 14.

An appropriate ATIM window size can optimize the energy efficiency. If the ATIM window is too small, not all the buffered traffic can be announced. The buffered traffic is accumulated and the FRT increases. If the ATIM window is too big, the time duration that all stations stay awake increases and the time interval for the transmission of the actual traffic becomes smaller. This also result in the increase of the FRT because the transmission of the announced traffic can not be finished in the current beacon interval and has to be postponed to the next beacon interval. Therefore, for a given independent BSS, we suggest to choose the ATIM window size that within the range of the low and stable FRT time and the stable PTD.

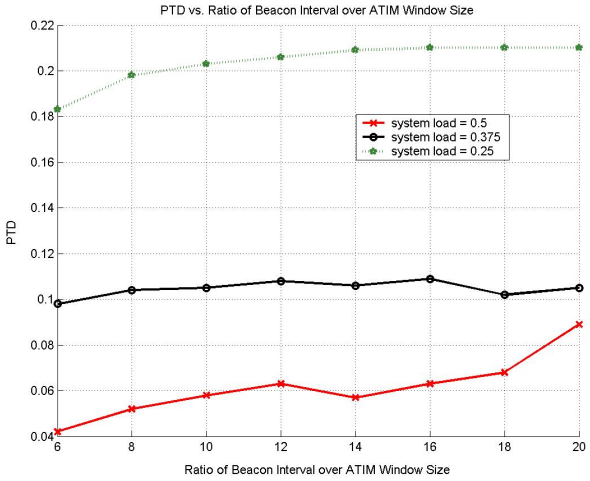

Figure 5.4: PTD with different ratios of beacon interval over ATIM window size.

5.2.3 Suggestions for system configuration

In conclusion, more energy can be saved by properly schedule the time a station in the Doze state. In an infrastructure BSS, the FRT increases linearly with the increase of a listen interval. We propose to select the largest listen interval under the requirement of the FRT. In an independent BSS, there is an optimal value for the ATIM window size. we propose to choose the ATIM window size that within the range of the small FRT and the stable PTD.

5.3 Enhanced power management for an infrastructure BSS

In this section, we propose an enhanced power management mechanism for an infrastructure BSS. First, we give our observation about the focus we want to improve for the power management in the IEEE 802.11. Then, we introduce our enhanced power management mechanism. Next, we evaluate the performance of our scheme

analytical. Finally, we give our simulation results that shows the enhanced power management scheme outperforms the original one.

5.3.1 Observations

We observe that the poll scheme and buffer management of the IEEE 802.11 can be enhanced to save energy.

Poll scheme

In the IEEE 802.11, when a station retrieves its buffered packets from the AP, the station sends a poll frame to the AP and the AP responds with the buffered packets. When two or more stations send poll frames at the same time, collisions happen. We propose a scheme that eliminate these collisions to save energy.

Buffer management

In the IEEE 802.11, the AP temporarily buffers the traffic if the destination station is in the Doze state. The protocol does not specify how to manage the AP buffer. It says, " The algorithm to manage this buffering is beyond the scope of this standard." We investigate the buffer management of the AP and propose to set up multiple queues and let the AP schedule the transmission sequence.

5.3.2 Enhanced power management scheme

We enhance the original power management in the IEEE 802.11 based on our observation.

Poll scheme

In order to reduce the collisions when the stations poll the AP to retrieve the buffered traffic, we let the AP schedule the transmission sequence of the buffered packets. The AP uses a bit map to indicate which station should poll and retrieve the traffic next.

In the beginning of beacon intervals, the TIM messages piggybacks the bit map. In the middle of the beacon intervals, the bit map is piggybacked by the data frames that the AP forwards to the stations. In our new scheme, the stations that have buffered traffic listen to the bit map and decide the time to send a poll frame. The AP responds the poll frame with the buffered packets and the bit map that indicates the next station that should poll.

The new scheme reduces collisions and saves energy. On the other hand, when the AP controls the transmission sequence of the packets, the energy efficiency can be further improved by proper buffer management and good scheduling algorithms.

Buffer management and scheduling algorithms

In the original IEEE 802.11, the stations initiate the poll frames and compete for the medium. The transmission sequence of the buffered traffic is random and the simplest FIFO queue is always used. Here, we propose to set up multiple queues in the AP buffer, one queue for each station. A queue and a station is one to one corresponding. When all the traffic in a queue is transmitted, the corresponding station enters the Doze state. The time duration that a station stays in the Doze state can be scheduled through scheduling the transmission of the traffic in the queues.

Therefore, the structure of multiple queues and the central control ability of the AP provide a good stage for applying scheduling algorithms to control the transmission sequence and the time a station is in the Doze state.

Scheduling algorithms

In terms of scheduling the transmission sequence, we can either transmit all the packets in one queue each time or transmit one packet from different queues each time. Generally, the scheduling algorithms can be divided into two categories: the Queue-by-Queue algorithm and the Packet-by-Packet algorithm. We define their operations here.

(a). Queue-by-Queue:

The scheduler in an AP schedules all the traffic in one queue each time. In other words, it selects a queue and announces it by the bit map. The selected station continuously polls and retrieves the packets until the queue is empty.

After finishing the transmission for the selected station, the scheduler selects another queue and repeats the process, until all the buffered traffic for the awake stations is transmitted or the current beacon interval is used up.

(b). Packet-by-Packet:

Each time the scheduler selects a packet among the packets in the head of all the queues and announces it by the bit map. The picked station polls and retrieves the head packet. When the transmission finishes, the AP schedules another packet among all the head packets and transmits it, until all the buffered traffic for the awake station is transmitted or the current beacon interval ends.

Thus, the scheduling unit of the Queue-by-Queue algorithm is all the traffic in a queue for a station, while the scheduling unit of the Packet-by-Packet algorithm is only one packet.

We use the subscript 'q' to stand for the Queue-by-Queue algorithm, and the subscripts 'p' to describe the Packet-by-Packet algorithm. For example, FRT_q is the FRT under the Queue-by-Queue algorithm, and FRT_p is the FRT under the Packet-by-Packet algorithm.

5.3.3 Performance evaluation

We evaluation the performance of the enhanced power management scheme analytically.

FRT

The average FRT remains the same, although we change the transmission order of

the packets by applying scheduling algorithms. This can be explained from the workload view point. For certain amount of buffered traffic, the server is busy from the beginning of the serving the first packet to the end of the serving the last packet. No matter with what order the packets are served, the total amount of workload is the same, so does the total service time. Therefore, the average FRT in our system is the same with different buffer scheduling algorithms.

PTD

With the enhanced poll scheme, collisions are reduced and energy is saved. The average PTD of the enhanced scheme is better than that of the original one. Furthermore, the average PTD of the enhanced scheme is different under different scheduling algorithms.

Assume there are n stations in the system. Arrivals to each station is Poisson process with parameter $\frac{\lambda}{n}$. We denote the number of buffered packets for a station at the beginning of a listen interval by X. The probability that a station has buffered packets at the beginning of a listen interval is $P(X > 0)$. Let M denote the number of stations that have buffered packets at the beginning of a listen interval, $M = n * P(X > 0)$, where $M \leq n$. Also, we denote the time a station stays awake in a listen interval as U.

Assume the expected service time of a packet is $E[S]$, the average queue length in the AP for a station is $E[Q]$.

We analyze the average PTD under the Queue-by-Queue algorithm and the Packet-by-Packet algorithm respectively.

(a). Average U of Packet-by-Packet algorithm

Assume a packet is randomly selected with probability $\frac{1}{M}$ every time. We denote the response time of a packet as S', which consists of the waiting time and the transmission time. When the packets from other queues are selected with probability $1 - \frac{1}{M}$, the waiting time starts.

We have,

$$S' = \begin{cases} S & \frac{1}{M} \\ S + S' & \frac{M-1}{M} \end{cases}$$

$$E[S'] = E[S]\frac{1}{M} + (E[S] + E[S'])\frac{M-1}{M}$$

$$E[S'] = ME[S]$$

Therefore, the average time a non-empty queue stays awake during an interval is:

$$E[U]_p = E[S']E[Q] = ME[S]E[Q]$$

(b). Average U of Queue-by-Queue algorithm

We view all the packets of a queue for a station as a batch. Denote the workload of the batch as S''. We have $E[S''] = E[Q]E[S]$.

