The Hong Kong Polytechnic University
Department of Computing

Energy Efficiency in IEEE 802.11 Wireless Networks

by

Yi Xie

A thesis submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

August 2007

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ _____(Signature)

_____(Name of Student)

To my dear Parents and Aunt.

Abstract

Energy Efficiency in IEEE 802.11 Wireless Networks

by

Yi Xie

To achieve energy efficiency, various existing mechanisms investigate the tradeoff between communication quality and energy consumption. This thesis proposes two new schemes based on Transmission Power Control (TPC) and Power Saving Mode (PSM). Both of them are Access Point (AP) centric and support multiple clients in an IEEE 802.11 infrastructure network. The new TPC-based scheme uses a polling-based MAC, such as the point coordination function (PCF) in IEEE 802.11, to optimize energy efficiency of all clients. The new PSM-scheme, on the other hand, achieves the same goal by using distribution coordination function (DCF).

The problem of optimizing energy efficiency in a polling-based network is to determine optimal transmission power allocations (therefore the transmission rates) for all clients. The AP notifies the optimal transmission power to all clients at the beginning of each polling cycle. Using this optimal AP-centric TPC, the total energy consumed by all clients in a polling cycle will be minimized when their buffer queues are guaranteed stable. We have solved this problem by first formulating it as a stability-constrained optimization problem and then solving it using an iterative algorithm. The problem is tackled by using two polling schedules: Phase Grouping (PG) and Mobile Grouping (MG). Our extensive experiment results have shown that the optimal power allocations can improve the energy efficiency of a random power allocation by saving energy as much as

four times. After using the AP-centric TPC, the optimal MG scheme is more energy efficient than the optimal PG scheme, because each client in the PG scheme spends more energy during the reception period. We have also shown that there are incentives for the clients to adopt the optimal power allocations.

The problem of optimizing energy efficiency in a DCF network using PSM is to determine the optimal PSM parameters, e.g., beacon interval (BI), listen interval (LI) and minimal contention window (CW) for each client. The centralized PSM (C-PSM) deploys these optimal parameters on the AP and clients. The BI and LIs are optimized to reduce major energy consumption due to unnecessary wake-ups and channel contention. Both simultaneous wake-ups and infrequent wake-ups will cause channel contention during which the contending clients waste much of their energy in the idle mode. In addition, we have proposed a first-wake-up schedule to further reduce simultaneous wake-ups. The CWs, on the other hand, are tuned to balance the clients' access probabilities, such that a shorter backoff period is assigned to a client waking up with a lower frequency. Our extensive simulation results have shown that compared with the standard PSM C-PSM reduces power consumption by up to 76%, increases energy efficiency as high as 320% and shortens AP buffering delay as much as 88%. Moreover, the first-wake-up schedule saves the energy efficiency by up to 22%. C-PSM's advantage also increases with the number of clients in the network.

Acknowledgements

I would like to express my gratitude to my supervisor Rocky K. C. Chang for his continuous support, encouragement and guidance. His unique combination of vision, technical knowledge and kindness has been an inspiring role model in my life. Moreover, I thank Prof. Chan, Prof. Misic, Prof. Muppala and Prof. Chang for their insightful comments and being helpful members of my committee. I thank all other faculty who have been my teachers and mentors over these years.

I would like to thank Xiapu Luo for his collaboration in my research work. Also, Sam Lam, Samantha Lo, Edmond Chan, Steve Poon, Yu Le, Kathy Tang and Wenchao Zhao gave me their constructive suggestions and a lot of support. My heartfelt thanks to them all. Besides, I wish to express thanks to Jing Zhou, Suyun Zhao, Xinli Geng, Yuan Zheng, Jin Yang, Weigang Wu, Wing Ng, Jiaodo Li, Peng Huo and Hui Cheng for their kind help.

I would like to give my special thanks to my family for their unconditional love and support.

The research conducted in this thesis was partially supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 5146/01E).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Wireless networks have been widely used because of its flexibility and mobility. An increasing number of integrated services, such as multimedia, real-time applications, as well as monitor and control systems either have been or will be deployed. These integrated services generally require a high capacity of computation and communication, as well as a good support for energy-constrained wireless devices. Therefore, it is important that the protocols (or architectures) designed for wireless network should optimize energy consumption and satisfy the required quality of service.

A wireless device (client) takes on communication and computational workloads with limited energy. The client's main sources of energy consumption include microprocessor, display device, hard disk, system memory and wireless network interface (WNIC) [93]. There are two ways to save energy. The first is to reduce energy consumed on computation and management using new hardware designs, such as the variable block speed CPUs [69], the blocking-aware processor voltage scheduling [182], the flash memory [120] and the disk spindown [56]. The other is to optimize the usage of energy on WNIC (i.e., the operation of

wireless transceiver) for a given communication task [124]. It is vital to reduce energy consumption on communication, because WNIC consumes up to 50% of the total energy storage in a wireless client. Therefore, the problem of efficiently consuming energy on communication is an increasing concern.

All wireless networks encounter this energy efficiency problem. According to the statement [124] "the energy-intensive nature of wireless communication motivates us to reduce the energy consumed by incorporating energy conservation strategies with network protocols", we focus on the trade-off between energy consumption and communication performance. The design of wireless communication protocols should consider the trade-offs between energy consumption and various performance metrics, such as throughput [181], system capacity [148], delay [83, 155, 131, 132], network utility [46] and error rate [146].

This thesis concentrates on the problem of energy efficiency in an IEEE 802.11 infrastructure wireless network. We explore the function of *access point* (AP) to improve energy efficiency for all wireless clients. We have designed a new AP-centric *transmission power control* (TPC) for polling-based *media access control* (MAC) protocol. This TPC minimizes the sum of clients' energy consumption subject to network stability. We have also designed a centralized *power saving mode* (PSM) to improve clients' energy efficiency. This PSM is an AP-centric deployment of the IEEE 802.11 PSM where the AP determines the optimal parameters for itself and all PSM-enabled clients.

We discuss the technical challenges of implementing energy efficiency in IEEE 802.11 wireless network in Section 1.1 and then present the motivations to design two AP-centric energy-efficient schemes in Section 1.2. We give an overview of our main contributions in Section 1.3 and the outline of this thesis in Section 1.4.

1.1 The Energy Efficiency Problem in IEEE 802.11 Wireless Networks

There are two different ways to configuring IEEE 802.11 wireless networks: There are two different ways to configure IEEE 802.11 wireless networks: infrastructure (centralized) and ad-hoc (distributed). Infrastructure wireless network requires a fixed infrastructure to relay data among wireless clients. One successful example is a mobile phone network. A mobile phone is a portable telephone which receives and makes calls through a cell site (i.e., a *base station*, BS). Another example is a *Wireless Local Area Network* (WLAN). With the development of the IEEE 802.11 standards [86], an increasing number of Wi-Fi hotspots are installed in libraries, classrooms, hotels, shops, airports and so on. In these applications, wireless clients (e.g., laptops, PDAs and mobile phones) using the IEEE 802.11 infrastructure mode connect to the Internet through an *access point* (AP). On the other hand, the IEEE 802.11 ad hoc mode does not require a fixed network infrastructure, such as the AP and the wireless clients route packets among themselves. Their typical applications include military usages, environmental sensor networks, car-based ad hoc networks and biomedical networks.

The energy efficiency problem has become a major problem to tackle in wireless networks. A wireless client is required to take on heavy workloads; but its energy storage is always limited, and nearly half is consumed on communication. In order to increase clients' battery life time, it is vital to reduce energy which is consumed on communication. Therefore, the wireless network protocols should consider the issues of energy saving and communication performance.

It is challenging to design an energy-efficient wireless network. In order to implement energy-saving schemes, energy-saving technologies on each layer must

take performance metrics into account, such as delay, throughput and network utility. For example, the energy-saving design on physical layer tries to support the CMOS circuit by lower power while satisfying the requirements of transmission rates (i.e., throughput). When wireless clients are allowed to sleep with very low power, the MAC protocol should let clients sleep as long as possible while satisfying the requirements of packet delay or throughput. Reducing transmission power saves energy consumed on the packet transmission at the expense of increasing the transmission time [131]. The adjustment of transmission power to a proper level can save energy while satisfying the requirements of packet delay. Packet retransmissions due to transmission error or packet loss consume much energy. The energy-saving designs on logical link sublayer and transport layer should reduce transmission errors and packet loss. In order to increase the life time of wireless network, energy-aware routing protocols on the network layer may choose the nonshortest path to bypass the poorly powered wireless clients.

Moreover, reducing energy consumption in wireless networks requires the cooperation of multiple network layers. The energy-saving design on one layer often influences wireless client's performance and energy consumption on other layers. Some hardware techniques in physical layer (PHY layer) are designed to reduce the power consumption of wireless clients [84]. They operate well only if there are suitable control mechanisms on the upper layers. For example, the dynamic power management in WNIC provides variable radio frequency (RF) power. It is helpful to save energy, reduce channel interference and increase throughput when the *transmission power control* (TPC) on MAC layer is adopted. Some WNICs' circuit supports a sleeping mode. The sleeping client can save energy, because the power consumed in sleeping mode is an order of magnitude less than that in an active mode. Yet the sleeping client cannot receive or send any packet.

An unoptimized sleeping period will result in interrupted transmissions, packet loss, long packet delay and low throughput. In order to use the sleeping mode efficiently, MAC protocol should adopt a special mechanism to control the mode transition of clients, for example, the *power saving mode* (PSM) in IEEE 802.11 standard.

In an infrastructure wireless network, the AP offers flexibility for designing energy-efficient protocols. There are two main reasons for this. First, the AP is usually not energy-constrainted as the clients do and has more computation power. Many energy-efficient protocols reduce energy consumption of wireless clients by letting the AP take over some tasks for clients and operate some scheduling algorithms. Second, wireless clients only need to transmit data to the AP or receive data from the AP, which are called uplink and downlink respectively. In contrast, ad hoc network cannot benefit from any fixed infrastructure. Wireless clients have to take up additional tasks such as forwarding packets and managing routing tables.

Some energy-efficient schemes [105, 147, 127, 167] in infrastructure wireless networks are distributed in each client (i.e., user-centric schemes). In order to save energy, the client individually determines its settings and notifies the AP of them. This kind of energy-efficient schemes could adapt to traffic variance, reduce or even minimize the client's energy consumption while satisfying its performance requirement. Their disadvantages include the increased workload of wireless client and the possible degradation of other clients in terms of energy saving and communication performance. In contrast, some energy-efficient schemes are regarded to be AP-centric (e.g., [111, 78, 116]), since the AP determines the operation or important setting of all wireless clients in the network. In stead of improving one client's energy efficiency, the AP-centric schemes usually optimize

energy consumption of all clients while satisfying their performance requirements in an integrated way. Two AP-centric approaches proposed in this thesis concentrate on achieving the energy efficiency of clients in infrastructure wireless network while exploring the advantage of the AP. The AP has a global knowledge of clients and takes on most tasks to implement the energy efficiency of all wireless clients.

1.2 Motivations for the AP-centric Approach

We aim at increasing energy efficiency for all wireless clients in an infrastructure wireless network. We consider the balance of energy consumption and communication performance of clients in an integrated way, instead of designing an energy-efficient distributed scheme for each client. Here, we assume that the positions of wireless clients are relatively fixed in the network. Therefore, the energy-saving designs (e.g., [54, 35]) based on the mobility model of clients are not considered.

Two new energy-efficient schemes proposed in this thesis are deployed in an AP. Since the AP relays packets for all wireless clients, it is able to obtain the channel condition and the traffic of each client. Therefore, the AP could act as a control knob to achieve energy efficiency for all clients. In the new AP-centric TPC, the AP determines the optimal transmission power allocation of clients in polling-based MAC protocol. It minimizes the total energy consumption of clients in a stable wireless network. In the centralized PSM, the AP determines the optimal parameters in the IEEE802.11 DCF and PSM for itself and all PSM-enabled clients. It improves the energy efficiency of all clients in an infrastructure wireless network.

1.2.1 AP-centric TPC for polling-based MAC

Medium Access Control (MAC) protocols [36] control the access to a shared medium by defining rules that allow wireless clients to communicate with each other in an efficient manner. The design of MAC protocol is critical for communication performance and energy consumption of wireless clients. In general, an infrastructure wireless network employs one of centralized MAC protocols. A typical example is polling-based MAC protocol. The AP sends and receives data by polling clients in a fixed order. Therefore, we need to design a polling-based MAC protocol for the network in which an AP serves as the relay for a number of wireless clients.

Previous studies considered a tradeoff between energy consumption and communication quality on MAC protocol [44]. Most of them designed an MAC mechanism to minimize energy consumption of one or more wireless clients taking into account specific performance metrics (e.g., delay, throughput and network utility). However, they did not consider the influence of one client's energy-saving action on other clients. In order to address the energy efficiency problem of all wireless clients, a network model and a fundamental metric, stability, are necessary. The analytical model is a queueing system where each client is a queue. It is possible to study the energy consumption and performance of each queue and the interactions of queues. By stability for a network, all clients are stable when the number of buffered packets in each client does not grow up unboundedly.

The PSM and TPC are our main approaches to achieve energy efficiency of all wireless clients in the AP-centric TPC for polling-based MAC protocol. PSM reduces the energy consumption of wireless clients, because it lets clients sleep with very low power consumption. When a client is polled to communicate with

the AP, it is active. When a client is not polled, it is allowed to enter into the sleeping mode to save energy. TPC enables wireless clients to consume less energy on transmission by using a lower transmission power. However, the lower transmission power results in a longer transmission time. If the transmission power is too low, the transmission period will prolong, thus increasing the clients energy consumption. In a sense, the much low transmission power degrades the stability by increasing packet delay without saving energy. Therefore, there is a trade-off among transmission power, energy consumption of clients and network stability.

It is easy to see that the proper reduction of transmission power and the increase of sleeping period are both helpful to save energy. However, these two issues conflict with each other. If the transmission power is too high, much energy is consumed on transmission. If the transmission power is too low, the extended transmission period will prevent a client from sleeping, thus increasing the energy consumption. Therefore, the transmission power is a control knob to minimize energy consumption. We should use TPC to reduce the transmission power of clients within an appropriate range which is effective to save energy while guaranteeing the stability of network.

We use stability as an important constraint to minimize total energy consumption of wireless clients, because it is a fundamental requirement of all networks. An energy-efficient mechanism will not be useful if it unstabilizes one or more clients. However, stability is seldom considered in the studies of energy efficiency problem. We only find three examples. One was the resource allocation policy [174], which minimizes the total energy consumption while satisfying the network stability and the QoS requirements of link. It is suitable for wireless sensor networks instead of wireless infrastructure networks. Another two examples

are [151] and [152]. The authors investigated the effect of their energy-efficient mechanisms on the stability region, but they did not guarantee network stability.

Therefore, we formulate an optimization problem about the transmission power of clients which minimizes the total energy consumption of all wireless clients while guaranteeing that each stable client handles packets with a limited delay. The optimization problem is based on the analysis of queueing systems which models polling-based MAC protocols. The AP solves this optimization problem and obtains an optimal transmission power allocation of all clients by a proposed interactive algorithm. Each wireless client simply deploys its own optimal transmission power. The optimal TPC is mainly implemented in the AP and denoted as the AP-centric TPC in polling-based MAC protocols.