The average time a batch spends in the system is:

$$E[U]_q = \frac{(1 + 2 + \ldots + M) * E[S'']}{M} = \frac{M+1}{2}E[S]E[Q]$$

(c). Comparison of the average PTD

We compare the average time a station stays awake in a system under the two different schemes.

$$E[U]_p - E[U]_q = ME[S]E[Q] - \frac{M+1}{2}E[S]E[Q]$$
$$= \frac{M-1}{2}E[S]E[Q]$$

If $M = 0$, it means that no traffic in the buffer of the AP; if $M \geq 1$, $E[U]_p \geq E[U]_q$.

The average percentage of time a station stays in the Doze state can be expressed as:

$$E[PTD] = \frac{M(T - E[U]) + (n - M)T}{nT}$$

Where, the first item in the numerator is the total time in the Doze state of all the stations that have traffic at the beginning of a listen interval; the second item in the numerator is the total time in the Doze state of those stations that have no traffic at the beginning of a listen interval; the denominator is the total time spend by all the stations.

Since $E[U]_p \geq E[U]_q$, we obtain, $E[PTD]_p \leq E[PTD]_q$. The Queue-by-Queue algorithm is better than the Packet-by-Packet algorithm in terms of the energy efficiency performance.

In conclusion, we suggest to set up multiple queues in the AP buffer and let the AP schedules the transmission sequence with the Queue-by-Queue algorithm.

5.3.4 Simulation results

In this section, we simulate our enhanced power management scheme and compare its performance with the original power management scheme in the IEEE 802.11.

Comparison of poll schemes

In our scheme, multiple queues are maintained in an AP, one queue for each station. Assume the buffer is large enough and there is no packets lost. Because in the original power management, the traffic is transmitted packet by packet, we use the Random Packet algorithm to compare the performance.

Fig.5.5 shows the average FRT increase linearly with the increase of the listen interval under different power management schemes. The curves are almost the same. This means the average FRT keeps the same under different schemes.

From Fig.5.6, we observe that the average PTD of the Random Packet algorithm in which the AP schedules the transmission is better than that of the original power management scheme.

The price for the improvement is that the operation of the AP becomes complicated.

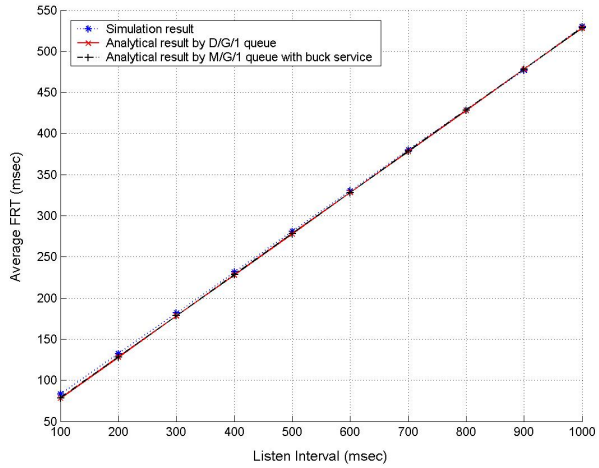

Figure 5.5: Comparison of the average FRT.

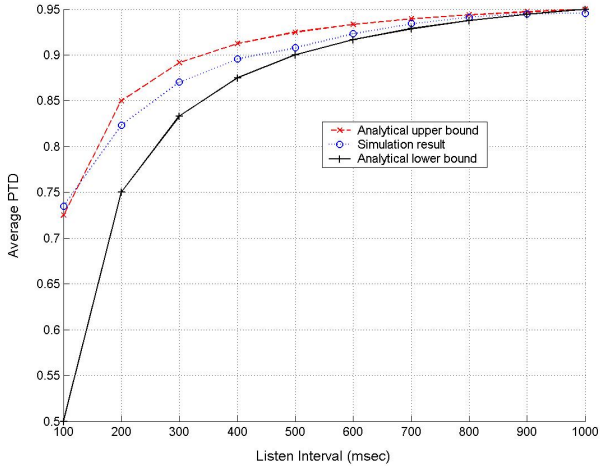

Figure 5.6: Comparison of the average PTD.

Table 5.1: Scheduling algorithms used in the simulation

Random Packet	Randomly pick a head packet from any queue each time
Random Queue	Each time randomly select a queue among all the queues
EDD (Earliest Due Date) Queue	The queue whose head packet arrives the earliest is transmitted first

Comparison of enhanced and original power management schemes

We apply all the enhanced features (poll scheme, buffer management, scheduling algorithms) and compare their performance with the original power management scheme in the IEEE 802.11 protocol. Table 5.1 lists the scheduling algorithms used in our simulation.

Fig.5.5 confirms our analysis that the average FRT remains the same. Fig.5.6 shows the relation between the average PTD and the listen interval of a station with different scheduling algorithms under the. The average PTD also increases with the larger listen intervals. Among the four power management schemes, we observe that $E[PTD]_{original} < E[PTD]_{RandomPacket} < E[PTD]_{EDDQueue} < E[PTD]_{RandomQueue}$.

Based on Fig.5.5 and Fig. 5.6, we have two observations.

(a). The AP-scheduled poll scheme has better energy efficiency than the station-initiated poll scheme.

(b). The Queue-by-Queue algorithm (including EDD Queue and the Random Queue) are generally better than the Packet-by-Packet algorithm (Random Packet).

Here,we simulate the EDD Queue scheduling algorithm with two purposes. First, we check if the average FRT changes or not when we pick the queue whose head packet arrive the earliest each time. From the simulation results, the average FRT remains the same. The delay jitter, which is not the performance metrics in this paper, might be improved. Second, we check the impact of this algorithm on the

average PTD. Because the queue whose packet arrives the earliest tends to be longer. So under the EDD Queue algorithm, we tend to select the longer queue first, this leads to lower average PTD than that of the Random Queue. This indicates that if the longer queue is served first, the waiting time of other queues becomes longer. So the average PTD decreases. This also indicates that we can serve the shorter queue first when the system load is light to save more energy.

In conclusion, the simulation results also shows that setting up multiple queues in the buffer of an AP and scheduling the transmission of the buffered traffic with appropriate algorithm can improve energy efficiency in WLAN. Among the scheduling algorithms, the Queue-by-Queue algorithm is generally better than the Packet-by-Packet algorithm.

5.4 Enhanced power management for an independent BSS

Collision and idle listening are two main reasons of energy inefficiency. In this section, we propose an enhanced power management scheme to reduce collisions in an independent BSS.

In the IEEE 802.11 protocol, DCF is used in both ATIM window for traffic announcement and in the interval following the ATIM window for traffic exchange.

In the ATIM window of a beacon interval, stations compete for media to announce its traffic. The announcement message itself is small. Even if collisions happen, the energy wasted is not significant. However, the size of the traffic packets transmitted in the interval following the ATIM window is much larger than the announcement message. If collisions happen, the energy wasted is also much larger. Therefore, reducing the latter collisions can save more energy.

We propose the contention free media access for the transmission of the actual

traffic in the interval following the ATIM window. Since keeping DCF in the ATIM window is the design requirement of a distributed mechanism, we suggest that DCF is applied only in ATIM window for traffic announcement. In the interval following the ATIM window, the traffic packets are transmitted according to scheduling.

The simplest scheduling algorithm is to transmit the traffic with the same order as the ATIM message announcement. That means, the first announced message is transmitted first, the second announced message is transmitted the second, and so on.

In order to implement the contention free traffic transmission, we add several new features to the system.

(a). Add a broadcast message after each ATIM announcement/ATIM ACK pair. The message, which is sent by the source station, contains the address of the source and destination station. In the ATIM window, every station stays awake to listen to the broadcast.

(b). Add a table in each station to record the order of the ATIM announcement/ATIM ACK pair. A station also need to maintain a pointer, which points to the next transmitting/receiving pair in the table.

(c). In the interval following the ATIM window, after the transmission of the traffic of a source/destination pair, the source station broadcasts a message which tells all the other stations that the current transmission is finished.

(d). After receiving the broadcast of the transmission completion notice, the pointer of the table in a station is moved to the next transmitting/receiving pair. If the current source station is the station itself, the station sends out all the buffered traffic for the destination station. If the current source station is other station, the station listens to the next transmission completion message.

(e). In a station, when a pointer moves to the end of the table, all the announced

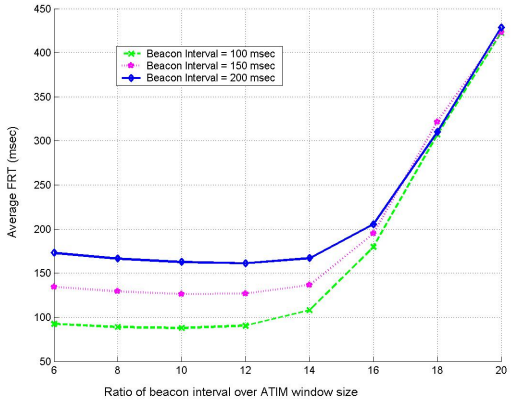

Figure 5.7: Average FRT of the enhanced power management scheme.

traffic are sent out. If the current bacon interval still has not finished, and there are some new incoming traffic in the buffer, and the station knows the destination station is also awake by checking the source/destination pair in the table, the station can send the newly buffered traffic with the DCF mechanism.