As we shall see in this thesis, the energy-efficient polling-based MAC protocol can achieve energy efficiency by optimizing energy consumption of all PSM-enabled clients when two polling schemes are considered: *phase grouping* (PG) and *mobile grouping* (MG). The PG schemes have been widely deployed. The typical example is the energy conserving MAC (EC-MAC) protocol [44] which uses an explicit transmission order with reservations. In the broadcast (downlink) phase, the wireless device listens to the downlink for the transmission order. Many MG polling-based protocols have also been proposed to reduce mode transitions and control packet transmissions [73], such as the disposable token MAC Protocol (DTMP) [72], the E^2MaC protocol [75] and the mobile grouping schedule [77].

In summary, we have designed the AP-centric TPC for polling-based MAC protocols which provides a stable network and consumes energy efficiently. It is a new way to achieve a trade-off between energy consumption and stability. When wireless applications require specific performance requirements (e.g., short

delay or large throughput), other energy-efficient schemes may be applied as a complement to this design. However, the optimal transmission power allocation is obtained under an assumption that the traffic of each wireless client is arrived according to Poisson distributions. We should explore our energy-efficient MAC protocol to support a general wireless network with different traffic distributions.

1.2.2 Centralized PSM

In this thesis, we concentrate on the PSM for IEEE 802.11 infrastructure networks and will show that significant improvement can be obtained from a new deployment of the PSM.

The IEEE 802.11 PSM is one main mechanism which could save energy. The wireless clients could stay in an active or dozing state. The active client consumes high power when it transmits data, receives data and keeps idle. However, the dozing client operates with much lower power by entering a sleep mode. Using IEEE 802.11 PSM, the client goes into the sleep mode whenever it is idle. The client also wakes up periodically to check whether they have frames buffered at the access point (AP) through its periodic transmissions of beacon frames. The AP broadcasts a beacon frame every *beacon interval* (BI), and each client's wake-up frequency is determined by a PSM parameter *listen interval* (LI). Both BI and LI are configurable, and their settings can directly influence the PSM's performance. However, the PSM does not prescribe how the BI and LI should be configured; therefore, default values are often used. Hereafter, we use PSM to refer to the IEEE 802.11 PSM.

Obviously, the PSM cannot adapt to the traffic and configuration dynamics inherent in typical wireless networks. For example, the studies [127, 13] have

shown that the typical BI of 100ms cannot save significant energy in many practical scenarios. Worse yet, the PSM is reported to have an adverse impact on application performance, such as short TCP connections [105].

To address these shortcomings, a number of new power-saving schemes based on putting idle clients into sleep have been proposed. A class of them (e.g., [127, 147, 105]) enables each client to save energy by reducing the number of unnecessary wake-ups, i.e., designing an optimal wake-up schedule. These user-centric schemes, however, do not address energy consumption from channel contention which, as we will show, is another major source of energy wastage. Additionally, these user-centric schemes should be deployed in each client, which increases the workload of clients and the complication of MAC design.

Another class is based on an AP-centric approach which relies on the AP to improve the energy efficiency of all clients in the network. The AP is a control knob to determine the operations of wireless clients based on the global knowledge of channel condition and the traffic of clients. Each client does not do anything but follows the proposed scheme directly. One approach is for the AP to design a packet transmission schedule to minimize channel contention (e.g., [116, 78]). Another approach is for the AP to redesign beacon frame and poll clients one by one, which totally avoids channel contention [111]. Most AP-centric schemes are not compatible with the standard PSM scheme or difficult to implement because of the precise transmission schedule.

In this thesis, we propose a *centralized PSM*, or C-PSM, which deploys the PSM using the AP-centric approach. The AP in C-PSM chooses the best BI and LIs for all clients based on the frame patterns arriving at the AP. These intervals are chosen to reduce energy consumption due to both unnecessary wake-ups and channel contentions. The energy wasted in channel contentions could be very

significant, because the clients involved cannot go to sleep throughout the contention period which could be very long. To further reduce energy consumption, C-PSM includes an optional first-wake-up schedule to reduce the simultaneous wake-ups of clients. When frame collisions occur, the AP assigns congestion windows to the clients involved, such that a client that wakes up less frequently will be able to retransmit earlier.

C-PSM is different from other AP-centric schemes in three important aspects. First, C-PSM conforms to the PSM, whereas other AP-centric schemes, such as [20], do not. The only additional mechanism required for C-PSM is to notify the clients of their LIs which could be accomplished through the beacon transmission channel. Second, C-PSM does not rely on computational-expensive frame scheduling which is employed in [78, 116, 111]. Instead, the AP in C-PSM simply observes the statistics of the frame arrival patterns. Third, C-PSM is designed independent of the upper-layer protocols. Therefore, it could be used for any mix of network protocols. However, other AP-centric schemes, such as [12], are designed only for TCP traffic.

In summary, C-PSM improves the energy efficiency of all PSM-enabled clients by reducing energy consumed on unnecessary wake-ups and channel contention. It is accomplished by the AP which completely determines the optimal beacon interval, listen intervals, contention windows, as well as schedule of the clients' first wake-up times. As a result, C-PSM is compatible with the PSM except that the parameter setting of each client is notified by the AP. In order to use C-PSM, we should know or measure the distributions of inter-frame arrival times. Future works include extending the system model to data sources with feedback channels, such as TCP. Moreover, we will consider using path measurement to aid the AP's determination of the optimal parameters.

1.3 Main Contributions

This thesis focuses on how to achieve the energy efficiency of all wireless clients in an infrastructure wireless network. The main contributions include the following:

Firstly, we have proposed an AP-centric TPC by solving optimization problems of energy efficiency in an infrastructure wireless network using the polling-based MAC protocol. PSM and TPC have been successfully combined to conserve energy. A clients actions, such as reducing the transmission power and going to sleeping mode, greatly affect other clients in terms of energy consumption and communication performance. The interactions of clients are complicated. Therefore, we have formulated the optimization problem to minimize the total energy consumption of all wireless clients in a polling cycle. At the same time, all clients remain stable. The stability constraints are derived from a queueing analysis. In this situation, the queue lengths would not go unboundedly. We have designed an iterative algorithm to solve the stability-constrained optimization problems. The optimal solution is the optimal transmission power allocation for all wireless clients when two polling schemes—PG and MG are considered. The simulation results show that:

* The AP-centric TPC uses the optimal transmission power allocation to minimize the total energy consumption of clients while guaranteeing network stability.

* The MG scheme with the AP-centric TPC (i.e., the optimal MG scheme) is more energy efficient than the PG scheme with the AP-centric TPC (i.e., the optimal PG scheme).

* The adverse effect of the AP-centric TPC on delay is not significant. The balance between energy consumption and delay is considered to be beneficial as a whole.

* There are incentives for wireless clients to adopt the optimal power allocations. An uncooperative client, which adopts a random transmission power instead of the optimal transmission power, does not benefit from his self-governed action. If the transmission power of the uncooperative client is lower than the optimal one, no client reduces energy consumption or shortens delay. Otherwise, all clients shorten delay at the cost of higher energy consumption.

Secondly, we have performed extensive simulations to analyze the performance of the standard PSM in an infrastructure wireless network. In this network, one or more PSM-enabled clients receive data from the AP, and four different distributions of inter-frame arrival times are considered. We have verified that the adjustment of BI and LIs can improve energy efficiency when it avoids clients' frequent unnecessary wake-ups and alleviate channel contentions. Then the simultaneous wake-ups of clients and channel collisions should be decreased. Motivated by the analysis results, we have proposed the centralized PSM (C-PSM) to increase the energy efficiency of all PSM-enabled clients. The AP in C-PSM employs mechanisms reduce these issues which results in the increase of energy consumption.

C-PSM avoids the frequent unnecessary wake-ups of one client by adapting its LI according to the average inter-frame arrival time of the traffic destined for it. On the other hand, C-PSM reduces the channel contentions by choosing the optimal BI and LIs jointly, such that the simultaneous wake-ups and collisions

of clients are reduced. Moreover, the scheme assigns different CWs to the clients according to their optimal LIs, such that the client with a large LI has a higher probability to retrieve its frames. To further reduce the simultaneous wake-ups, C-PSM could optionally schedule the first-wake-up time of each client, such that two or more clients wake up in one BI with a low probability. And this first-wake-up schedule is effective to improve energy efficiency of all clients when two or more optimal LIs are the same or have the same common factor. The simulation results lend a strong support to the usefulness of C-PSM. Compared with the PSM, C-PSM reduces significantly more energy (up to 76%), achieves higher energy efficiency (up to 320%) and reducing AP buffering delay (up to 88%). Moreover, C-PSM saves more energy when the number of clients increases. C-PSM can obviously increases clients' energy efficiency only when the client's wake-up energy consumption is small (e.g., $\leq 0.0066J$ in our simulation), and the ratio of idle power to sleeping power is large (e.g., $\geq 1000\%$ in our simulation).

Finally, we would like to point out our contributions to the design of simulators. We have designed a simulator using MATLAB [169], which fully implements the 802.11 MAC protocol with PSM in an infrastructure network. It simulates the operations of AP and wireless clients slot by slot, including frame creation, frame exchange, frame collision, power mode transition and energy consumption. Furthermore, this simulator can implement a given PSM scheme if it is compatible with the 802.11 PSM. Accordingly, we can modify some input variables to simulate the proposed scheme, C-PSM.

1.4 Outline of Thesis

The rest of this thesis is organized as follows.

Chapter 2 introduces the background knowledge of IEEE 802.11 wireless network and some energy-saving schemes on it. Section 2.1 talks about the architecture and MAC protocols on IEEE 802.11 network. Section 2.2 discusses the energy-saving schemes proposed in IEEE 802.11 wireless network on different layers. Section 2.3 and Section 2.4 individually review two main energy-saving schemes—PSM and TPC, as well as some energy-efficient schemes based on them.

Chapter 3 proposes a new approach to achieve a tradeoff between energy consumption and stability. In Section 3.1, we describe polling-based MAC protocols and two types of poling schemes-Phase Grouping (PG) and Mobile Grouping (MG). Then they are modeled as cyclic-service queueing systems in Section 3.2. In Section 3.3, we formulate stability-constrained optimization problems for these two schemes, which minimize total energy consumption of all wireless clients in a infrastructure wireless network. Then we present an iterative algorithm to compute the optimal transmission power allocation for the AP-centric TPC in Section 3.4. After that, we evaluate the optimal solutions and compare the optimal PG and MG schemes using the AP-centric TPC in Section 3.5. The effects of the two optimal schemes on delay and energy-delay tradeoff are also evaluated in this section. Furthermore, Section 3.6 studies the effect of an uncooperative client on the network performance when others follow the optimal scheme.

In Chapter 4, a new AP-centric PSM scheme, centralized PSM (C-PSM), is proposed to achieves energy efficiency of wireless clients. Section 4.1 presents the network model, and Section 4.2 introduces how we use MATLAB to simulate this network, in which the AP and the wireless clients operate according to the 802.11 MAC with PSM. Next, Section 4.3 investigates the effects of the beacon interval and listen intervals on several performance metrics, such as power, energy consumption and delay. The simulation results show that the appropriate choices

of BI and LIs are helpful to save energy. Accordingly, Section 4.4 presents the design of a main algorithm and a first-wake-up schedule (FWS) in the AP to obtain the optimal PSM parameters (e.g., BI, LIs, CWs and FWS) according to the traffic patterns. Then, we evaluate C-PSM by four performance indexes based on the standard PSM scheme in Section 4.5. Furthermore, we study the effects of power consumption model on C-PSM's performance in Section 4.6.

Finally, Chapter 5 summarizes the main results and presents several directions on the future works.

Chapter 2

IEEE 802.11 Networks and Energy-saving Schemes

This chapter introduces the background knowledge of IEEE 802.11 wireless networks [86], such as the architecture (infrastructure and ad hoc mode) and *Media Access Control* (MAC) protocols. We also summarize various mechanisms and approaches proposed to increase energy efficiency in IEEE 802.11 wireless networks, including the power saving mode and transmission power control.

2.1 IEEE 802.11 Networks

2.1.1 The Infrastructure Mode and Ad Hoc Mode

The building block of an IEEE 802.11 wireless network is known as the *basic service set* (BSS), which is a group of wireless stations that communicate with each other in a *basic service area* [65]. There are two different ways of configuring a BSS: independent (ad hoc or distributed) and infrastructure (centralized), as

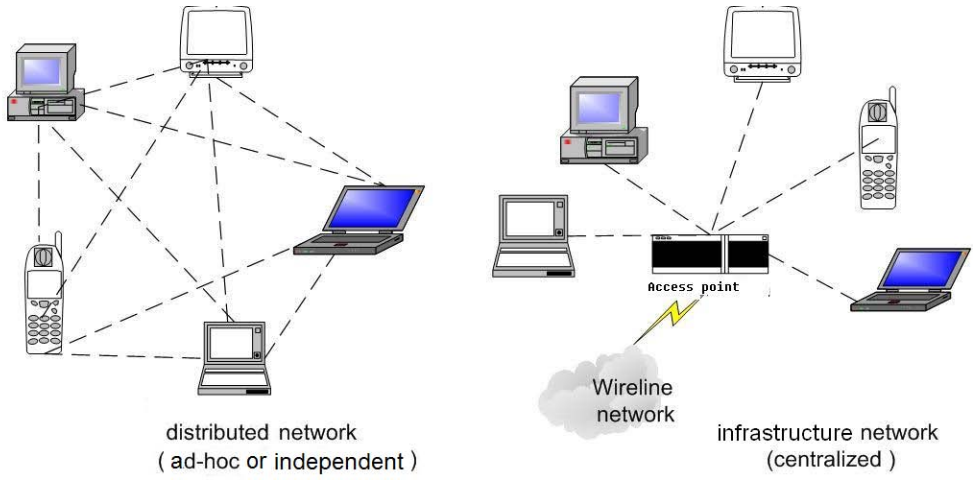


Figure 2.1. The distributed and infrastructure configurations in the IEEE 802.11 wireless networks.

shown in Figure 2.1.

An *ad hoc wireless network* consists of one or more *independent BSS* (IBSS) without any central administration. Wireless stations in an IBSS directly communicate with each other. Several IBSSs may build a network topology, allowing multi-hop connectivity. Each wireless station behaves as a host and router simultaneously. If a receiver is not within a sender's transmission range (i.e., these two stations are located in different IBSSs), one or more stations between them will forward packets to the receiver through a multi-hop path. In this case, each station involved discovers the path to the destination dynamically.