We simulate the enhanced power management and compare it with the original power management in an independent BSS. The system load is 50 %. The beacon interval 100 *msec*, 150 *msec*, and 200 *msec* are simulated.

Fig.5.7 and Fig.5.8 shows the performance of the enhanced power management scheme. Fig.5.9 and Fig.5.10 illustrates the performance of the original power management mechanism.

We compare the FRT performance by Fig.5.7 and Fig.5.9. All the curves are concave. The curves in the range when the ratio is between 6 and 14 is stable. The optimal ATIM window size is selected in this range. Within the stable range, the average FRT of the enhanced scheme is lower than that of the original power management scheme. Also, the difference between the curves in Fig.5.7 is smaller than that in Fig.5.9. This means the enhanced scheme is less sensitive to the change of the beacon interval and thus more robust.

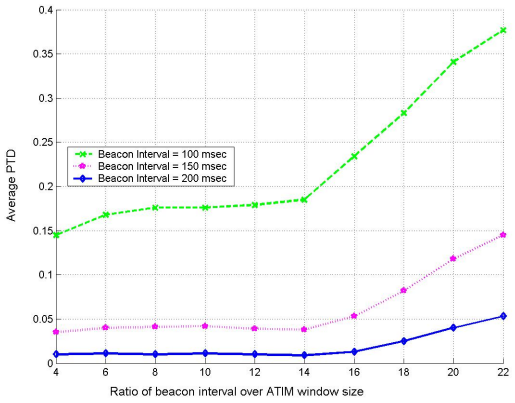

Figure 5.8: Average PTD of the enhanced power management scheme.

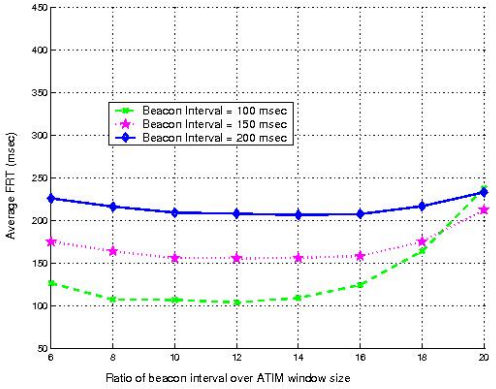

Figure 5.9: Average FRT of the original power management scheme.

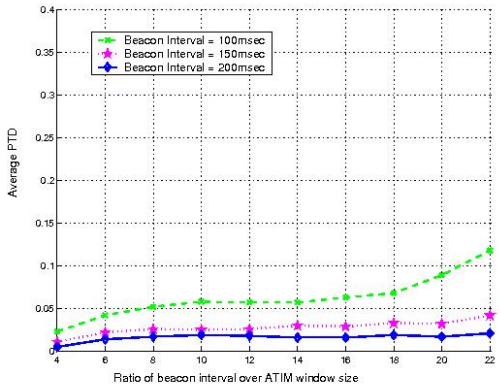

Figure 5.10: Average PTD of the original power management scheme.

We compare the PTD performance by Fig.5.8 and Fig.5.10. The curves are stable when the ratio is between 6 and 14. when the ratio is greater than 14, although the PTD becomes larger, the FRT also increases dramatically. Thus, the ratio between 6 and 14 is good for ATIM window size. Within the stable range, the PTD of the enhanced scheme is larger than that of the original scheme.

In conclusion, the enhanced power management mechanism has lower FRT and larger PTD. It outperforms the original power management scheme.

Chapter 6

Power Saving Mechanism in UMTS

Energy efficiency is becoming a critical issue in the 3G systems. UMTS provides higher data rate, packet-switched multimedia traffic in addition to traditional circuit-switched voice traffic. Multimedia traffic has the characteristics of busty traffic, asymmetric throughput, variable rates, and different traffic patterns of different applications. These features bring in the challenge of designing a good energy efficiency mechanism. We study the power saving mechanism in UMTS and give proposals to improve it.

In this chapter, we introduce the power saving mechanism in UMTS, identify the two key parameters DRX cycle and inactivity timer, and discuss the difference and similarity of power saving mechanism between WLAN and UMTS. Also, we describe the traffic models used in our study.

6.1 Power saving mechanism

6.1.1 Radio resource control states

Managing power of mobile equipments is one of the objectives of the radio resource control in UMTS. UMTS defines four RRC states [21], which provide different transmission speeds and consume different power. Fig.6.1 shows the RRC states and their

transitions defined by UMTS. A UE can save energy by appropriately selecting states and making transitions. Our study focus on the operation of UE in the connected mode.

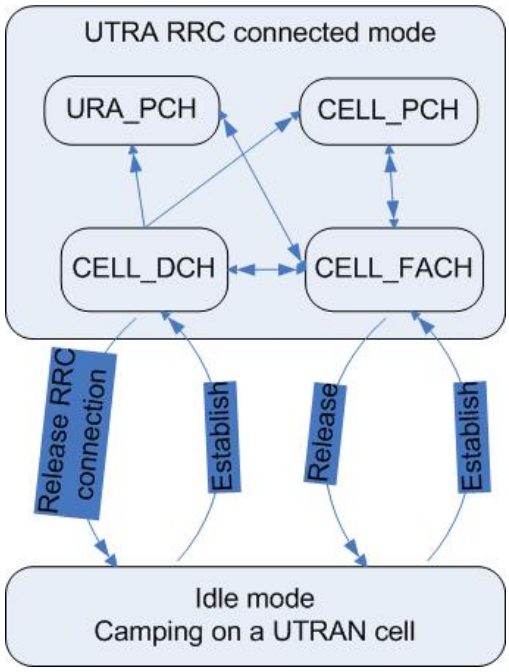

Figure 6.1: RRC states and state transition in UMTS.

We introduce the characterization of each state in the connected mode.
CELL_DCH (Dedicated channel) state:

- A dedicated physical channel is allocated to the UE in uplink and downlink.
- The UE is known on cell level according to its current active set.
- Dedicated transport channels, downlink and uplink (TDD) shared transport channels, and a combination of these transport channels can be used by the UE.
- The bit rates of DCH can be up to 2Mbps. A DCH is established to continuously transmit large amount of data.

CELL_FACH (Forward access channel) state:

- No dedicated physical channel is allocated to the UE.
- The UE continuously monitors a FACH in the downlink.
- The UE is assigned a default common or shared transport channel in the uplink (e.g. RACH) that it can use anytime according to the access procedure for that transport channel.
- The position of the UE is known by UTRAN on cell level according to the cell where the UE last made a cell update.
- The bit rates of RACH/FACH is low. If a small amount of data is to be transmitted, a UE selects CELL_FACH state.

CELL_PCH (Paging channel) state:

- No dedicated physical channel is allocated to the UE.
- The UE selects a PCH with the algorithm, and uses DRX for monitoring the selected PCH via an associated PICH.
- No uplink activity is possible.
- The position of the UE is known by UTRAN on cell level according to the cell where UE last made a cell update in CELL_FACH state.
- A UE maintains necessary update information of networks and does not transmit or receive user traffic.

URA_PCH (Paging channel) state:

- No dedicated channel is allocated to the UE.
- The UE selects a PCH with the algorithm, and uses DRX for monitoring the selected PCH via an associated PICH.

- No uplink activity is possible.
- The location of the UE is known on UTRAN registration area level according to the URA assigned to the UE during the last URA update in CELL_FACH state.
- A UE maintains necessary update information of networks and does not transmit or receive user traffic.

The power consumption in these states are different. In Cell_PCH state, both the receiver and transmitter of a UE are active. In the Cell_FACH state, only the receiver of a UE is active. In the Cell_PCH or URA_PCH state, the receiver of a UE can use DRX mechanism.

With DRX mechanism, the receiver in a UE is turned off at most time and turned on periodically to listen to the network. If there is incoming traffic, a UE enters the Cell_FACH or Cell_DCH state. Else, it enters another sleep cycle. The sleep cycle is called DRX cycle. The state transitions are triggered by inactivity timers.

The values of the typical power consumption in the Cell_DCH state are 200-400mA, in the Cell_FACH state are 100-200mA and in Cell_PCH or URA_PCH state are below 5mA [47]. It is obvious that putting a UE into Cell_PCH or URA_PCH state can save lots of energy.