An *infrastructure wireless network*, on the other hand, consists of infrastructure BSSs. It is also known as the last-hop network, centralized network and cellular network. The key characteristic of an infrastructure BSS is the provision of an *access point* (AP). An AP serves as a central unit to relay data from one wireless station/client to another. The wireless client therefore does not maintain any route to its neighbors but communicates with the AP instead. Before

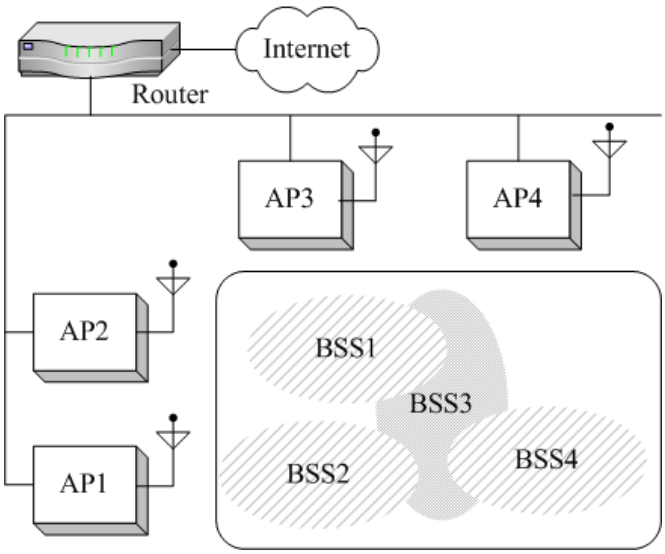


Figure 2.2. An extended service set for the infrastructure networks.

obtaining network services, a client must *associate* with an AP. There are two directions of communications: downlink (from the AP to wireless clients) and uplink (from wireless clients to the AP).

An infrastructure network can also be scaled by configuring multiple BSSs into an *extended service set* (ESS). An ESS is a set of IBSSs and infrastructure BSSs, which may connect to a backbone network through several APs, as depicted in Figure 2.2. All APs in an ESS share the same *service set identifier* (SSID). The wireless stations belonging to an ESS use the SSID as the identity of the network. Moreover, the overlaps among BSSs are inevitable in wireless environment.

The network services provided by the IEEE 802.11 standard are classified into three categories [65]:

1. The station services are provided by all 802.11-compliant devices, including wireless/mobile stations and AP. These services include: (1) *MAC Service Data Unit (MSDU) delivery service* that supports the message delivery

among wireless stations or between wireless station and AP; (2) *authentication service* that establishes associations between sender and receiver; (3) *deauthentication service* that terminates an authenticated relationship and clears the authentication information; and (4) *confidentiality service* that protects messages when they traverse the vulnerable wireless link (e.g., Wired Equivalent Privacy (WEP) protocol, encryption schemes, the user-based authentication and key management of IEEE 802.11i [87]).

2. The spectrum management services enable 802.11-compliant devices to change their radio settings dynamically in response to the wireless environment. The main mechanisms include the *transmission power control* (TPC), *transmission rate adaptation* and *dynamic frequency selection* (DFS). The TPC allows wireless stations and APs to reduce their transmission power which could help decrease channel interference and save energy in the wireless stations. The IEEE 802.11 standard also provides multi-rate capabilities through a rate adaptation algorithm, such as ARF [96], RBAR [82], AARF [110], SampleRate [26], ONOE [2], CARA [99] and the gradient based scheduling [135]. On the other hand, DFS allows wireless devices to dynamically select a good quality communication channel, spreads the load across available channels and ensures coexistence with existing radio systems.

3. The distribution system services are provided only in infrastructure wireless networks, which allow an AP to connect to a distributed system (i.e., the wireless side). The AP provides *distribution service* and *integration service* to extend the services from the wired side to the wireless side. For example, in receiving a frame, the AP uses the distribution service to deliver this frame to its destination within a BSS. In order to manage mobile stations,

the AP maintains associated data and location information of each station. The *association/reassociation service* is used for (re-)establishing the association of one wireless station with an AP. The *disassociation service* is used to terminate an existing association between a wireless station and an AP.

2.1.2 MAC Protocols: DCF and PCF

An MAC protocol [36] defines the rules for wireless clients to share the channel resources. It is composed of two components: (1) access arbitration, which determines when a particular station should transmit data and (2) transmission control, which determines the transmission duration of a particular station and the related parameters, such as the transmission power and transmission rate.

Depending on the network architecture, wireless MAC protocols can be broadly classified into two categories: distributed and centralized MAC protocols. Distributed MAC protocols are a collection of distributed random access and dynamic reservation/collision resolution protocols [106]. ALOHA and slotted ALOHA are classic examples of the distributed random access protocols without carrier sensing [67, 128]. A station in ALOHA networks accesses the channel with a certain probability whenever it is ready. However, as the number of stations increases, considerable collisions will occur.

To reduce collisions, a carrier-sensing component is introduced to detect any ongoing transmissions in the distributed MAC protocols. An example is the *carrier sensing media access* (CSMA) in the IEEE 802.11 standard. When a station detects a busy channel using the carrier sensing, it will defer its transmissions to avoid collisions. However, the location-dependent CSMA suffers from the

problems of hidden stations and exposed stations [65]. To mitigate the adverse impact of these problems, a number of dynamic reservation and collision resolution protocols have been proposed, including the out-of-band approach (e.g., Busy Tone Multiple Access (BTMA) [103] and RI-BTMA [177]), the handshaking approach (e.g., the RTS/CTS control in Multiple Access with Collision Avoidance (MACA) [97], MACAW [24], Floor Acquisition Multiple Access [63]) and a hybrid of them (e.g., Dual BTMA [71]).

Centralized MAC protocols, on the other hand, designate a central entity which is very often known as an AP to manage the channel resource through admission control, bandwidth assignment and channel access control. The central control by the AP is generally more effective to avoid collisions, assign channel resource and satisfy the performance requirements of all wireless clients. Examples include the point coordination function (PCF) in the IEEE 802.11 networks, HIPERLAN/2 of ETSI [92] and numerous wireless asynchronous transfer mode (ATM) proposals [138].

Centralized MAC protocols may use random access mechanisms, guaranteed access mechanisms or their hybrid access mechanisms. Random access mechanisms let the AP instead of wireless client undertake the tasks of carrier sensing and collision detection. The examples include Idle Sense Multiple Access (ISMA) [179], Reservation ISMA [178] and Slotted ISMA [62]. Furthermore, the AP could adopt an efficient algorithm to assign resources or avoid collisions, for example, Resource Auction Multiple Access (RAMA) [9] and Fair-RAMA [139]. Guaranteed access mechanisms guarantee bandwidth or satisfy the quality-of-service (QoS) requirements. In general, the AP lets wireless station access the medium channel in an orderly manner, usually in a round-robin fashion. In this way, the collisions among clients are mostly avoided. *Polling Protocol* [183] [5] [39]

and *token-passing protocol* (e.g., *Disposable Token MAC Protocol* [72] and *ATP-MAP* [42]) are two kinds of implementations.

Other centralized MAC protocols use hybrid access mechanisms. They let clients use a random access protocol to send AP a request for its data exchange, and then the AP will allocate the service time slots for each client (which is more precise than the accessing ordering). The typical examples include Packet Reservation Multiple Access (PRMA) [176], Centralized PRMA [25], Independent Stations Algorithm (RRA-ISA) [29] and Distributed-Queueing Request Update Multiple Access [98].

The Distributed Coordination Function

The distributed coordination function (DCF) can be used in infrastructure mode or ad hoc mode, because it allows multiple wireless stations to communicate with each other without any central control. The DCF must be implemented in all 802.11-compliant wireless devices, since it is the basic media access mechanism in the 802.11 standard.

The DCF employs the *carrier sensing multiple access with collision avoidance* (CSMA/CA) mechanism for channel access. The carrier sensing is performed using both a physical medium-dependent method (e.g., *Channel Clear Assessment*) and a virtual carrier sensing mechanism called *Network Allocation Vector* (NAV). Stations sending packets set the NAV to the time for which it expects to use the medium, while other clients count down from the NAV to zero during the packet transmission. When the NAV becomes zero, the carrier-sensing function indicates that the medium is idle/free.

Before sending a data packet, the station first uses carrier sensing to determine

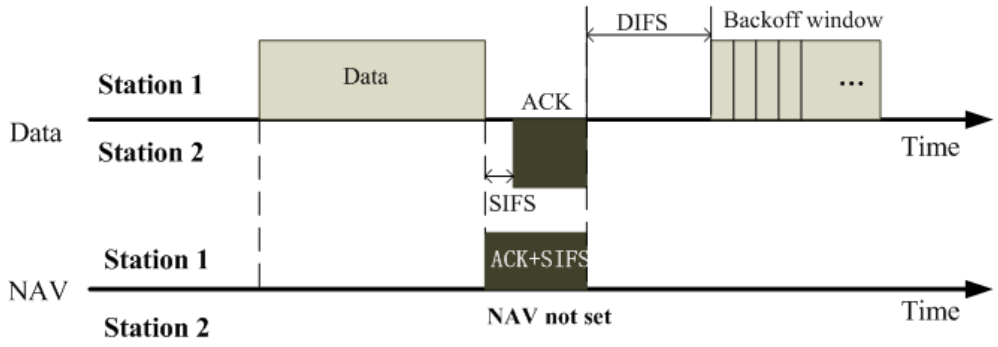


Figure 2.3. The basic mechanisms of the DCF.

whether the medium is free for transmission. If the medium has been idle for a *DCF interframe space* (DIFS), the station will transmit data immediately. Otherwise, the access will be deferred, and the station will keep detecting the channel continuously. If the packet is destined for a single station, the *unicast* packet must be positively acknowledged to ensure reliability. The receiver responds with an acknowledgment frame after it has successfully received the packet with a short delay given by the *Short interframe space* (SIFS). The basic transmission process of a single packet is shown in Figure 2.3 [65], where $DIFS > SIFS$.

When the packet transmission has completed and a DIFS has elapsed (if the previous transmission contained errors, the medium must wait for a *Extended interframe space* (EIFS)), the station will send another packet after a *backoff window/timer* by slots, which is randomly selected by

$$BackoffTimer = slotTime \times INT(CW * Random()).$$

Here, the slot time is medium-dependent: higher-speed physical layers use shorter slot times. The *contention window* (CW) is also divided into slots whose size is always 1 less than a power of 2 (e.g., $7, 15, 31, \ldots, 1023$). The CW is initially set as the minimal CW and will not increase over the maximal CW. When several stations are attempting to transmit data, the station which picks

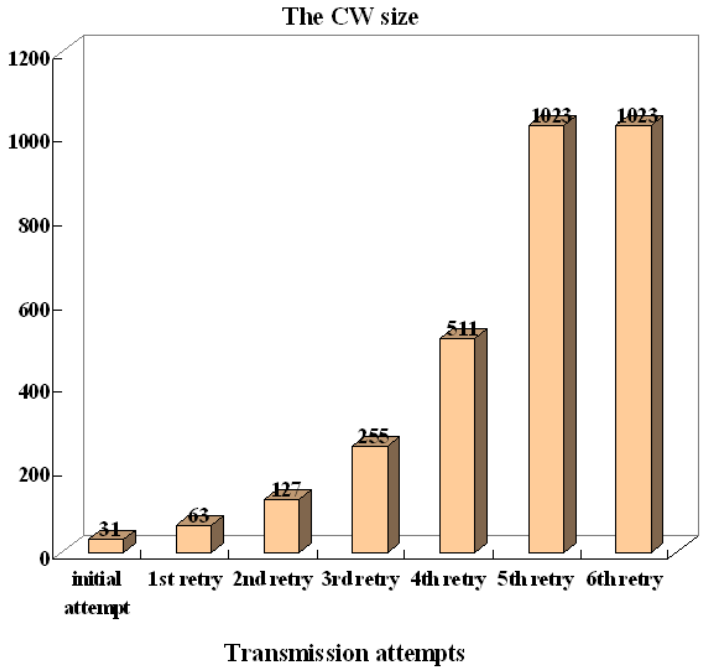


Figure 2.4. The growth of the CW size ($31 \leq CW \leq 1023$).

a shorter backoff timer will win. Therefore, simultaneous data transmission from multiple stations could occur in the DCF which will result in channel collisions. Then the stations involved in collision will use an exponential backoff mechanism of CW to resolve this problem. Each of them resets the backoff timer after double its current CW. When the CW reaches its maximal size, it remains. Moreover, the CW will not go back to the minimal size until the packet is transmitted successfully or discarded. Figure 2.4 illustrates the growth of the CW by the number of transmission attempts.

The DCF may optionally use the *Request to Send* (RTS) and *Clear to Send* (CTS) frames to guarantee reservation of the medium and uninterrupted data transmission, which could further reduce frame collisions [65]. Figure 2.5 shows this process. Before sending a data packet, a sender sends an RTS to request the right of transmission when the medium has been idle for a period of DIFS

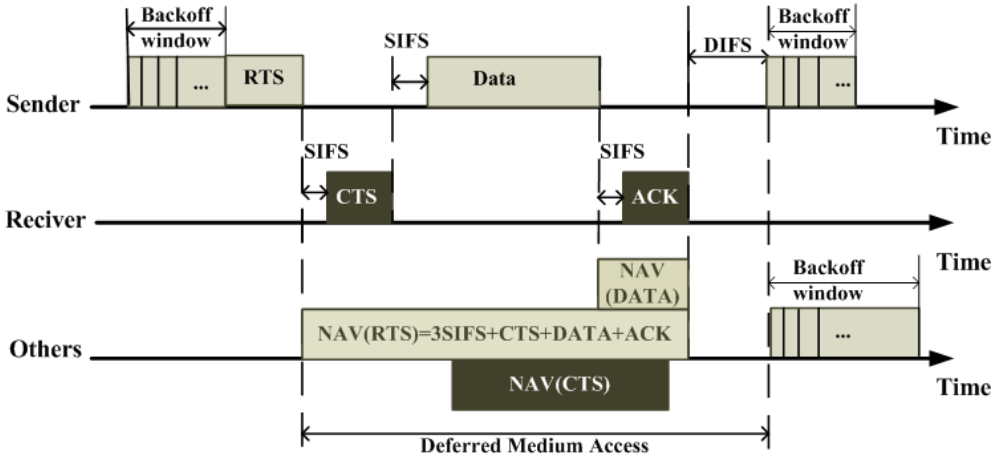


Figure 2.5. The process of DCF with RTS/CTS.

and one backoff window. After the receiver has received the RTS and a SIFS has elapsed, it will send back a CTS frame. After receiving the corresponding CTS, the sender will begin data transmission after a SIFS. Each successful reception of a data frame will be acknowledged after a delay of SIFS. Meanwhile, other stations will defer their medium access. When the medium has been idle for a DIFS according to the NAV, each station will pick a backoff window according to the updated CW individually. The station with a shorter backoff window will win the channel contention and begins the next round of data transmission. For example, the sender in Figure 2.5 will continue sending another frame, because its backoff window is less than others.