6.1.2 Comparison of power saving mechanisms in UMTS and WLAN

In wireless networks, the basic idea of power saving is to put a mobile equipment into the low power consuming state when it is idle and to wake it up periodically to listen to the network. We call the low power consuming state the sleep state and the high power consuming state the active state.

The longer sleep period saves more energy with the price of extra delay. Optimal parameters need to be found to achieve maximum energy saving with the satisfaction of other system requirements.

In WLAN, the listen interval is used to control the time that a mobile stays in the sleep state. In UMTS, when a UE is in CELL_PCH or URA_PCH state, DRX mechanism is applied to save energy. The parameter to control the sleep period is DRX cycle. Also, considering that UMTS provide multiple services and applications, the inactivity timer parameter is used to maintain the smooth of data bursts.

The similarity of the power management schemes in WLAN and UMTS is the basic mechanism, which puts a mobile into the sleep state when it is idle and wakes it up periodically.

The are several differences between the power management schemes in WLAN and UMTS.

- The signaling in UMTS is more complicated and a mobile may travels across cells from time to time. The overhead of maintaining and changing the corresponding parameters is higher.

- There is more mobile states in UMTS (CELL_DCH, CELL_FACH, CELL_PCH, and URA_PCH) than that in WLAN (Doze state and Awake state).

- The state change from the active state to the sleep state is smoother in UMTS than in WLAN. In WLAN, a mobile goes to sleep immediately after it finishes retrieving the buffered traffic. In UMTS, in addition to the parameter DRX cycle, the other parameter, the inactivity timer, is used to ensure the smooth transmission of data bursts, and furthermore to maintain the stability of network traffic.

6.1.3 Model of power management mechanism

In this study, we define the Cell_DCH and Cell_FACH state as the active state, in which data can be transmitted and received. The Cell_PCH and URA_PCH state are defined as the sleep state, in which a UE stays in the lower power consumption state and wakes up periodically.

Fig.6.2 shows the active and sleep pattern of a UE under power saving mechanism. In the active state, a UE transmit/receive (busy) first. After it finishes transmitting/receiving, it remains in the active state and stays idle for some time, then enters the sleep state. The duration of idle time is controlled by an inactivity timer. If there is new traffic arriving before the inactivity timer expires, the UE becomes busy again. In the sleep state, a UE wakes up periodically to listen to the network. If there is traffic to receive, the UE enters the active state, else it goes to another sleep period. The sleep period is determined by the DRX cycle.

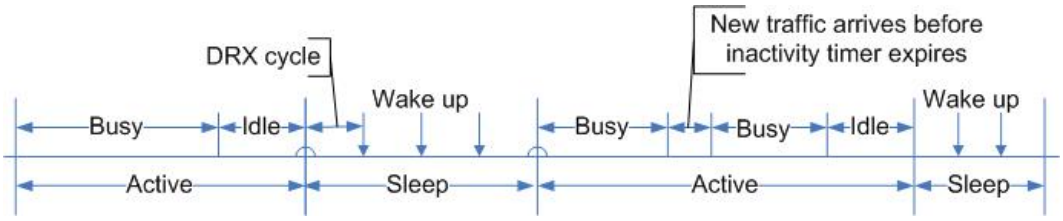

Figure 6.2: Active and sleep pattern.

Setting an inactivity timer can keep the continuity of bursty data, but it consumes more energy because of idle waiting. For DRX mechanism, too small cycle size leads to unnecessary waking up and too large cycle size causes large delay. A suitable DRX cycle size and inactivity timer value that can maximize energy efficiency and satisfy other system requirements will be investigated for several traffic models in this paper.

6.2 Performance metrics of power management in UMTS

We use packet delay, energy consumption, and percent of complete calls as the performance metrics of power saving mechanism in UMTS. We give detailed description of these performance metrics in the chapter of simulation.

6.3 Traffic models

6.3.1 Review of traffic models

The first generation and the second generation wireless networks mainly provide voice communications. The third generation wireless networks serves more variable traffic types, such as data, video, etc. In traditional voice communication systems, the traffic arriving is modeled as a Poisson process. The Poisson traffic presents irregular characteristic in small scale and becomes smoother when the time scales increase. However, according to extensive measurement and analysis of traffic, many data and video traffic display self-similar property, i.e., the irregularity regardless of time scales. Therefore, the Poisson traffic model is not suitable for analysis of multimedia traffic in 3G networks. New traffic models are needed for analyzing the performance of 3G networks.

Modeling network traffic is an important topic in the performance evaluation of networks. Both theoretical analysis based and measurement based approaches are used [48] [49]. The typical parameters of a traffic model include mean value, higher moments, autocorrelation function, marginal density, etc. Generally, traffic models are divided into the following several categories:

- Short range dependent models, where the autocorrelation function decays exponentially. They include Renewal models, Markov models, Autoregressive mod-

els, TES (Transform-Expand-Sample) models, etc.

- Long range dependent models, where the autocorrelation function decays very slowly. They include Heavy-tail on-off model, Monofractal models, Fractal Brownian model, Distorted Gaussian model, etc.
- Multifractal models, which are the generalization of self-similar process.
- Fluid traffic model, which views the traffic as a fluid like stream. The size of the individual unit (such as packet) is not explicitly modeled. This model makes the analysis and simulation easier.

Traffic models sketch the characteristics of real traffic in networks. Different traffic models suit different traffic types. It is important to select good traffic models to analyze the performance of a network.

6.3.2 Traffic models used in energy saving performance evaluation

UMTS defines four QoS classes in terms of the transfer delay, guaranteed bit rate, traffic handling priority and allocation/retention priority. The four classes are conversational class, streaming class, interactive class and background class.

The energy saving mechanism is to put a UE into the sleep state when it is idle. Since the conversational class and the streaming class are real time applications and demand tight delay, it is better to always keep the mobile in the active state when the UE is in the connected mode. The interactive class and the background class tolerate delays. It is a good choice to use energy saving mechanism for these two classes. The DRX mechanism and inactivity timer can help a UE to save energy when it transmits non-real time traffic.

Since the background class has the lowest priority, it can be done whenever the network and a UE have extra resources. Here, we study the UMTS power management scheme for the interactive class and find the optimal parameters.

The typical traffic models of the interactive class suggested by 3GPP specifications include FTP model, Modified gaming model, Web browsing model, etc [50]. In our study, we use FTP model and Modified gaming model to evaluate the performance of power saving mechanism in UMTS.

FTP Model

In FTP applications, a session includes a series of file transfers. Between file transfers, there are reading times. Fig.6.3 shows the FTP traffic model.

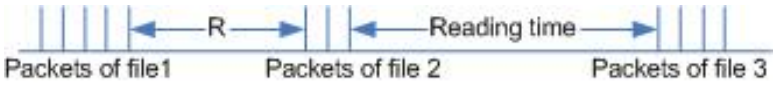

Figure 6.3: FTP traffic model.

Here is the typical parameters for FTP applications. Denote the size of a file to be transferred as F. The distribution of F is truncated lognormal with the mean of 2MB, the standard derivation of 0.722MB and the maximum size of 5MB. A big file is always fragmented into many small packets to transmit. Denote the reading time as R, which is the time interval between the end of downloading previous file and the user request for the next file. The distribution of R is exponential with the mean of 180 seconds. The above parameters are based on measurement and statistical calculation, and proposed by 3GPP specification.

Modified Gaming Model

A gaming session consists of a sequence of packet call arrival processes, separated by reading times. Within a packet call, there is a series of datagram arrival processes.

Fig.6.4 describes the modified gaming traffic model.

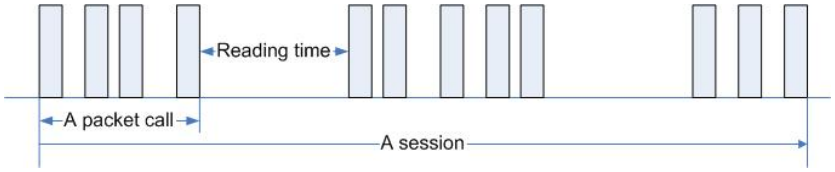

Figure 6.4: Modified gaming traffic model.

Table 6.1 gives the parameters of a typical gaming model.