The Point Coordination Function

The point coordination function (PCF) is typically based on a polling access control scheme which polls the wireless clients in a fixed order. The PCF is therefore restricted to infrastructure networks, because a central unit is necessary for coordinating the medium access. It has not widely been implemented and is

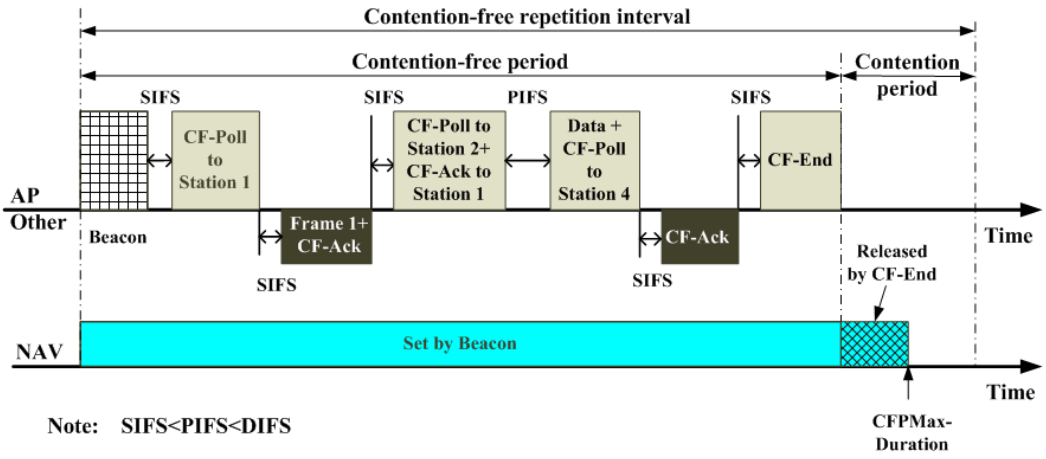


Figure 2.6. The mechanisms of the PCF in the IEEE 802.11 networks.

usually built on the top of the DCF, for example in the network interface of PCMCIA [3], the AP of Aopen WarpLink AOI-706 [66] and some advanced implementations of 802.11 MAC (e.g., HCF [88], Enhanced MAC architecture [100] and UPCF [47]). SIEMENS also provides *Industrial Point Coordination Function* (iPCF) [89], which is a proprietary alternative to PCF in the access point of SCALANCE W788 RR.

In the PCF, the AP sends beacon frames at regular intervals (every 100ms by default). Between two consecutive beacon frames, the PCF operates in two periods: *Contention Free Period* (CFP) and *Contention Period* (CP). The DCF is used during the CPs, whereas the AP polls the stations by sending *Contention Free-Poll* (CF-Poll) frames during the CFPs. Obviously, the PCF gives AP more control over access to wireless medium and helps guarantee the performance of wireless stations. It best suits delay sensitive data transmissions, for example real-time audio/video.

Figure 2.6 [65] depicts data transfer using the PCF. At the beginning of CFP, the AP transmits a beacon frame that announces the maximum duration of the

CFP, *CFPMaxDuration*. All stations set the NAV according to the *CFPMaxDuration* and lock out the DCF-based access. The AP controls the wireless medium and polls respective stations according to a polling list. It sends a CF-Poll frame to give Station 1 the right to transmit a single packet. After receiving the CF-Poll, Station 1 sends Frame 1 and a *CF-ACK* frame which acknowledges the frame receipt when no more data will be transmitted. Then the AP sends another CF-Poll to Station 2 and returns a CF-ACK to Station 1 after Frame 1 has been received with a delay of SIFS. However, Station 2 which is on the polling list fails to respond to the CF-Poll because of the loss of responding packet or other exceptions (e.g., being out of the range of this AP). In order to maintain the central control, the AP polls the third station on the polling list after waiting for a *PCF interframe space* (PIFS), which is longer than a SIFS. Then the AP sends the frame of *Data+CF-Poll* to transmit data to Station 4. Station 4 receives this frame and returns a CF-ACK after waiting for a SIFS. When the duration for polling all stations in the list is less than the *CFPMaxDuration*, the AP sends a *CF-End* frame to end the CFP and release the medium control by resetting the NAV to zero. After entering the CP, the medium is controlled the by DCF. Alternating periods of CFP and CP repeat at regular intervals, which is called the *contention-free repetition interval*.

To effectively support real-time applications, a new coordination function, known as *Hybrid Coordination Function* (HCF), is proposed in IEEE 802.11e [88]. There are two channel access methods: *Controlled Channel Access* (HCCA) and *Enhanced Distributed Channel Access* (EDCA). The EDCA enhances the DCF function for contention-based services, whereas the HCCA enhances the PCF function for contention-free services. Both the EDCA and HCCA operate based on traffic classes (TC). The traffic class that has a higher priority will be given

a higher priority to access the channel. For example, network traffic from email and FTP applications is often assigned with a lower priority, whereas that from Voice over Wireless LAN (VoWLAN) is assigned with a higher priority.

2.2 Energy-saving Schemes in IEEE 802.11 Networks

2.2.1 Energy-saving Schemes on the Physical Layer

The PHY layer consists of radio frequency circuits, modulation and channel coding system. Early approaches save wireless client's energy concentrates on this layer. A most straightforward approach is to develop a good circuit to reduce the power consumption in different operation states. For example, Chandrakasan et al. discussed many special techniques to minimize power consumption in CMOS circuits [37]. They further investigated low-power hardware techniques to reduce the power consumption of CMOS digital circuit while maintaining throughput [38]. Moreover, the design of minimal logic circuit makes the PSM scheme possible. Using the PSM scheme, the wireless network interface card (NIC) can shut off its power for everything except timing circuit, such that it consumes very little power in the sleeping mode. The support for multiple levels of transmission power and various PHY transmission rates are the building blocks for energy-efficient mechanisms on other layers.

On PHY layer, voltage scaling, convolution code strength and radio transmission power are regarded as three crucial knobs for energy efficiency (e.g., [123]). In this thesis, the control of transmission power is considered one of the two main

schemes to improve energy efficiency, which will be discussed in Section 2.3. The modulation level in the error coding scheme also gives rise to a tradeoff between power and performance. Dynamic modulation scaling (DMS), dynamic code scaling (DCS) and dynamic modulation-code scaling (DMCS) are three typical mechanisms to achieve energy efficiency.

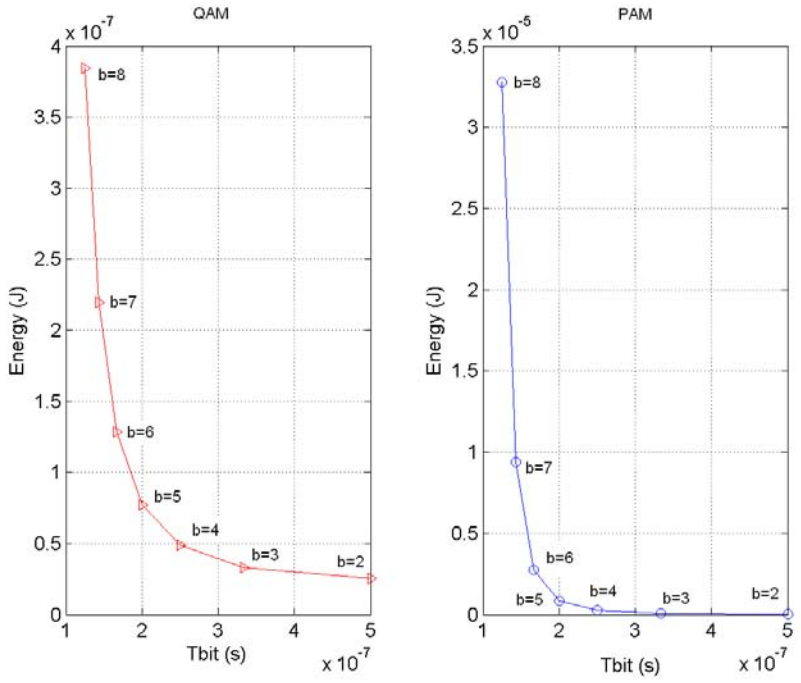


Figure 2.7. Energy-delay tradeoff achieved by the dynamic modulation scaling method.

Using the DMS [154] as an illustration, the modulation level b is the control knob which specifies the number of bits encoded into one symbol. Let R_S be the symbol rate. The average time for transmitting one bit over a wireless channel is then given by $T_{bit} = 1/(bR_S)$. The energy required for transmitting one bit is given by $E_{bit} = [C_S f(b) + C_E]/b$, where the constants C_S and C_E and $f(b)$ are given in [149]. The convex energy-delay curves in Figure 2.7 shows the relationship between E_{bit} and T_{bit} when *quadrature amplitude modulation* (2^b-QAM)

and *phase shift keying* (2^b-PAM) are used. Obviously, b has a significant effect on both T_{bit} and E_{bit}. We can adjust b to achieve energy efficiency by balancing energy consumption and transmission time. Moreover, b has a close relationship with the *bit error rate* (BER). Therefore, the DMS can jointly adjust b and R_S to efficiently schedule real-time, energy-aware packet transmissions over wireless channels under the control of BER (e.g., [155]). Furthermore, E^2WFQ [149] applies the DMS to the leaky bucket regulated input stream and fairly schedules packets to achieve energy efficiency.

2.2.2 Energy-saving Schemes on the MAC Sublayer

The traditional MAC protocol is designed to maximize packet throughput, minimize latency and provide fairness among wireless stations. It determines the communication performance and energy consumption of wireless stations. Many researchers focus on proposing efficient MAC protocols to balance performance and energy consumption. The energy-saving design in the MAC sublayer addresses two main problems: How to reduce collisions and how to schedule data transmission to save energy.

Since the packet retransmissions caused by collisions consume energy and increase transmission delay, reducing collisions is a viable way to reduce energy consumption. For example, Singh et al. [158] proposed the PAMAS protocol to achieve energy efficiency by reducing collisions and enhancing the IEEE 802.11 PSM. It separates the channels for RTS/CTS control packets and data packets. Each wireless client senses the signaling channel to determine when to turn on the transceiver for communication and when to turn it off for sleeping. The reliable control of RTS/CTS greatly reduces the probability of collisions in the

channel of data transmission. Monks et al. [126] proposed the PCMA protocol as another way to avoid collisions by using the control of *request-power-to-send* (RPTS) frame and *acceptable-power-to-send* (APTS) frame. Each sender sends a RPTS before sending a data packet, and the receiver replies with an APTS packet when it is available. The RPTS/APTS exchange determines the suitable transmission time slot for the sender and avoids collisions. At the same time, it also determines the minimal transmission power which guarantees the successful packet transmission.

Packet transmission scheduling is another common approach of reducing energy consumption. Some transmission schedulers adjust transmission power to effectively save energy. The technique of transmission power control will been discussed in detail in Section 2.3. Some extended energy-efficient packet schedulers [132, 90] also take the property of energy recovery in battery into consideration. Other transmission schedulers extract relevant parameters from the PHY layer to enhance the energy efficiency of MAC protocols. For example, the PHY-MAC access mechanism [7] adopts an automatic transmission rate adaptation to increase energy efficiency, which is based on the traffic information on the MAC sublayer and the channel condition on the PHY layer (e.g., BER).

When a PSM scheme is used, a wireless client could save energy at the cost of increasing complexity at the MAC protocol. For example, the *Energy-efficient Distributed Access* (EDA) scheme [16] modifies the virtual carrier sense mechanism and the backoff function specified in IEEE 802.11 DCF. As a result, the wireless client achieves energy efficiency by staying in a low-power state, instead of spending energy on overhearing channel during channel contention. The PSM scheme in the IEEE 802.11 standard saves energy by allowing the clients to stay in a low-power state. For example, the IEEE 802.11e proposes an *Automatic*

Power Save Delivery (APSD) [88] based on the basic PSM. In the APSD, the delivery of downlink frames is scheduled by a *Hybrid Coordinator* (HC) in a given service period (e.g., the *Power Save Feedback Based Dynamic Scheduler* (PS-FEBDS) [28]). In Section 2.4, we will discuss it and its previous studies in detail.

Especially, the energy-efficient designs of polling based MAC protocols have been paid attention, since these protocols are widely used in infrastructure wireless networks. In general, polling based MAC protocols deploy two grouping schemes [76]: *phase grouping* (PG) and *mobile grouping*(MG). The PG polling-based protocols have been proposed to improve energy efficiency by reducing or avoiding channel collisions. The typical example of PG scheme is the *energy conserving MAC protocol* (EC-MAC) [44], which uses an explicit transmission order with reservations. In the broadcast (downlink) phase, the wireless client listens for a transmission schedule. The AP's centralized scheduler tries to reduce energy consumption and support QoS in the uplink phase. Many MG polling-based protocols have also been proposed to minimize energy consumption by reducing the number of mode transitions and by controlling the packet transmission [73], such as the *disposable token MAC Protocol* (DTMP) [72], the E^2MaC *protocol* [75] and the *mobile grouping scheme* [77].

2.2.3 Energy-saving Schemes on the Upper Layers

Besides PHY and MAC layers, researchers have also proposed many energy-saving schemes on upper layers. Next, we introduce some of them on logical link control (LCC) sublayer, network layer and transport layer.

The Logical Link Sublayer

A link layer protocol is used to establish a reliable logical link over unreliable wireless channels. But the BER in a wireless link is usually high. It could be as high as 10^{-3} in adverse environments. Therefore, the wireless link error control is extremely important, and the energy-efficient design in the LLC sublayer concentrates on reducing the number of retransmissions caused by errors. In general, there are two approaches.

One approach is to improve error control schemes in terms of energy efficiency [74]. Existing schemes can be categorized into two classes: *automatic repeat request* (ARQ) and *forward error correction* (FEC). Using ARQ, a feedback is expected by a sender after sending a packet. If the feedback is lost, a receiver has to wait for a long time and retransmit the entire packet. It is easy to see that ARQ schemes do poorly when channel condition is bad, since clients have to spend much energy and time on retransmission. FEC attaches redundant bits to packet and allows each receiver to detect and even correct errors. When channel is bad, FEC preemptively deals with the erroneous packets instead of dropping them immediately. This scheme efficiently reduces energy consumed on retransmissions. However, when channel is good, FEC incurs a fixed overhead of every packet in terms of energy and delay. In summary, ARQ schemes are suitable for a network with good channel condition while FEC schemes are better for a network with bad channel condition.

Some data link designs various error control schemes according to various channel conditions. These adaptive error control schemes reduce the number of retransmissions and avoid persistent retransmissions. It is helpful to reduce energy consumed on retransmission and IDLE mode. For example, the AIRMAIL

wireless link-layer protocol [14] uses the hybrid ARQ/FEC design to obtain a better performance in terms of end-to-end throughput and latency in an unreliable wireless channel. At the same time, wireless stations spend less energy on retransmission. Furthermore, Lettieri et al. [114] adaptively controlled the frame length in concert with adaptive error control, which is also a hybrid of FEC and ARQ.

The other approach is based on channel detection. It helps wireless device avoid wasting energy on transmitting error packets. When the channel condition is detected to be poor, transmission will be inhibited or deferred. To do that, when a sender detects a lack of an ACK, it will enter probing mode and periodically transmit a probing packet. The probing continues until an ACK is received successfully, which is a sign of a good channel condition. After that, this sender resumes transmitting its packets. In this way, the transmission errors are mostly avoided, and the retransmissions are greatly reduced to save energy.

Furthermore, the combination of these two approaches is also used to improve energy efficiency. For example, energy efficient probing ARQ protocol has been investigated in [184], which slows down retransmissions during fades. Soni et al. [161] modeled a multi-path fading channel by a first-order Markov chain. After analyzing throughput and energy efficiency of UDP on the link layer, they proposed the link layer backoff algorithms to improve energy efficiency.