Table 6.1: Typical parameters of a modified gaming model

Parameter	Value	Distribution
Mean packet call duration	5s	Expo
Mean reading time	5s	Expo
Datagram size	576 bytes or 1500 bytes	Fixed
Mean datagram interarrival time	40ms	Log-normal 40ms std

Chapter 7

Simulation of Power Saving Mechanism in UMTS

In this chapter, we investigate the impact of inactivity timer and DRX cycle on the performance metrics and find the optimal configurations for power saving parameters by simulation. From the simulation results, we propose to adaptively configure the power saving parameters based on real time measurement.

7.1 Simulation environment

7.1.1 Simulation configuration

A UE communicates with others through a RNC as shown in Fig.7.1. The UE communicates with Node B through wireless media and the Node B communicates with RNC through wired or wireless connection. The function of Node B is to relay the transmission, i.e., to execute the command from RNC and report the information coming from UE to RNC. We assume perfect channel condition and the Node B can be viewed as transparent. Therefore, we simplify the problem by studying the communication between the UE and the RNC. The buffer of uplink traffic locates in the UE and the buffer of downlink traffic is in the RNC. RNC controls the RRC states of the UE, their transitions, as well as their parameters configuration.

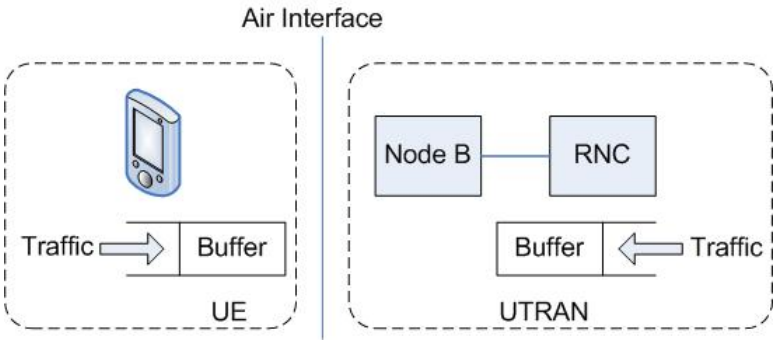

Figure 7.1: Simulation environment.

We develop the simulator with C/C++ programming language. We assume the power consumption is 300mA in the active state and 5mA in the sleep state, and that each wake-up check lasts 40ms.

7.1.2 Input and output

Input

- System parameters: DRX cycle and inactivity timer.
- Traffic models: We use the FTP model and Modified gaming model as input to evaluate the performance of power management scheme for the interactive class traffic. The parameters of the traffic models are introduced in the previous section.

Output

The performance metrics in our study include packet delay, energy consumption, and the percent of complete calls.

- Packet delay: We mainly consider the delay caused by the UE staying in the sleep state. In the uplink, the delay is the time interval between the instant that a packet is generated and the instant that the packet is transmitted by the

UE. In the downlink, the delay is the time interval between the instant that a packet arrives at the RNC buffer and the instant that the packet is transmitted by the RNC. We simplify the analysis by assuming perfect channel condition and ignoring the propagation delay.

- Energy consumption: We measure the energy efficiency by energy consumption with the unit of Joule per bit. The smaller the energy consumption, the better the energy efficiency. We calculate the energy consumption by dividing the total energy used during certain period with the total traffic transmitted during that period. The total energy can be computed by multiplying the total time a UE stays in the active state and the power consumption of a UE in the active state. The total traffic can be obtained through the summation of the size of all the packet transmitted.

- Percentage of complete packet calls: For an FTP model, every time a file is fragmented into small packets to transmit. It can be viewed that all the packets arrive at the same time, which means the file can be transmitted without interruption by the sleep state of a UE. However, in a Modified gaming midel, a packet call consists of many datagrams. The time interval between two arriving datagram is a random number with much smaller time scale than reading time. The packet call might be broken by DRX cycles. It is desired to maintain the continuity of datagrams within a packet call. Therefore, another performance metric, the percentage of complete packet calls is studied.

7.1.3 Results collection and analysis

We collect the results when the system reaches the steady state. The simulation is run long enough to make sure the half width of the 95% confidence interval of a variable is within 5% of its average value.

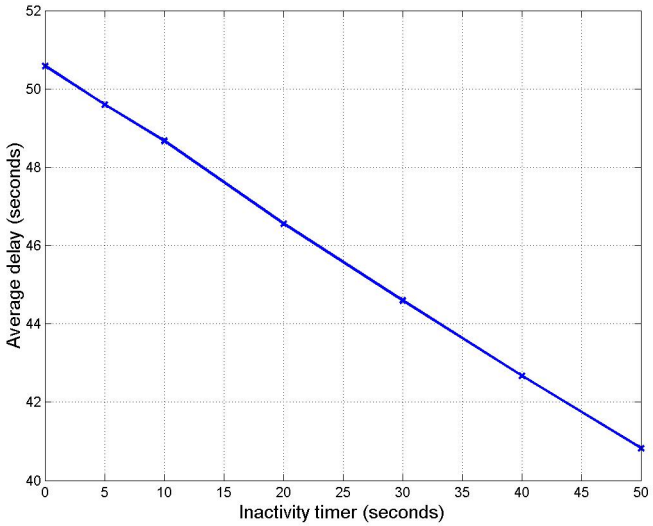

Figure 7.2: Impact of inactivity timer on the average packet delay in a FTP session.

7.2 Energy saving with FTP applications

In this section, we discuss the impact of energy saving mechanism on FTP applications, i.e., the configuration of DRX cycle and inactivity timer on the delay and energy consumption performance metrics. We also propose the optimal way to configure the input parameters.

First, we fix the DRX cycle at 100 seconds and check the impact of the inactivity timer on the average packet delay and energy consumption of an FTP session.

Fig.7.2 shows that the average packet delay decreases linearly when the inactivity timer increases. The inactivity timer increases from 0 to 50 seconds and the average delay decreases from around 50.5 seconds to 40.7 seconds. Thus, larger inactivity timer can reduce delay.

Fig.7.3 shows that the energy consumption increases linearly with the increase of the inactivity timer. When the inactivity timer increases from 0 to 50 seconds, the energy consumption dramatically increases by 5 times, from 0.00002 Joule per bit to 0.0001 Joule per bit. The smaller inactivity timer consumes lower energy.

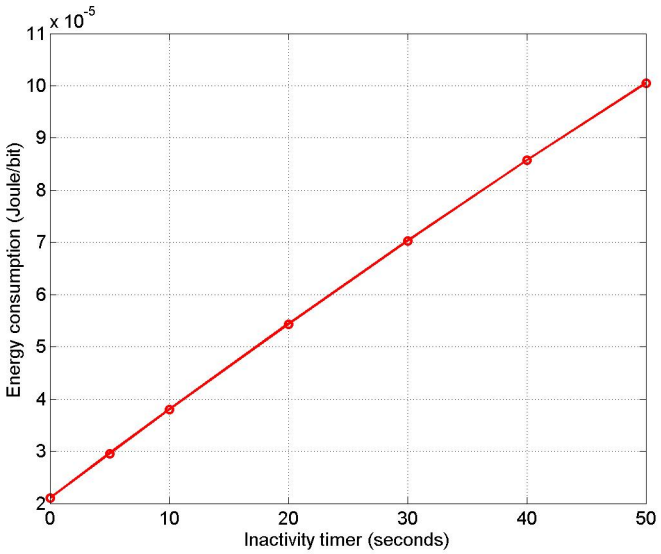

Figure 7.3: Impact of inactivity timer on the energy consumption in a FTP session.

Therefore, the inactivity timer reduces delay with the cost of energy.

In Fig.7.4 and Fig.7.5, we set the inactivity timer as 0 and check the impact of DRX cycle on the performance metrics.

Fig.7.4 shows that the average delay of packets increases linearly when DRX cycle increases. We observe that the average packet delay is about half of the DRX cycle size. For example, when the DRX cycle is 100 seconds, the average delay is 50 seconds; when the DRX cycle is 300 seconds, the average delay is 150 seconds. This is because the distribution of reading time in the FTP model is exponential. Within the DRX cycle, the end of reading time is uniformly distributed and the average ending time is half of the DRX cycle.

Fig.7.5 shows that when the DRX cycle increases, the energy consumption almost remain the same (around 0.000021 Joule per bit). It means that DRX cycle has little impact on the energy consumption. This is because DRX mechanism does not involve idle waiting. After a UE wakes up and listens to the network, it either enters the active state to transmit/receive traffic or goes back to the sleep state immediately.