The Network Layer

The network layer is responsible for routing, congestion control and mobility management. Here, we discuss the energy-efficient routing in ad hoc wireless network. Usually, it cannot be applied into infrastructure network where an AP

relays traffic for all wireless stations.

In an ad hoc wireless network, the typical metric of shortest-hop in wired network is not applicable generally. There are two reasons for that. First, it may consume more energy or overuse energy of a small set of wireless clients, which decreases the life time of an entire network. Second, the experiment results from wireless test-beds have presented that quite a few minimum hop-count paths suffer from poor throughput [52]. Next, we discuss three categories of routing in ad hoc wireless networks.

Minimum energy routing minimizes total energy consumed on forwarding one packet (e.g., [55, 175, 173]). This type of routing protocols search for a forwarding path composed of several low-energy hops, which may consume less energy than a single hop path. Then it is time to make a choice. The path with one high-energy hop may consume more energy but the probability of a successful transmission is high. The path with several low-energy hops reduces energy for each hop at the cost of routing failures and packet errors (i.e., with a higher BER). It may spend much energy on packet retransmissions and thus degrade the communication performance when channel condition is bad. To deal with this, two approaches are proposed to find an energy-efficient path. The proactive approach finds route paths for all stations once and compares their energy consumption (e.g., *minimum energy disjoint path routing* in [162]). The reactive approach finds route path hop by hop, which is also called adaptive minimum energy routing. For example, the power-aware routing optimization protocol (PARO) [68] is implemented by adaptive operations.

The max-min routing selects a route which maximizes the minimal residual energy of stations. It may select a long route and even increase total energy consumption. But it balances energy consumption among stations to increase

the network lifetime as long as possible. For example, SPAN [41] avoids network traffic going through poorly powered wireless clients. Every station determines whether it should act as a coordinator to relay packets after taking into account its own energy reserve and its neighbors' states. Furthermore, RandomCast [115] is designed to improve energy efficiency of *Dynamic Source Routing* (DSR) in MANETs in which stations consistently operate in the PS mode. It also balances energy among stations without significantly affecting the routing efficiency. In comparison, minimum energy routing does not maximize the network lifetime, although it performs better than max-min routing in terms of throughput and delay.

Minimum-cost routing reduces the cost of forwarding packet at each hop and selects the route which minimizes the total cost of hops. It operates on the basis of the cost function which is generally a monotonic increasing function of energy consumption. For example, the shortest cost path routing algorithm in [40] uses a function of link cost that reflects both the communication energy consumption and the residual energy levels at stations. Moreover, in order to find a reliable route with minimum energy, the cost function may also reflect the expected cost on reliably forwarding a packet (e.g., MRPC [125]).

In summary, energy-aware routing protocols maintain the connectivity of network (e.g., PARO [68]) and increase the life time of network.

The Transport Layer

The transport layer provides reliable or unreliable end-to-end data communication service. The most common protocol is *Transmission Control Protocol* (TCP) [140]. TCP performs well in wired networks, but its performance signif-

icantly degrades in wireless networks. In a wireless channel, the probability of successful transmissions is low due to a relatively high BER. As a result, TCP frequently invokes the congestion control algorithm, because it cannot distinguish between the wireless link errors and the packet loss caused by the channel collision (or the handoff). It results in reducing the sender's *congestion window* (CWND) and induce unnecessary packet retransmissions. Therefore, the traditional TCP operating over a wireless network could suffer from low throughput, long delay and high energy consumption.

One approach to increase TCP's performance over wireless networks is to split the TCP connection into the wired part and wireless part (e.g., Indirect-TCP [17], Berkeley Snoop Module [18] and M-TCP [31]). The AP is responsible for splicing the wireless TCP connection and the wired TCP connection. The wired TCP connection is based on a typical TCP implementation, while the wireless TCP connection often uses a modified version of TCP that enhances performance over a wireless channel and saves energy of wireless stations.

Some researchers proposed wireless TCP protocols which enhance the typical TCP in response to the features of wireless communication. The selective acknowledgement allows a TCP sender to recover from multiple packet losses (e.g., TCP Westwood [121] and ATCP [117]). The explicit loss notification aids a TCP sender to distinguish between congestion, channel collision and other forms of loss. The sender will not update CWND or waste energy on retransmitting packets if the loss is due to the bad channel condition. In a similar way, the explicit link failure notification techniques in [81] can significantly improve TCP performance in terms of *expected throughput* and energy saving.

However, some researchers believe that the typical TCP is fundamentally inappropriate for wireless networks. They have designed entirely new transport

protocols, instead of using TCP variants, to provide reliable data transmission over a wireless link. One example is the ad-hoc transport protocol (ATP) [164]. It adopts rate adaptation, quick-start during a connection initiation and coarse grained receiver feedback. The congestion detection and control is supported by network, and the timeout of retransmission is not considered. Therefore, the ATP avoids retransmitting TCP packet when packet loss is due to the error or collision of wireless channel. As a result, much energy is saved in wireless stations, especially when the wireless channel is bad. The other example is the energy-conscious transport protocol (JTP) [150], which is designed to achieve a given objective of network-wide energy efficiency while incorporating application requirements in an energy-constrained environment.

2.3 The TPC-based Energy-efficient Schemes

2.3.1 The TPC Operation

Transmission power control (TPC) saves communication energy by transmitting packets over a longer duration, i.e., reducing the transmission rate. Tarello et al. [168] performed TPC according to a linear and a strictly convex rate-power curve, which minimizes the overall expected energy consumed in transmitting a fixed amount of data while satisfying a deadline constraint. According to a given power-rate function, the TPC may also determine the most energy-efficient transmission strategy for data frames, for example, the optimal rate-power combination table in Miser [144].

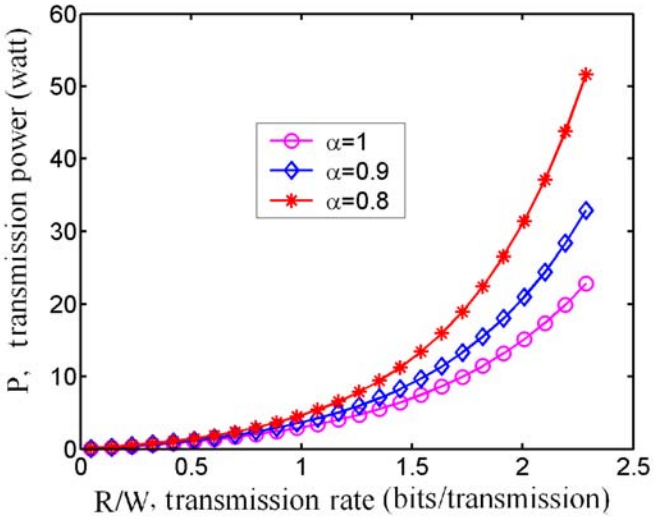
According to Shannon's theorem, the ideal channel capacity is given by $C_{max} = W \times \log_2(1+\frac{S}{N})$ bps, where W is the channel bandwidth in Hz, N is the Gaussian

noise power, and S is the signal power. The transmission rate R is given by αC_{max} bps, where α ranges from 0 to 1. By introducing an attenuation factor A which is the ratio of the transmission power P to S, we obtain the power-rate relationship (2.1). Note that a lower transmission power corresponds to a lower transmission rate. Figure 2.8(a) illustrates the relationship between the transmission power P and the rate $\frac{R}{W}$ by the unit of bits per transmission.

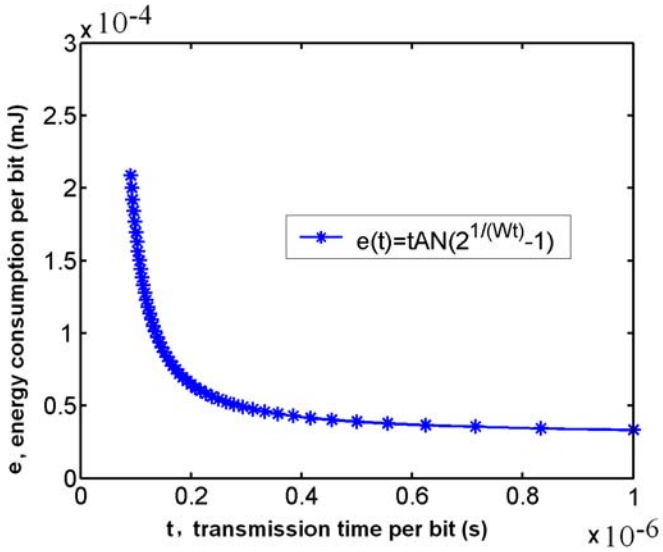
$$R = \alpha W \log_2(1 + P/(AN)). \tag{2.1}$$

As will be seen later, energy efficiency depends on both channel conditions and power consumption characteristics. It can be measured by the energy required to communicate (transmit and receive) one bit of information [153, 80]. In Figure 2.8(b), the energy spent on transmitting one bit, $e(t) = tP$, is a convex function of the time for transmitting one bit, $t = \frac{1}{R}$. The energy-time curves clearly show that $e(t)$ can be effectively reduced by lowering P initially, since the lower transmission power results in a longer transmission time. However, when t is further prolonged, the gain in reducing the energy efficiency starts diminishing. Once entering the tail zone, a further reduction of transmission power can only save an insignificant amount of energy but greatly increases transmission time. In fact, having the client transmit faster and then go into the sleep mode will be more energy efficient than the TPC in this region of operation.

The proper reduction of transmission power and the increase of sleeping period are both helpful to save energy. But they conflict with each other. If the transmission power is too high, much energy is consumed on transmission. If the transmission power is too low, the extended transmission time will prevent client from sleeping and then increase energy consumption. Therefore, there exists an



(a) Power-rate curves.



(b) Energy-time curves.

Figure 2.8. Tradeoffs between energy and transmission rate.

optimal transmission power when taking the TCP and sleeping into consideration. Moreover, since the client is more prone to instability when t increases, the TPC must operate under the constraint that the client is stable (i.e., the transmission rate is higher than the incoming rate of data frames).

In this thesis we measure the energy efficiency by bits per joule which has been widely used in [151, 122, 15, 85]. It is defined by the number of bits communicated by consuming one joule of energy, denoted as $z = 1/e(t)$. Another widely used metric is the average total transmission energy required to transmit a single packet reliably to the destination [19]. We do not choose this metric, since it considers the energy spent on retransmissions which may be caused by others reasons, such as channel errors and channel collisions.

2.3.2 TPC-based Schemes

Reduction of transmission power in a wireless client may reduce the energy consumption for transmission, but it also yields two conflicting results: (1) the degradation of quality which results in a great deal of retransmission due to errors or collisions, (ii) the mitigation of the interference among other clients and reduction of errors and collisions. Accordingly, Broustis et al. [30] conducted experiments to quantitatively study the implications of power control on interference and channel contention. The purpose of the study is to understand the transmission power trade-off: decreased power consumption, reliability and performance on one client verses decreased levels of interference caused to other clients. D. Qiao et al. [145] provided a rigorous analysis of the relationship between different radio ranges and TPC's effects on the interference in 802.11 systems. He further proposed that control frames (e.g., CTS frames) should be transmitted at a stronger power level to ameliorate the TPC induced interference.

Some studies, such as [6], adopt the basic TPC scheme in IEEE 802.11, which uses different power levels to send RTS-CTS and DATA-ACK packets. RTS-CTS packets are sent by a maximal transmit power, while DATA-ACK packets

are sent by a minimal transmit power to save energy. However, Jung et al. [95] have shown that this technique may degrade network throughput and even result in high energy consumption on retransmissions. In this case, collisions increase greatly, because the RTS-CTS packets are insensitive to low power transmissions. Therefore, they proposed *power control MAC* (PCM) [95] to reduce collisions and save energy. This is similar to the basic MAC except that it periodically increases the transmit power to the maximum during the data packet transmission.

TPC is also suitable for balancing energy conservation and performance metrics, such as delay (or transmission time), throughput and network utility. Some packet transmission schedules use TPC to minimize energy consumption subject to a deadline. For example, the optimal offline/online schedules [27] and Move-right algorithm [64] provide the optimal transmission power of each packet (which is derived from the optimal transmission time of each packet). They let the single transmitter uses the minimal energy to transmit packets within a given amount of time. Nuggenhalli et al. designed a delay-constrained transmission schedule that maximizes the lifetime of a transmitter [131]. Instead of considering power and delay individually, Bettesh and Shamai [23] used the weighted sum of the average power and the average delay as the performance metric. Then they employed a dynamic programming approach to optimize this new metric. What is more, Zhang and Chanson [181] have obtained an optimal transmission power to minimize the energy consumption subject to the given throughput constraints, or maximize the throughput subject to a fixed energy storage. Chiang and Bell [46] defined a network utility based on all user utilities and maximized it subject to energy constraints. The optimal transmission powers and optimal transmission rates are obtained for all wireless clients in a WLAN.

Some TPC schemes are integrated with the algorithms of TCP congestion

control and the adaptation of transmission rate to increase communication performance and energy efficiency. For example, Chiang [45] provided the global optimum of joint power control and congestion control (e.g., TCP Vegas) to balance the data rate demand (i.e., throughput which is regulated through TCP) and supply (i.e., energy consumption which is regulated by TPC). The SP-TPC [101] and the cooperative/non-cooperative media control [8] [70] adopt the TPC and PHY rate adaptation jointly to optimize the network throughput. MiSer [144, 145] computes offline an optimal rate-power combination table, then at runtime, a wireless station determines the most energy-efficient transmission strategy for each data frame transmission by a simple table lookup. The energy-efficient PCF [146] allows a wireless client to adaptively select the best transmit power level as well as the proper PHY rate to transmit an uplink data frame. The combination of TPC and adaptive PHY rate selection minimizes the energy consumption with/without the constraints of goodput. Berry and Gallager [21] adapted the user's transmission power and rate based on the channel state information as well as the buffer occupancy. They also provided several characteristics of the optimal power/delay tradeoff curve and found all Pareto optimal power/delay operating points.

In this thesis we deploy the AP-centric TPC on the PSM-enabled wireless clients in an infrastructure wireless network when it uses a polling-based MAC protocol. A polling system model is set up to study the communication performance and energy consumption of all clients. The AP obtains the optimal transmission power allocation for all clients to minimize the total energy consumption of clients subject to stability instead of a specific metric. The PSM-enabled client follows the existing polling-based MAC protocol and transmits packet with its optimal transmission power.

WNIC	802.11.*	Sleep	Idle	Reception	Transmission
Dell MiniPCI 1150	b	0.10 W	0.66 W	0.76 W	1.09 W
Cisco Aironet 350	b	0.24 W	1.41 W	2.61 W	3.69 W
Interl PRO/WL2100	b	0.91 W	1.82 W	3.50 W	3.35 W
Intersil	b	0.46 W	1.95 W	8.84 W	7.42 W
Agere	b	0.14 W	0.98 W	10.6 W	9.37 W
Atheros AR50001X+	a	0.19 W	1.14 W	2.60 W	2.61 W
	b	0.87 W	2.59 W	5.85 W	6.15 W
	g	0.92 W	2.56 W	8.45 W	8.36 W

Table 2.1. Power consumption in different operation modes.