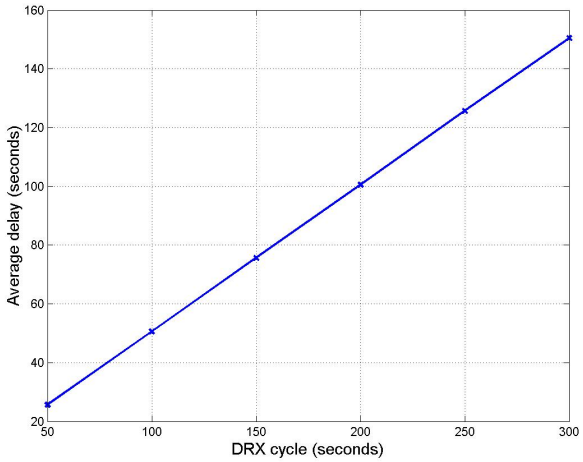

Figure 7.4: Impact of DRX cycle on the average packet delay in a FTP session.

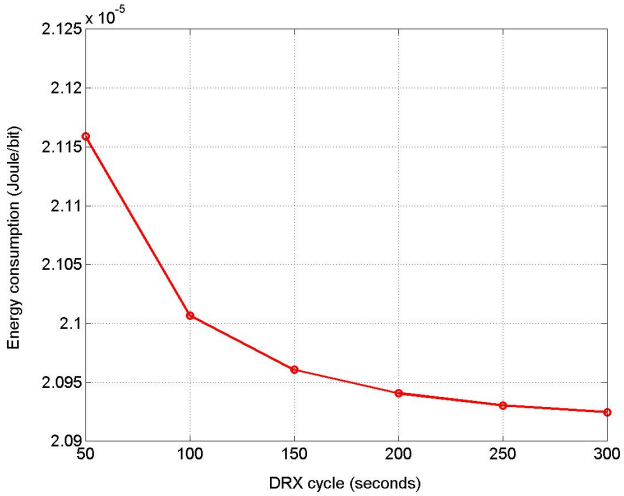

Figure 7.5: Impact of DRX cycle on the energy consumption in a FTP session.

Based on observation, we find that it is better to set the inactivity timer to 0 for FTP sessions. The inactivity timer can reduce delay, but it costs more energy. We can use a smaller DRX cycle to reduce delay without sacrificing energy consumption since energy consumption doesn't change much with DRX cycle. For example, In Fig.7.2, the average delay is reduced to 41 seconds by increasing the inactivity timer to 50 seconds, however, in Fig.7.3 the energy consumption jumps by 5 times, from 0.00002 Joule per bit to 0.0001 Joule per bit. We can achieve the same average delay by setting DRX cycle as 82 seconds in Fig.7.4 and keeping the energy consumption at 0.00002 Joule per bit in Fig.7.5.

The simulation results indicate that the optimal value of the inactivity timer is 0 for an FTP session. This is because the reading time of FTP applications is far greater than the traffic transmission time. Within the value of an inactivity timer, the possibility of the arrival of next file is very small. The inactivity timer costs extra energy without bringing much benefits. Therefore, we suggest setting the inactivity timer to 0 for FTP applications. The value of a DRX cycle should be configured as twice as the largest delay permitted by the system requirement.

7.3 Energy saving with Modified gaming applications

In this section, we discuss the performance of power saving mechanism with a Modified gaming model, i.e., the impact of DRX cycle and inactivity timer on the delay, energy consumption and percentage of complete packet calls of a Modified gaming session.

Compared with FTP applications, modified gaming traffic is more delay sensitive. In a Modified gaming model, the idle time that may be used for energy saving includes the reading time, which is long and comparable to that of the FTP model, and the

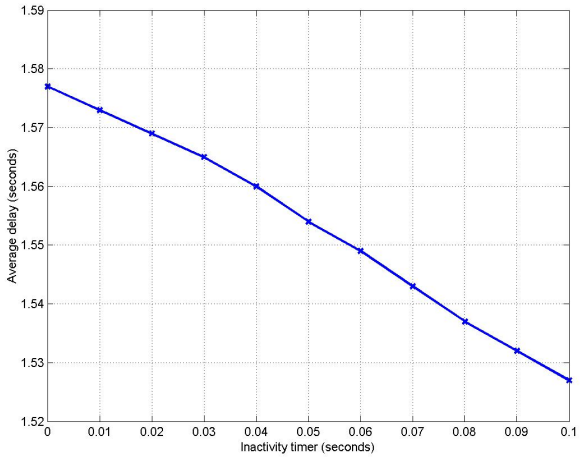

Figure 7.6: Impact of inactivity timer on the average packet delay in a Modified gaming session.

interarrival time between datagrams in a packet call, which is very short.

When DRX cycle is fixed at 3 seconds, we study the impact of inactivity timer on the average packet delay and the energy consumption in a Modified gaming session. Fig.7.6 shows that the average delay decreases when the inactivity timer increases. Fig.7.7 illustrates that the energy consumption increases with the increase of the inactivity timer. This means larger inactivity timer can help reduce the delay with the cost of energy.

To check the impact of DRX cycle on the average delay and energy consumption in a Modified gaming session, the inactivity timer is fixed at 40ms and the results are shown in Fig.7.8 and Fig.7.9.

Fig.7.8 shows that the average delay increases when the DRX cycle increases. Here, unlike an FTP session, the average delay is higher than the half of the DRX cycle. Because there is interval between the arrivals of datagrams, the earlier arrived datagrams experiences more delay when the UE is in the sleep state, which increases the average delay.

Fig.7.9 shows that when the DRX cycle increases, the energy consumption stays

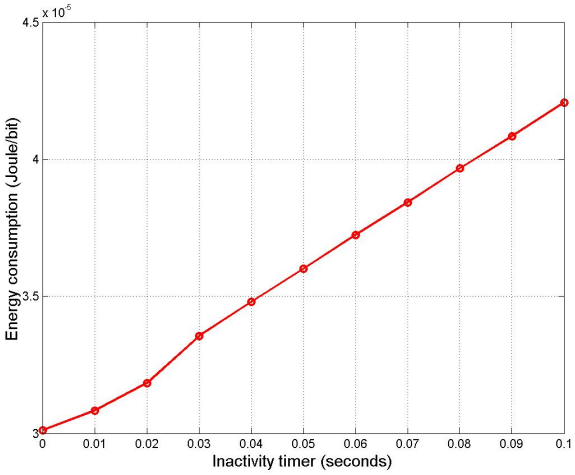

Figure 7.7: Impact of inactivity timer on the energy consumption in a Modified gaming session.

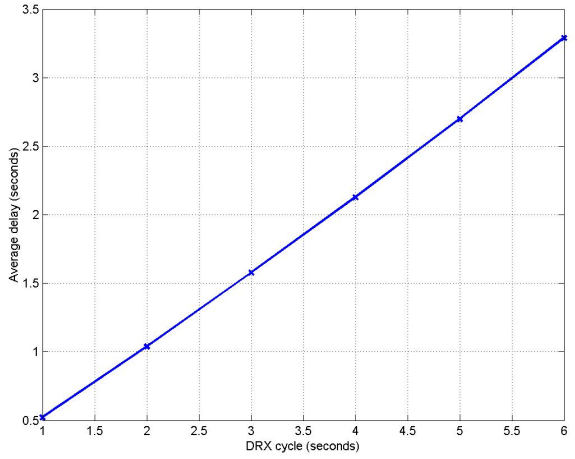

Figure 7.8: Impact of DRX cycle on the average packet delay in a Modified gaming session.

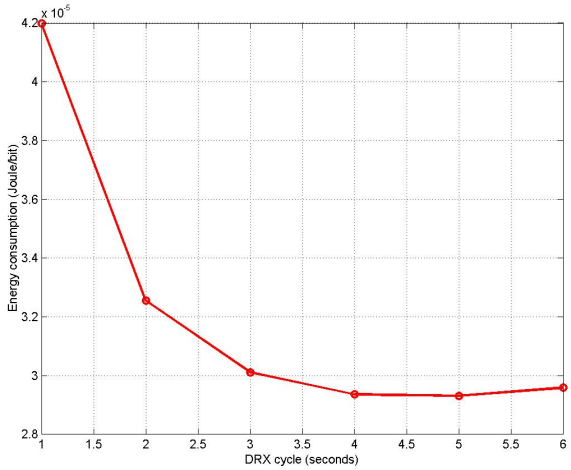

Figure 7.9: Impact of DRX cycle on the energy consumption in a Modified gaming session.

approximately the same, around 0.0003 Joule per bit, with the DRX cycle being 3 - 6 seconds. That means the impact of DRX cycle on the energy consumption can be ignored when the DRX cycle is 3 - 6 seconds. When the DRX cycle is too small, the UE frequently wakes up and listens the networks, which consumes more energy.