2.4 The PSM-based Energy-efficient Schemes

2.4.1 The PSM Operation

The IEEE 802.11 standard [86] allows a wireless client to enter a *sleep* mode that incurs a very low power consumption. In this mode, the client need not transmit or receive data or sense wireless channels. In contrast, the wireless client in an *active* mode is able to exchange data and overhear traffic. It consumes a high power during the states of transmission, reception and idleness. Note that, the power consumption during a transition from sleep mode to active mode is as high as the one in active mode [105]. Its transition time is in the order of a few milliseconds. But the opposite transition is assumed to be cost-free because of its extremely short transition time. As an example, Table 2.1 [4, 10] shows the average power consumption of the wireless NICs (WNICs) under different operating modes.

In an wireless infrastructure network, each wireless client informs the AP whether it utilizes a PSM scheme. The AP buffers the frames which are addressed to the sleeping clients. It also periodically broadcasts beacon frame and announces the buffer status of all PSM-enabled clients in *Traffic Indication Map*

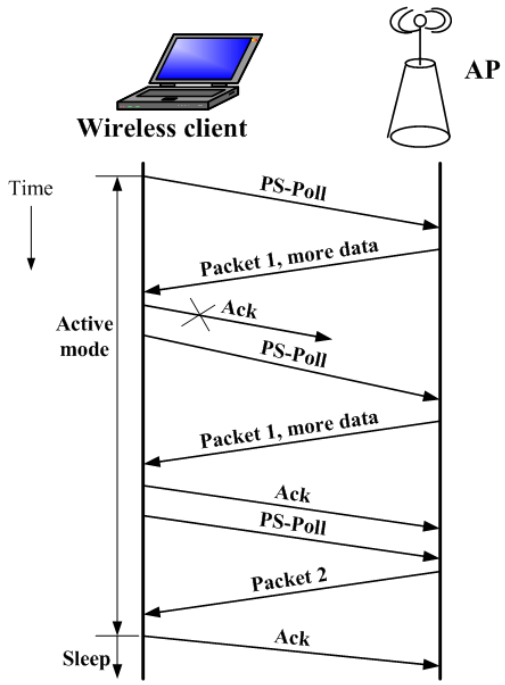


Figure 2.9. PS-Poll frame retrieval.

(TIM). The period of beacon broadcast is Beacon Interval (BI), denoted as β in milliseconds (ms). PSM-enabled client also periodically wakes up to hear one beacon and receives buffered frames according to TIM. Each PSM-enabled client's *Listen Interval* (LI) is a multiple of BI; therefore, the BI actually determines the LI's granularity. Let the value of LI be $\gamma \times \beta$ milliseconds, where $\gamma \geq 1$. For the PSM, the default settings are $\beta = 100$ ms and $\gamma = 1$. If an awakened PSM-enabled client hears a beacon whose TIM indicates the presence of its buffered frames in the AP, this client will stay active and send a PS-Poll frame to retrieve the first frame. After receiving one frame, the client sends out an ACK. If the *More Data bit* in the received frame is not set, the PSM-enabled client will return to sleep; otherwise, it will send another PS-Poll. The frame exchange goes on until all buffered frames are sent. After that, the PSM-enabled client returns to the sleep mode. Figure 2.9 [65] illustrates the process. However, if a PSM-enabled

client wants to send data, it may wake up any time to initiate a transmission according to the IEEE 802.11 DCF.

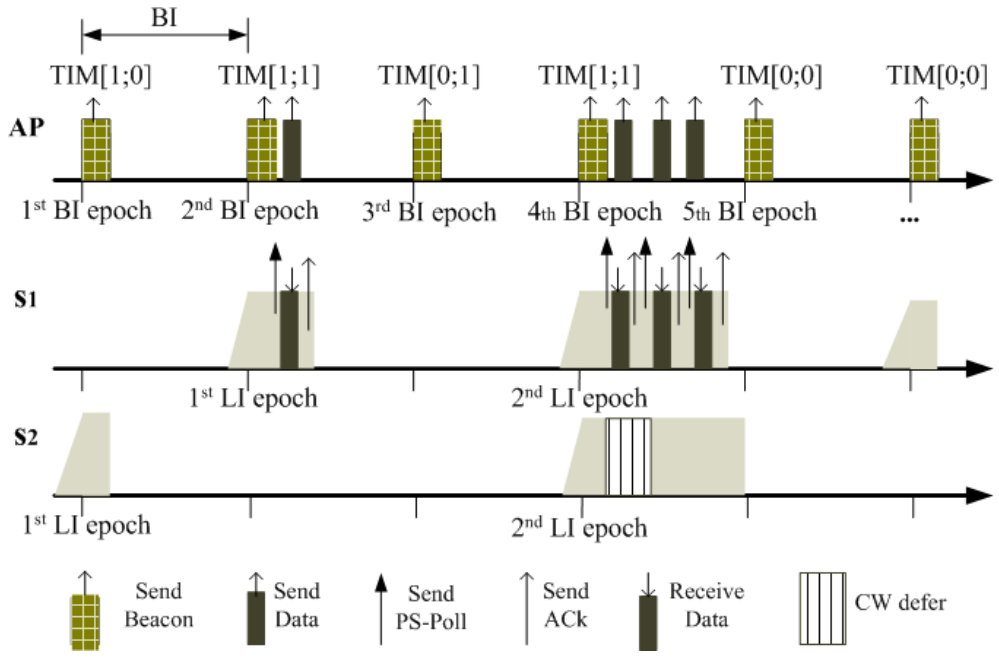


Figure 2.10. Buffered frame retrieval process.

Next, we discuss the buffering and delivery process of PSM when the AP supports more than two PSM-enabled clients. Figure 2.10 [65] gives us an example including an AP and two clients–s_1 and s_2. s_1 has a LI of 2 while s_2 has a LI of 3. The first (second) bit in the TIM indicates the buffer status for client s_1 (s_2). The shaded region shows that the client is in the active state. At the first BI epoch, there are frames buffered only for s_1. s_2 wakes up at its LI epoch and immediately return to sleep, since no frame is buffered to it. s_1 keeps sleep, since it does not wake up. At the second BI epoch, the TIM indicates that there are frames buffered for both clients. It is the LI epoch for s_1 to wake up and receive every frame by sending a PS-Poll frame. After receiving all of its buffered frames, s_1 returns to sleep. Both clients are asleep during the third BI. At the fourth BI epoch, two clients wake up and are aware of buffered frames according to the

TIM of beacon. Both s_1 and s_2 prepare to transmit a PS-Poll frame according to the 802.11 DCF. s_1 finally wins the channel, and the AP transmits its buffered frame first. During the transmission, s_2 defers. Since the channel is always busy in this BI, s_2 cannot send any PS-Poll to retrieve data, therefore it must remain awake to compete for the channel by sending PS-Poll until the next TIM. If the AP has run out of buffer and has discarded the buffered frames of s_2, s_2 finally return to a low-power mode. We will come back to this figure later to analyze the energy consumption sources.

The PSM scheme also supports broadcast and multicast. The AP buffers the broadcast and multicast frames and transmits them immediately following a beacon frame with *Delivery TIM* (DTIM). These beacon frames with DTIM are transmitted every DTIM interval which is a multiple of BI. If a PSM-enabled client is not interested in receiving broadcast and multicast frames, it can skip DTIM beacon. Otherwise, it should not miss any DTIM beacon, in which case the DTIM interval must be a multiple of LI.

In a MANET, a client in the IEEE 802.11 PSM scheme periodically wakes up during the packet advertisement period, called the *Ad hoc Traffic Indication Message (ATIM)* window, to see if it has any data to receive. The ATIM determines when to sleep and when to wake up for packet reception. Since the size of the ATIM window has a significant impact on energy saving and throughput, some adaptive mechanisms choose ATIM dynamically to save energy without degrading throughput, such as DPSM in [94] and three power management protocols in [170]. In this dynamic distributed network without any central unit, the complexity of routing, the difficulties of synchronization and packet transmission control definitely increase. This thesis will not discuss the PSM in a MANET further. However, many papers refer to it and its development, such as [170, 41, 115].

Finally, we use Figure 2.10 to analyze two main sources of energy wastage: unnecessary wake-ups and channel contention. Clearly, the individual LI has a direct impact on the number of unnecessary wake-ups. Overly-frequent wake-ups will consume a significant amount of energy. For example, s_2 wakes up the first time to find no frames buffered for it. Energy wastage due to channel contention, however, is more complicated. For example, s_2 wakes up the second time to find frames buffered at the AP. However, it has lost to s_1 after contending for the channel during the PS-Poll transmissions. Client s_2 then stays in the IDLE mode and consumes high power consumption during the channel contention period. The worse case happens once the channel collisions happen, for example, when s_1 and s_2 send the PS-Poll frames at the same time. If then, two clients consume more energy in the IDLE mode during the backoff process which could take a long time according to IEEE 802.11 DCF. Therefore, rescheduling the active clients' wake-ups to different times will help reduce this source of consumptions (i.e., nonoverlapping of LIs). Note that reducing the BI value will also help, because the BI value determines the LI's granularity.

Therefore, the choice of BI and LI of clients has a great impact on energy consumption and communication performance [113, 137]. In order to save energy, the PSM scheme should let clients sleep as long as possible by using large BI and LIs. It also avoids PSM-enabled wireless clients wasting much energy on frequent unnecessary wake-ups. However, the AP may buffer many frames for multiple clients after a long BI and LIs and then indicate these clients in the TIM of Beacon. If then, the frame delay will increase, and two or more awaken clients which send PS-Polls to the AP for retrieving data will be involved in the channel contention. That is, the serious waste of energy occurs, since these clients spend much time on IDLE simultaneously. Therefore, the appropriate values of BI and

LIs should avoid frequent unnecessary wake-ups of clients and alleviate channel contentions by reducing simultaneous wake-ups of clients.

2.4.2 PSM-based Schemes

The IEEE 802.11 PSM is called static, because BI and LI do not adapt to dynamic conditions of network or traffic. The static PSM performs quite well in saving energy in limited scenarios, but results in a significant performance cost, for example, the adverse impact on the short TCP connections reported in [105]. Many studies have been carried out to balance energy saving and performance (e.g., [111, 147, 105, 112, 116, 127]), as well as adapt to the traffic patterns (e.g., [11, 13, 10, 127]). We divide the PSM schemes into two main categories—*AP-centric PSM* and *user-centric PSM*—according to the component performing the major enhancement. Note that the PSM scheme is different from the energy conservation mechanism in [156], which intelligently powers off wireless stations.

AP-centric PSM

An AP-centric approach lets AP deploy the PSM operations. The AP determines the operations of wireless clients under AP-centric PSM schemes.

Some AP-centric PSM schemes are designed in a network including an AP and a single wireless client. The power-saving management in [12] saves energy in a network with the following traffic pattern: data-transfer phases are characterized by *traffic bursts* interleaved by *idle* phases. The AP (i.e., PS-WiFi) follows the Indirect-TCP approach [17] and includes the very accurate estimator for short idle times (within a burst), the simple estimator for long idle times (between two bursts) and the Power Saving Algorithm (PSA) for predicting the idle duration.

Therefore, the AP accurately estimates when and how long the wireless interface will be idle and lets it sleep or become active accordingly. In a similar network scenario, the cross-layer energy manager (XEM) [13] allows wireless clients to switch on/off this power-saving management based on the detection of burst and idleness. These two power saving mechanisms are energy efficient but strictly limited by the burst traffic pattern under TCP of the singer client.

Many AP-centric schemes support multiple clients, and they are usually designed to reduce or totally eliminate channel contention. Lin et al. [116] have proposed several mechanisms to save energy by reducing or avoiding collisions among wireless clients using PSM. One is the wake-up schedule which balances the number of active stations at each BI. As a result, the wireless client has a lower collisions probability and saves more energy. Other mechanisms concentrate on how to poll the active client to avoid channel contentions. Using the beacon frame with redesigned TIM, the AP may schedule a single active client, or provide a sequence of several active clients based on *the smallest association first* or *the smallest queue length first*. Moreover, Lee et al. [111] pointed out that the energy consumption of each client and the delay of packet depend not only on the number of packets buffered in the AP but also on the schedule of packet transmission. Then they proposed two energy-efficient schedulers to minimize the average total power consumption of all the PSM stations per *beacon period* (BP). In this network, several PSM wireless clients connect to an AP, and they are required to wake up every BP. And the optimal BP for each schedule is obtained to guarantee the maximum allowable average packet delay.

Some schedule algorithms are deployed in AP to improve energy efficiency of all PSM-enabled clients. The AP in the scheduled PSM [78] assigns the time slices of a BI for the clients' buffered frames. All clients learn the slice assignment

information from the redesigned TIM. Accordingly, each station controls the state of the wireless interface, because it knows whether or not it has data buffered at the AP, as well as the timing of the expected data delivery procedure. The proposed scheme is near theoretical optimal for power saving in the sense so that it greatly reduces the interactions among stations by scheduling every transmission. The simulation results show that the energy saving is achieved at the cost of only a slight degradation of the one-way delay performance. Lee et al. [112] pointed out that maximizing throughput is an ultimate goal even for the wireless client using PSM. They considered throughput maximization in the first place and then minimized energy consumption as much as possible. The proposed scheme consists of an opportunistic power-saving mode and a scheduler called OEES (Opportunistic Energy Efficient Scheduler). First, the channel probing scheme and the TIM mechanism are combined to efficiently utilize the channel. Second, OEES chooses the client and schedules packets for the TIM period, which then decides the transmission order to minimize energy consumption on sending those packets.

Belghith et al. [20] discovered two reasons for the PSM inefficiency under high traffic loads–a large number of PS-Poll frames and the unknown information of actual power mode of wireless clients. Accordingly, they proposed two different enhancements called *Once Poll PSM* (OP-PSM) and *State Aware PSM* (SA-PSM). In OP-PSM, the frames buffered at the AP are forwarded upon the reception of a single PS-Poll. A station transmits a single PS-Poll after receiving a beacon with a TIM indicating some buffered data frames. If More Data field is set in a received data frame, the station will remain active, and the AP will keep forwarding the rest of the buffered frames. Upon receiving a data frame with the More Data bit unset, the station goes into the sleep state. In SA-PSM, a station

sends a sleep request (a Sleep-Request management frame) to the AP to enter the sleep state. If and only if the AP responds positively (by a positive Sleep-Confirm management frame) does the station go into sleep. Therefore, the AP knows the power mode of each wireless client. Accordingly, it directly forwards incoming data frames destined to stations currently in the wake-up state and stores only data frames destined to stations currently in the sleep state. Although these two schemes are effective to save energy, they have abandoned the frame retrieving process of PSM.

User-centric PSM

Several enhancements have been proposed for the PSM from the client side [105, 10, 180, 127, 167]. They belong to the user-centric PSM category. That is, each client determines when it will sleep and wake up.

Some user-centric PSM scheme could adapt to network conditions when they aim at the energy efficiency of a single wireless client. Yan et al. [180] let the wireless client predict when packets will arrive and keep the wireless interface in high-power mode only when necessary. Otherwise, the wireless client goes to sleep. This energy saving method is not flexible, since the accuracy of traffic prediction is easily impacted by various factors. PSM-throttling [167] effectively explores the power saving potential by controlling instead of detecting traffic pattern. Without using beacon frames, a wireless client wakes up at the beginning of each traffic burst, because it can identify bandwidth throttling connections and reshape the TCP traffic into periodic bursts. Therefore, PSM-throttling can minimize power consumption on TCP-based bulk traffic by effectively utilizing available Internet bandwidth. Furthermore, it is application-independent, client-centric and does not degrade the application's performance perceived by

users. Bounded Slowdown Protocol (BSD) [105] is proposed to minimize energy consumption while guaranteeing a delay bound based on the RTT of a TCP connection. With BSD, the wireless client sends a request and then stays awake for a certain period before sleeping. In order to reduce unnecessary wake-ups, the wireless client also increases its LI when the period of idleness increases. The experiment results show that BSD is suitable for TCP connections with short RTT. Moreover, Qiao et al. [147] have improved BSD by Smart PSM (SPSM), which minimizes energy consumption subject to delay performance constraints. After solving the optimization problem, each client determines whether it will enter into the PSM depending on the traffic condition. After the client enters into the PSM, the LI can be dynamically adjusted.

In order to adapt to the various requirements of different wireless clients, Nath et al. [127] proposed a dynamic wake-up period in which each client chooses its optimal LI according to the round-trip time of its current TCP connection. Accordingly, the AP sends a beacon with a TIM to each client according to its own LI. This technique saves a great deal of energy and has similar download times of Web pages compared with the standard 802.11 PSM. The distributed algorithm of selecting an optimal LI in one PSM-enabled client is simply extended to the network scenario when multiple wireless clients connect to the AP. However, this algorithm does not consider the channel condition of WLAN in general and specially neglects the channel competition among wireless clients. Self-tuning power management (STPM) [10] is also a typical user-centric PSM to support multiuser, which adapts to the characteristics of network interface, wireless clients and applications. In handholds and laptops, the STPM module implemented in the Linux kernel determines when to transit the wireless client to CAM (i.e., a continuously-aware mode without any power management) and when to transit

it back to PSM. The limitation of STPM is that the application code must be modified accordingly.

Table 2.2 compares the performance and features of some PSM schemes discussed above, where c is the number of PSM-enabled clients in the wireless network and $j \in \{1, \ldots, c\}$. All user-centric PSM schemes are designed to improve the energy efficiency of one target client. They do not address the power consumption due to channel contention. Moreover, it is not clear whether these schemes remain effective when some clients do not employ them. Although most AP centric PSM schemes generally perform well, the cost is a computation-intensive scheduling algorithm for the wake-up and frame transmissions.

Our new AP-centric PSM scheme, the centralized PSM (C-PSM), is designed to achieve energy efficiency by avoiding frequent unnecessary wake-ups and alleviating channel contentions. They have several features. First, unlike user-centric PSM schemes, C-PSM aims for the energy efficiency of all wireless clients instead of a single client. Second, C-PSM is traffic-aware, and we do not fix any traffic pattern. Many traffic distributions of clients are considered. Third, unlike using precise schedules of clients or packet transmission in other AP-centric PSM schemes, C-PSM updates a set of optimal PSM parameters to improve energy efficiency without modifying the standard operations of PSM scheme. The AP calculates accordingly the optimal parameters and updates the optimal BI. The wireless client updates its optimal LI and *contention window* (CW) which is notified in the beacon. Fourth, C-PSM proposes the first-wake-up schedule to further improve the energy efficiency.

Table 2.2. A summary of the PSM-based schemes.

Name	Centric	c	Target	Contention reduce	BI (ms)	LI period (ms) for client j	Delay	Adapt	Traffic
802.11 PSM	user	≥ 1	single	no	fixed, 100 by default	fixed, LI_jBI, $LI_j = 1$ (default)	degraded	no	All
BSD PSM[105]	user	≥ 1	single	no	100	dynamic, LI_jBI	bounded	yes	TCP
SPSM[147]	user	1	single	no	fixed, 100 by default	null, optimal action sequence	constrained	yes	TCP
DBP[127]	user	≥ 1	single	no	20	optimal, $LI_j^* > 1$, LI_j^*RTT	degraded	yes	HTTP
PSM-throttle[167]	user	≥ 1	single	no	null	null, sleep/wake up by prediction	not-degraded	yes	bulk TCP
EE-packet schedule[111]	AP	≥ 1	all	no	optimal, BI^*	BI^*	reduced	no	Poisson traffic
PS schedule[116]	AP	≥ 1	all	yes	100, scheduled TIM	fixed, LI_jBI	reduced with c	no	CBR
Scheduled PSM[78]	AP	≥ 1	single	yes	100, redesigned TIM	null, scheduled service period	degraded	yes	CBR UDP
Our PSM	AP	≥ 1	all	yes	optimal, $BI^* \geq 10$	optimal, $LI_j^*BI^*$	reduced	yes	traffic-aware

Chapter 3

New AP-centric TPC for Polling-Based MAC Protocol

Energy efficiency is one of the important design issues to consider in wireless networks. Various mechanisms have been proposed to balance energy consumption and the quality of communication [43]. In this chapter we mainly concentrate on the media access control (MAC) sublayer in a wireless infrastructure network where an *access point* (AP) serves as a relay for a number of wireless clients. The polling-based MAC protocols are investigated in this network, where the AP serves multiple clients according to the polling list, for example the *point coordination function* (PCF) in IEEE 802.11 standard. They provide a number of advantages over the random access MAC protocols. For example, they can guarantee a certain level of quality-of-service to each client even when the channel is accessed simultaneously by multiple clients.

As we shall see, the polling-based MAC protocol can also be used to optimize the energy consumption of the entire WLAN. In the beginning of each polling

cycle, the AP sends a beacon for time synchronization and the notification of the polling list according to the packets buffered in the AP. The client is notified when it should wake up to communicate, when it should sleep. Here, we assume that the polling list with the PSM scheme has been ready. Then we employ *transmission power control* (TPC) which is coupled with *power saving mode* (PSM) to achieve energy efficiency. It is the AP that calculates and notifies the optimal transmission power allocation to minimize energy consumption by balancing two conflicts energy-saving factors: the reduction of transmission power and the increase period of sleeping (refer to 2.3.1). Therefore, we design an AP-centric optimal TPC scheme to minimize the total energy consumption of all clients in an infrastructure wireless network with the constraint that all clients are stable.

We have conducted extensive experiments to validate the optimal transmission power allocations. Moreover, we have measured the improvements of the schemes with optimal TPC in terms of energy saving and a bit-per-joule metric. We have presented the effect of optimized schemes on the delay performance. Finally, the effect of an uncooperative client shows that all clients are incentive to follow the optimal transmission power allocation.

3.1 Two polling schemes

The polling list may use two types of polling schemes–*phase grouping* (PG) scheme and *mobile grouping* (MG) scheme. In the PG scheme, the uplink and downlink phases alternate as illustrated in Figure 3.1. The wake-up schedule of clients is broadcasted in *Traffic Control Slot* (TCS) which locates at the beginning of the downlink phase. Also in the downlink phase, the AP broadcasts packets to all wireless clients $M1$, $M2$, ... Thus, all clients are active to receive these

packets. Each client is polled individually in the uplink phase. Therefore, using the PSM, a client can be put into the sleeping mode when it is not polled and be waken into the active mode when it is polled. On the other hand, the MG scheme groups uplink and downlink phases for each wireless client (e.g., Disposable Token MAC protocol [72] and Learning Automata-Based Polling protocol [130]). The wake-up schedule is broadcasted in TCS at the beginning of each cycle. Each client stays in the sleeping state except when it is polled by the AP for message reception and transmission. Note that we do not include TCS in the following queueing model because of its trivial overhead. But a transition from the sleeping mode to the active mode also consumes an amount of energy.

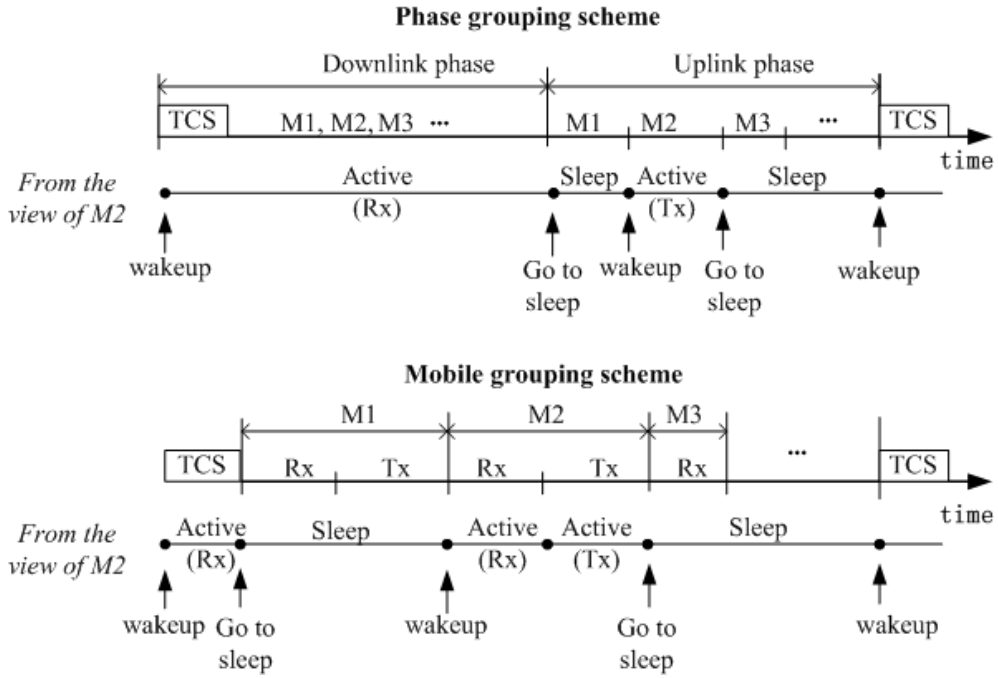


Figure 3.1. Grouping schemes of the polling-based MAC.

3.2 System Model

We model the MG and PG schemes using a cyclic-service queueing model [59]. We first describe the common model elements here and then continue the model description that differs for these two schemes. Consider that there are c wireless clients in the network, which are serviced by an AP according to a fixed polling scheme. Each wireless client is modeled as a queue with infinite buffers, denoted by q_i $(i = 1, \ldots, c)$, and the AP is modeled as a server. We use the gated service discipline in two polling schemes, since the wake-up schedule is required beforehand. This schedule wakes up the next queue and lets the server leave the polled queue when the buffered packets addressed to the polled queue at the beginning of the current polling cycle have been serviced. Packets are generated at the queues for uplink transmissions to the AP, and packets are generated at the AP for downlink transmissions to the queues. The uplink packets and downlink packets are buffered individually for each queue. The AP may sense the arrival processes of the uplink traffic and estimate the arrival processes of downlink traffic. These processes are assumed to be independent Poisson processes, with the mean arrival rates equal to $\lambda_{u,i}$ for the uplink traffic and $\lambda_{d,i}$ for the downlink traffic in q_i. The total arrival process of q_i is therefore also Poisson with rate $\lambda_i = \lambda_{u,i} + \lambda_{d,i}$. The uplink (downlink) service time processes, which are assumed to be independent and generally distributed, are denoted as $B_{u,i}(B_{d,i})$ for q_i, with mean equal to $b_{u,i}(b_{d,i}) < \infty$. Note that the mean service times are not constants because of the TPC mechanism.

In the model of the MG scheme, the server visits each queue in a deterministic and cyclic order: q_1, q_2, ..., q_c, q_1, ... Moreover, there is a nonzero walk time involved in switching between queues. During this time, the client wakes up

and prepares to communicate with the AP while the AP switches the antenna to receive or transmit packets. The walk time processes from q_{i-1} to $q_{mod(i,c)}$ denoted by SW_i are random and independent of each other, $i = 2, \ldots, c+1$. They are assumed to be generally distributed with mean $s_i < \infty$ [1]. In addition, we define several quantities that are useful for our discussion later. Consider a queue q_i in the system. From q_i's point of view, the server is either serving it or is on vacation, as shown in Figure 3.2. We define the period, in which the server is serving q_i, as q_i's *busy period*, denoted by Θ_i. The busy period is further divided into an uplink period Tx_i and a downlink period Rx_i. The uplink packets and downlink packets can be dealt with according a priority order (e.g., the downlink traffic is assigned a higher priority over the uplink traffic in Figure 3.2) or a random order on each client. We omit the trivial overhead of client's mode transition between transmission and reception in the queueing model. On the other hand, the period in which the server is away from q_i is called the *vacation period* (from q_i's viewpoint), denoted by V_i. The *cycle time* C_i is given by a sum of Θ_i and V_i.

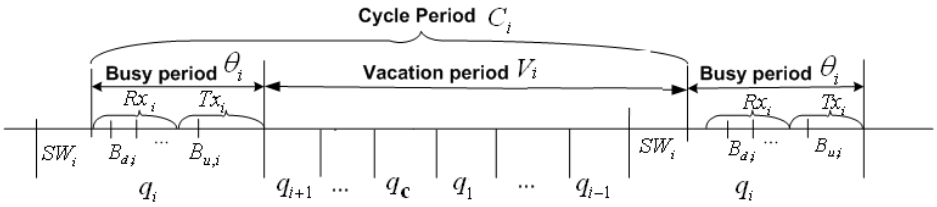


Figure 3.2. A queueing model for the MG scheme.

The model of the PG scheme depicted in Figure 3.3 is similar to the one for the MG scheme, except that we separate the downlink phase and uplink phase. The sum of all Tx_i and SW_i is regarded as the uplink phase, during which the server

[1] The walk time processes between queues (SW_i) are generally distributed, since they include the random time spent on the mode transition of clients. If SW_i are assumed to be constant, the optimization problems do not change, because the analysis results of the polling system have the same expressions [51].

visits the queues in the fixed order. The uplink periods Tx_i are defined similarly as before. The walk time process SW_i models the time spent on switching from one queue to another. Unlike the MG scheme, the AP transmits all packets with the same transmission power $P_{AP,i} \equiv P_{AP}$ $(\forall i)$. Therefore, the downlink phase is modeled as another queue with traffic generated by the AP. A walk time SW_0 is spent before all wireless clients are ready to receive data while the AP switches the antenna to broadcast, which is generally distributed with the mean s_0. The length of the downlink phase equals to the sum of the downlink period Rx and SW_0. The cycle time C is defined as the sum of the uplink phase and the downlink phase.

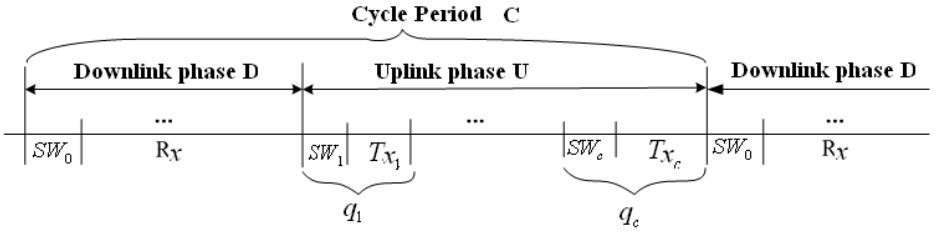


Figure 3.3. A queueing model for the PG scheme.