Another performance metric for a Modified gaming model is the percentage of complete calls. Fig.7.10 and Fig.7.11 illustrate the the impact of inactivity timer and DRX cycle on it in a Modified gaming session.

The percentage of complete calls increases with both the increase of inactivity timer and DRX cycle. The increase of inactivity timer connects datagrams within a packet call. The increase of DRX cycle makes all the datagrams of a packet call arrive within a DRX cycle, which delays the whole packet call, but it ensures the continuity of the call. Here, the impact of inactivity timer is favored. This is because the arrival interval between datagrams is very small (with mean 40ms), new datagram arrivals within the inactivity timer value can be served immediately, which reduces delay and keeps the traffic flow together within a packet call.

For a Modified gaming session, the requirements of both delay and the percentage

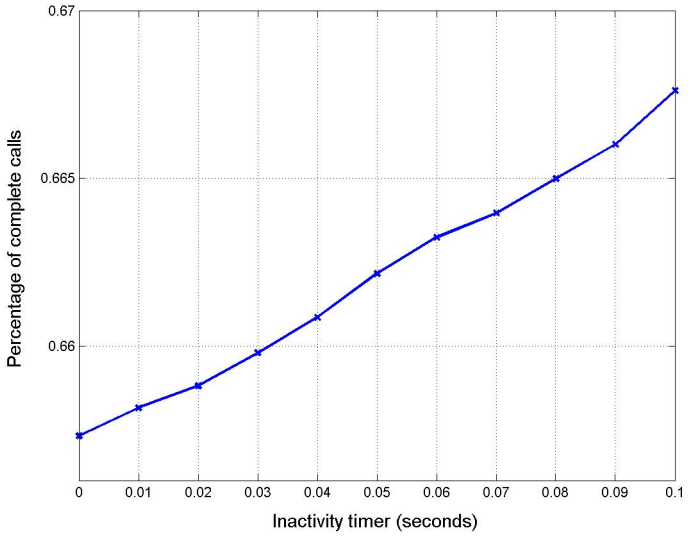

Figure 7.10: Impact of inactivity timer on the percentage of complete calls in a Modified gaming session.

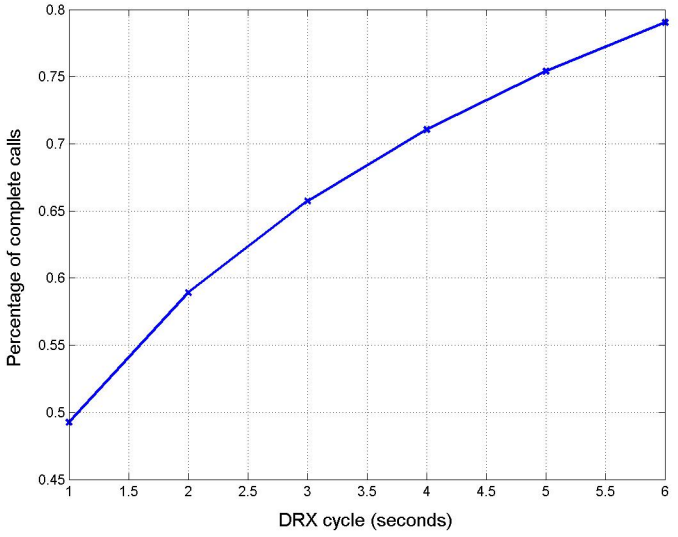

Figure 7.11: Impact of DRX cycle on the percentage of complete calls in a Modified gaming session.

of complete calls should be satisfied. Both inactivity timer and DRX cycle parameters are needed. We suggest selecting the inactivity timer to satisfy the percent of complete calls first, then choosing the DRX cycle to meet the requirement of delay. Here, the inactivity timer also help reduce delay.

7.4 Summary of the impact of power saving parameters

The DRX cycle saves energy, increases delay, has some impacts on the continuity of packet calls in a Modified gaming session. Also, the energy consumption almost keeps the same even when a DRX cycle changes.

The inactivity timer reduces delay, increases the continuity of a packet call in a modified gaming session and energy consumption.

The inactivity timer is not needed in an FTP application because of long reading time and the whole file, even it is fragmented for transmission. In terms of Modified gaming applications, an appropriate inactivity timer can maintain the continuity of bursty data and reduce packet delay. With the main goal of saving energy, DRX cycle is needed in both FTP and Modified gaming applications.

7.5 Adaptive power saving mechanism

A UE uses different applications from time to time. Each application demands different optimal values for inactivity timer and DRX cycle. Therefore, we suggest adaptively deciding the values of inactivity timer and DRX cycle according to real time measurement.

The measurement can be operated by monitoring the buffer size of UE and RNC. The increasing of buffer size indicates the arriving of new traffic. We can record the

traffic arriving moments to calculate interarrival time between two arriving traffic. The average interarrival time between bursty data can be used to decide DRX cycle. The average interarrival time between packets within a burst can be used to decide the inactivity timer. Generally, a DRX cycle is much larger than an inactivity timer in terms of time scale. By monitoring the buffer size, the DRX cycle and inactivity timer can be adjusted to acquire optimal performance for the system.

Chapter 8

Conclusion and Perspective

Energy conservation is one of the important issues in wireless networks. We study the power management in WLAN and UMTS systems.

In WLANs, we model the power management scheme as an M/G/1 queue with bulk service and a D/G/1 queue first. Using the analytical model, we calculate the delay and energy efficiency performance metrics. Then, we simulate the power management scheme. The simulation results are in good agreement with the analytical results. Based on both theoretical analysis and simulation, we give suggestion about how to optimally configure power management parameters. Also, we propose the enhanced power management mechanisms for both infrastructure and independent mode network. The performance of enhanced schemes outperform the original one by simulation.

UMTS uses DRX mechanism and inactivity timer to save energy. DRX scheme allows a UE to turn off its receiver when it is idle and wake up periodically to listen to the paging channel. Inactivity timer maintains the continuity of bursty data and helps reduce delay. We use traffic models specified in 3GPP as input to investigate the impact of DRX cycle and inactivity timer on the delay and energy efficiency performance. By simulation, we find that the optimal values for different traffic types are dramatically different. Therefore, we propose to adaptively configure DRX cycle and inactivity timer according to real time measurement of buffer size.

With the development of wireless communications, inter-networking between UMTS and WLANs becomes a trend [51] [52]. Both systems can provide high data rate and multiple services. UMTS supports mobility better and WLANs can serve as hot spots to relieve possible traffic congestion of UMTS.

When a user switches between UMTS and WLANs, the power management scheme and corresponding parameters need to be updated accordingly. Since the basic idea of power saving is the same for both UMTS and WLANs, which puts a user equipment into the low power consumption state when it is idle and wakes it up periodically. It is possible to achieve seamless transition for power management between UMTS and WLANs. The seamless transition will be a challenging topic and attract lots of interests.

Bibliography

[1] P. Lettieri and M. Srivastava, "Advances in wireless terminals," *IEEE Personal Communications*, vol. 6(1), pp. 6–18, 1999.

[2] M. Miles, "Recent advances in lithium battery technology," in *Proc. of Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, 2001, pp. 219–222.

[3] T. Starner, "Powerful change part 1: batteries and possible alternatives for the mobile market," *IEEE Pervasive Computing*, vol. 2(4), pp. 86–88, 2003.

[4] R. Balan, "Powerful change part 2: reducing the power demands of mobile devices," *IEEE Pervasive Computing*, vol. 3(2), pp. 71–73, 2004.

[5] K. Green and J. Wilson, "Future power sources for mobile communications," *Electronics and Communication Engineering Journal*, vol. 13(1), pp. 43–47, Feb. 2001.

[6] J. Lorch and A. J. Smith, "Reducing processor power consumption by improving processor time management in a single-user operating system," in *Proc. of International conference on Mobile Computing and Networking*, 1996.

[7] D. Bertozzi, L. Benini, and B. Ricco, "Power aware network interface management for streaming multimedia," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2002, pp. 17–21.

[8] D. Chan, T. Berger, and R. Bridgelall, "Energy efficiency of csma protocols for wireless packet switched networks," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2004, pp. 447 – 452.

[9] C. Chiasserini, P. Nuggehalli, and V. Srinivasan, "Energy-efficient communication protocols," in *Proc. of the 39th conference on Design automation*, 2002, pp. 824–830.

[10] C. Song, S. Turner, and H. Sharif, "An energy-efficient tcp quick timeout scheme for wireless lans," in *Proc. of Performance, Computing, and Communications Conference*, 2003, pp. 193 – 197.

[11] C. Takahashi, "Application aware scheduling for power management on ieee 802.11," in *Proc. of Performance, computing, and communication conference*, 2000, pp. 247–253.

[12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1999.

[13] H. Woesner, J.-P. Ebert, M. Schlager, and A. Wolisz, "Power-saving mechanisms in emerging standards for wireless LANs: The MAC level perspective," *IEEE Personal Communications*, vol. 5, pp. 40–48, June 1998.

[14] J. A. Stine and G. Veciana, "Improving energy efficiency of centrally controlled wireless data networks," *Wireless Networks*, vol. 8, pp. 681–700, Nov. 2002.

[15] L. Bononi, M. Conti, and L. Donatiello, "A distributed contention control mechanism for power saving in random-access ad-hoc wireless local area networks," in *Proc. of IEEE International Workshop on Multimedia Communication (MoMuC'99)*, San Diego, USA, Nov. 1999, pp. 114–123.

[16] E. Jung and N. H. Vaidya, "An energy efficient mac protocol for wireless lans," in *Proc. of INFOCOM 2002*, June 2002, pp. 1756–1764.

[17] M. Liu and M. T. Liu, "A power-saving scheduling for ieee 802.11 mobile ad hoc network," in *Proc. of the 2003 International Conference on Computer Networks and Mobile Computing, ICCNMC 2003*, Oct. 2003, pp. 238–245.

[18] S. Kwon, Y. Chung, and D. Sung, "Queueing model of sleep-mode operation in cellular digital packet data," *IEEE Transaction on Vehicular technology*, vol. 52(4), pp. 1158–1162, Jul 2003.

[19] Y. Lin and Y. Chuang, "Modeling the sleep mode for celluar digital packet data," *IEEE Communication letters*, vol. 3(3), pp. 63–65, Mar 1999.

[20] J. So and D. Cho, "On effect of timer object for sleep mode operation in cdma2000 system," in *Proc. of IEEE ICC*, June 2000, pp. 555–559.

[21] *Radio Resource Control (RRC) protocol specification*, 3GPP Std. TS 25.331 V6.4.0, 2004.

[22] J. Yeh, C. Lee, and J. Chen, "Performance analysis of energy consumption in 3gpp networks," in *Proc. of Wireless Telecommunications symposium*, May 2004, pp. 67–72.

[23] M. Chuah, W. Luo, and X. Zhang, "Impact of inactivity timer values on umts system capacity," in *Proc. WCNC*, Mar. 2002, pp. 897–902.

[24] C. Lee, J.Yeh, and J. Chen, "Impact of inactivity timer on energy consumption in wcdma and cdma2000," in *Proc. of Wireless Telecommunications Symposium*, May 2004, pp. 15–24.

[25] S. Yang and Y. Lin, "Modeling umts discontinuous reception mechanism," *IEEE Transaction on Wireless Communications*, vol. 4(1), pp. 312–319, Jan 2005.

[26] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen, "A survey of energy efficient network protocols for wireless networks," *Wireless Networks*, vol. 7, pp. 343–358, 2001.

[27] K. Sivalingam, J. chen, P. Agrawal, and M. Srivastava, "Designand analysis of low-power access protocols for wireless and mobile atm networks," *Wireless Networks*, vol. 6(1), pp. 73–78, 2000.

[28] S. Singh and C. Raghavendra, "Pamas: Power aware multi-access protocol with signalling for ad hoc networks," *Computer Communication Review*, vol. 28(3), pp. 5–26, 1998.

[29] M. Woo, S. Singh, and C. Raghavendra, "Power aware routing in mobile ad hoc networks," in *Proc. of ACM MobiCom*, 1998, pp. 181–190.

[30] v. Tiwari, S. Malik, and A. Wolfe, "Power analysis of embedded software: A first step towards software power minimization," *IEEE Transaction on Very Large Integration*, vol. 2(4), pp. 437–445, 1994.

[31] B. Bing, Ed., *Wireless Local Area Networks*. John Wiley and Sons, Ltd, 2002.

[32] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.

[33] T. S. Rappaport, *Wireless Communications, Principles and Practice*. Prentice Hall, 1996.

[34] K. Etemad, *CDMA2000 Evolution*. John Wiley and Sons, Ltd, 2004.

[35] D. Ferrari, "A scheme for real-time channel establishment in wide-area networks," *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 619–629, April 1990.

[36] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on Selected Area in Communications*, vol. 18(3), pp. 535–547, Mar. 2000.

[37] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the ieee 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Transactions on Networking*, vol. 8(6), pp. 785–799, Dec. 2000.

[38] Y. Tay and K. Chua, "A capacity analysis for the ieee 802.11 mac protocol," *Wireless Networks*, vol. 7, pp. 159–171, 2001.

[39] L. Kleinrock, *Queueing Systems, vol.1.* John Wiley and Sons, 1975.

[40] J. Medhi, *Stochastic Models in Queueing Theory*, 2nd ed. Academic Press, 2003.

[41] G. Bolch, s. Greiner, H. d. Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains, Modeling and Performance Evaluation with Computer Science Applications.* John Wiley and Sons, 1998.

[42] A. A. Aaby, *The Logical Foundations of Computer Science and Mathematics.* Lecture Notes, Walla Walla College, 2004.

[43] K. Lin and Y. Juang, "Extensions of rouche's theorem," *Journal of Control Systems and Technology*, vol. 4(2), pp. 127–131, June 1996.

[44] R. Nelson, *Probability, Stochastic Process, and Queueing Theory, The Mathematics of Computer Performance Modeling.* Springer, 1995.

[45] L. D. Servi, "D/G/1 queues with vacations," *Operation Research*, vol. 34, pp. 619–629, Jul/Aug 1986.

[46] W. D. Kelton, R. P. Sadowski, and D. T. Sturrock, *Simulation with Arena*, 3rd ed. New York: McGraw-Hill, 2003.

[47] H. Holma and A. Toskala, *WCDMA for UMTS*, 3rd ed. John Wiley and Sons, Ltd, 2004.

[48] M. Devetsikiotis and N. Fonseca, "Network taffic modeling," *The Wiley Encyclopedia of Telecommunications*, 2003.

[49] M. Schwartz, Ed., *Broadband Integrated Networks.* Prentice Hall, 1996.

[50] *Feasibility study for enhanced uplink for UTRA FDD (Release 6)*, 3GPP Std. TS 25.896 V6.0.0, 2004.

[51] I. F. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-ip-based wireless systems," *IEEE Wireless comunications*, August 2004.

[52] M. Lott, M. Siebert, S. Bonjour, D. V. Hugo, and M. weckerle, "Interworking of wlan and 3g systems," *IEE Comunications*, October 2004.

Appendix A

Acronyms

3GPP Third Generation Partnership Project

ACK Acknowledgement

AP Access Point

ARQ Automatic Repeat reQuest

AS Active State

ATIM Ad hoc Traffic Indication Message

BSS Basic Service Set

CCA Clear Channel Assessment

CDPD Cellular Digital Packet Data

CHS Control Hold State

CN Core Network

CSMA/CA Carrier-Sense Multiple Access with Collision Avoidance

CTS Clear To Send

DCC Distributed Contention Controller

DCF Distributed Coordination Function

DIFS DCF interframe space

DRX Discontinuous Reception

DS Dormant State

DS Distributed System

DSSS Direct Sequence Spread Spectrum

ECMAC Energy Conserving Medium Access Control

EDD Earliest Due Date

EDGE Enhanced Data Rates for GSM Evolution

EIFS Extended interframe space

ESS Extended Service Set

FCFS First Come First Serve

FEC Forward Error Correction

FHSS Frequency Hopping Spread Spectrum

FIFO First In First Out

FRT Frame Response Time

FSM Frame Synchronization Message

FTP File Transfer Protocol

GPS Generalized Processor Sharing

GPRS General Packet Radio Service

IFS InterFrame Space

IMT International Mobile Telecommunication

IR InfraRed

ISM Industrial Scientific and Medical

ITU International Telecommunication Union

LLC Logic Link Control

MAC Media Access Layer

OFDM Orthogonal Frequency Division Multiplexing

OSI Open System Interconnection

PAMAS Power Aware Multi-AcceSs

PC Pointer Coordination

PCF Point Coordination Function

PHY Physical

PIFS PCF interframe space

PTD Percentage of Time a station stays in the Doze state

QoS Quality of Service

RNC Radio Network Controller

RR Round Robin

RRC Radio Resource Control

RRM Radio Resource Management

RTS	Request To Send
SIFS	Short interframe space
SS	Suspended State
TBTT	Target Beacon Transmission Time
TCP	Transmission Control Protocol
TIM	Traffic Indication Map
TEI	Temporary Equipment Identifier
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network