In Table 3.1, we present three different energy-consuming modes for the MG and PG schemes. In the active mode, we separate the transmission and reception periods. Normally, the relative power consumption for different modes are given by: $P_{V,i} \ll P_{I,i} < P_{Rx,i} \leq P_{Tx,i}$, shown as Table 2.1.

It is not difficult to see that the PG scheme consumes more energy in one cycle than the MG scheme. First, the wireless clients consume more energy in receiving downstream traffic in the PG scheme, because all clients have to stay active for the entire reception period. Second, the PG scheme has more transitions in each wireless client, and the power consumption $P_{I,i}$ is also rather high. Therefore, the MG scheme should be more energy efficient than the PG scheme. This claim,

Table 3.1. Time periods in the MG and PG schemes.

Period	**Scheme**	**Notation**	**Mode**	**Power**
q_i's transmission	MG, PG	Tx_i	Active	$P_{Tx,i}$
q_i's reception	MG	Rx_i	Active	$P_{Rx,i}$
Receiving period	PG	Rx	Active	P_{Rx}
q_i's state transition	MG	SW_i	Wake-up	$P_{I,i}$
	PG	SW_0, SW_i		
q_i's sleep period	MG	$V_i - SW_i$	Sleep	$P_{V,i}$
	PG	$C - SW_0 - Rx$ $-SW_i - Tx_i$		

however, has never been validated through a rigorous analysis. We will next optimize these two grouping schemes by TPC. Especially, we will prove that the optimal MG scheme is more energy efficient than the optimal PG scheme.

3.3 Stability-constrained Optimization Problems

Let E_i be the energy consumed by q_i during a cycle and $E = \sum_{j=1}^{c} E_j$, the total energy consumed during a cycle. We are interested in computing the average of E for the MG and PG schemes. According to Law of Large Numbers, the statistical average converges to its expectation, $\mathbf{E}[E]$. As for the power constraint, the wireless clients are assumed to operate within $[P_{min}, P_{max}]$. Besides, let the AP's transmission power when sending data to q_i be $P_{AP,i}$ ($\leq P_{MAX}$). The second set of constraints is on the client (queue) stability which is defined in the following [165].

Definition. 3.3.1. *The client q is considered* stable *if the process of queue length $L(t)$ converges to some proper distribution F, i.e.,*

$$\lim_{t\to\infty} Pr\{L(t) < x\} = F(x), \lim_{x\to\infty} F(x) = 1.$$

q_i is called substable if $\lim_{x\to\infty} \lim_{t\to\infty} \inf Pr\{L(t) < x\} = 1$.

A network is considered stable if all the nodes in the network are substable; it is considered unstable otherwise. From the results in [166], a polling system with the gated/exhaustive service discipline will be stable, if and only if the total workload $\rho = \sum_{\forall i} \lambda_i b_i$ is strictly less than 1 when the buffer size of each queue is unlimited [2]. In order to apply the standard optimization problem formulation, we relax the stability constraint to the necessary stability condition of $\rho \leq 1$. The relaxation on the stability constraint does not affect the optimization results in practice, because all the experiments show that none of the optimal solutions are on the stability boundary.

3.3.1 The Mobile Grouping Scheme

Consider q_i in the MG scheme. E_i consists of the energy consumption during the sleep state, the transition from the sleep state to the active state and the active mode (to transmit and receive data). That is, $E_{MG,i} = P_{Tx,i}Tx_i + P_{Rx,i}Rx_i + P_{I,i}SW_i + P_{V,i}(V_i - SW_i)$. Therefore, the total energy consumption of all queues in one cycle is given by $E_{MG} = \mathbf{P}_{Tx}^T\mathbf{Tx} + \mathbf{P}_{Rx}^T\mathbf{Rx} + \mathbf{P}_I^T\mathbf{SW} + \mathbf{P}_V^T(\mathbf{V} - \mathbf{SW})$, where all vectors are of dimension c. The objective function then becomes

$$\begin{aligned}\mathbf{E}[E_{MG}] &= \mathbf{E}[\mathbf{P}_{Tx}^T\mathbf{Tx}] + \mathbf{E}[\mathbf{P}_{Rx}]^T\mathbf{E}[\mathbf{Rx}] + \\ &\quad \mathbf{E}[\mathbf{P}_I]^T\mathbf{E}[\mathbf{SW}] + \mathbf{E}[\mathbf{P_V}]^T(\mathbf{E}[\mathbf{V}] - \mathbf{E}[\mathbf{SW}]).\end{aligned}$$

Moreover, $\mathbf{E}[C]$ is the same for all queues which is given by $\sum_{j=1}^{c}(\mathbf{E}[Tx_j] + \mathbf{E}[Rx_j] + \mathbf{E}[SW_j])$. We combine the stability constraint with the power constraints

[2] We may not consider the limited buffers of queues. If the buffer size of q_i is K_i, the stability criterions will change to be $\mathbf{E}[L(q_i)] \leq K_i - 1$ $(i = 1, \ldots, c)$, where $\mathbf{E}[L(q_i)] = \sum_{j=0}^{K_i} L(q_i)Pr\{L(q_i) = j\}$ is the expectation of q_i's queue length.

and then formulate the optimization problem of the MG scheme as

$$\begin{aligned} \min_{\mathbf{b_d},\mathbf{b_u}} \quad & \mathbf{E}[E_{MG}] \\ s.t. \quad & \sum_{j=1}^{c}(\lambda_{d,j}b_{d,j} + \lambda_{u,j}b_{u,j}) - 1 \leq 0. \\ & P_{AP,i} \leq P_{MAX}, \quad \forall i. \\ & P_{min} \leq P_{Tx,i} \leq P_{max}, \quad \forall i. \end{aligned} \tag{3.1}$$

3.3.2 The Phase Grouping Scheme

For the PG scheme, q_i consumes energy during the entire downlink reception period, the mode transitions, the transmission period and the sleeping period. Therefore, $E_{PG,i} = P_{Rx,i}Rx + P_{Tx,i}Tx_i + P_{I,i}(SW_0 + SW_i) + P_{V,i}(C - SW_0 - Rx - SW_i - Tx_i)$. Similar to the previous case, we can obtain the objective function for the PG scheme as

$$\begin{aligned} \mathbf{E}[E_{PG}] \quad = \quad & (\mathbf{E}[\mathbf{P}_{Rx}]^T\mathbf{E}[Rx] + \mathbf{E}[\mathbf{P}_I]^T\mathbf{E}[SW_0])e_c + \mathbf{E}[\mathbf{P}_I]^T\mathbf{E}[\mathbf{SW}] + \mathbf{E}[\mathbf{P}_{Tx}^T\mathbf{Tx}] \\ & + \mathbf{E}[\mathbf{P_V}]^T \times (\mathbf{E}[C - SW_0 - Rx]e_c - \mathbf{E}[\mathbf{SW}] - \mathbf{E}[\mathbf{Tx}]), \end{aligned}$$

where $\mathbf{E}[C] = \mathbf{E}[Rx] + \sum_{j=1}^{c}(\mathbf{E}[Tx_j] + \mathbf{E}[SW_j])$ and e_c is a c-dimensional identity vertical vector, whose elements are all one. Therefore, the optimization problem of the PG scheme is formulated as

$$\begin{aligned} \min_{\mathbf{b_d},\mathbf{b_u}} \quad & \mathbf{E}[E_{PG}] \\ s.t. \quad & \overline{b_D}\sum_{j=1}^{c}\lambda_{d,j} + \sum_{j=1}^{c}(\lambda_{u,j}b_{u,j}) - 1 \leq 0. \\ & P_{min} \leq P_{Tx,i} \leq P_{max}, \quad \forall i, \\ & P_{AP} \leq P_{MAX}. \end{aligned} \tag{3.2}$$

Alternatively, the polling schemes may also adopt the k-limited service discipline when the number of packets which are served during a visit of the server

to queue i is at most k_i. That is, the wake-up schedule lets the server proceed to the next queue when either k_i packets have been served or queue i has become empty. The objective functions using the k_i-limited service discipline are the same as the ones using the gated service discipline, since the analysis results of the polling system are the same [172], shown as Table 3.2. However, the necessary and sufficient stability conditions are different, which are $\rho < 1$ and $\frac{E[C]}{\lambda_i} < k_i$, $i = 1, \ldots, c$ [61, 172, 53]. These new stability conditions increase the number of constraints. We will study these more complicated optimization problems in the future.

3.4 Optimal Power Allocations

In this section we solve the stability-constrained optimization problems formulated in the last section. First, let F_i be a random variable for q_i's packet size in bits. According to (2.1) and $SNR_i(P_{Tx,i}) \approx \frac{P_{Tx,i}/g_i}{N_i}$ derived in [45, 46], the uplink packet service time is given by

$$B_{u,i} = \frac{F_i/\alpha W}{\log_2(1 + SNR_i(P_{Tx,i}))}, \tag{3.3}$$

where g_i $(i = 1, \ldots, c)$ denotes the gain of the wireless channel between the AP and q_i. N_i is the noise power of the wireless channel for q_i. Similarly, the downlink packet service time $B_{d,i}$ is given by

$$B_{d,i} = \frac{F_i/\alpha W}{\log_2(1 + SNR_i(P_{AP,i}))}, \tag{3.4}$$

where $SNR_i(P_{AP,i}) = \frac{P_{AP,i}/g_i}{N_i}$.

Next, we find that the objective function of total energy consumption includes the exponential functions of decision variables which are not integers, since the

transmission power and the service time has an exponential operation. Then it is difficult to rewrite the objective function into a polynomial form or provide any analytical expression of solution. In order to solve these complicated nonlinear optimization problems, we have designed the iterative algorithms to search the optimal solutions for the PG and MG scheme.

3.4.1 The Mobile Grouping Scheme

We aim to allocate $\mathbf{P}^*_{AP}$ and $\mathbf{P}^*_{Tx}$ for the AP and all wireless clients. According to (3.4) and (3.3), these two transmission power allocations are derived from the vectors of optimal service times $\mathbf{b_u}^*$ and $\mathbf{b_d}^*$ which are optimal solutions to (3.1). We minimize the energy consumed by all clients while guaranteeing the stability. To make the following analysis tractable, we make the following assumptions:

1. Let $P_{max} = P_{MAX}$ and the transmission powers in both directions are the same (i.e., $P_i \equiv P_{Tx,i} = P_{AP,i} \in [P_{min}, P_{max}], \forall i$).

2. By (3.3) and (3.4), we denote $B_i \equiv B_{u,i} = B_{d,i}$ with mean $b_i = \frac{H_i}{\log_2(1+P_i/K_i)}$, where $K_i = g_i N_i$ and $H_i = \mathbf{E}[F_i]/\alpha W$. Since $SNR \gg 1$, we have

$$P_i(b_i) \approx K_i \times 2^{H_i/b_i}, \forall i. \tag{3.5}$$

3. The power consumptions during the reception, transition and sleep are the same for all clients. That is, they are denoted simply by P_R, P_I and P_V.

According to Table 3.2 [59], $\mathbf{E}[Rx_i] = \lambda_{d,i} b_i \mathbf{E}[C]$ and $\mathbf{E}[Tx_i] = \lambda_{u,i} b_i \mathbf{E}[C]$ if we regard that q_i consists of two queues for downlink and uplink traffic individually. Since the busy period of q_i is $\mathbf{E}[\Theta_i] = (\lambda_{d,i}+\lambda_{u,i}) b_i \mathbf{E}[C]$, we have $\mathbf{E}[Rx_i] = \beta_i \mathbf{E}[\Theta_i]$ and $\mathbf{E}[Tx_i] = (1-\beta_i)\mathbf{E}[\Theta_i]$, where the ratio of downlink traffic β_i in q_i is defined

Table 3.2. Known results for cyclic-service queueing model.

Cycle period	$\mathbf{E}[C] = \frac{s}{1-\rho}$, where $s = \sum_{i=1}^{c} s_i$, $\rho_i = \lambda_i b_i$, and $\rho = \sum_{i=1}^{c} \rho_i$
Busy period	$\mathbf{E}[\Theta_i] = \rho_i \mathbf{E}[C]$.
Vacation period	$\mathbf{E}[V_i] = (1-\rho_i)\mathbf{E}[C]$.

as $\frac{\lambda_{d,i}}{\lambda_{d,i}+\lambda_{u,i}}$. We refer the two special cases of $\boldsymbol{\beta} = \mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{0}$ to *pure downlink* and *pure uplink*, respectively. After applying the well-known results of the cyclic-service queueing systems summarized in Table 3.2 and $P_{Tx,i} = P_i(b_i)$ in (3.5), the optimization problem stated in (3.1) becomes

$$\begin{aligned} \min_{\mathbf{b}} \quad & \frac{s\sum_{j=1}^{c} \rho_j[(1-\beta_j)P_{Tx,j} + \beta_j P_R]}{1-\rho} + s(P_I + P_V \frac{c-1}{1-\rho}), \\ s.t. \quad & \sum_{j=1}^{c} \lambda_j b_j - 1 \leq 0. \\ & \frac{H_i}{\log_2(P_{max}/K_i)} - b_i \leq 0, \ \forall i. \\ & b_i - \frac{H_i}{\log_2(P_{min}/K_i)} \leq 0, \ \forall i. \end{aligned} \tag{3.6}$$

Proposition. 3.4.1. *In the pure downlink case, the optimal power allocation for the MG scheme is for the AP to transmit data to all wireless nodes with the maximal power.*

Proof. In this case, $\boldsymbol{\beta} = \mathbf{1}$ and $b_{d,i} = b_i, \ \forall i$. We adopt the Karush-Kuhn-Tucker (KKT) optimality conditions [60] to find the optimal vector of the downlink mean service times $\mathbf{b}_d^*$ and the corresponding optimal power allocation $\mathbf{P}_{\mathbf{AP}}^*$. For

a Lagrangian multiplier set $\{u^* \leq 0, \mathbf{v}^* \leq \mathbf{0}\}$, we have

$$\frac{\lambda_i s[P_R + P_V(c-1)]}{(1-\sum_{j=1}^{c}\lambda_j b_{d,j}^*)^2} - u^*\lambda_i + v_i^* = 0, \ \forall i. \tag{3.7}$$

$$u^*(\sum_{j=1}^{c}\lambda_j b_{d,j}^* - 1) = 0. \tag{3.8}$$

$$v_i^*(\frac{H_i}{\log_2(P_{MAX}/K_i)} - b_{d,i}^*) = 0, \ \forall i. \tag{3.9}$$

Note that the objective function is finite if $\sum_{j=1}^{c}\lambda_j b_{d,j}^* \neq 1$; therefore from (3.8), $u^* = 0$. Moreover, according to (3.7), v_i^* is strictly less than 0. In order to satisfy (3.9), it is then necessary that $b_{d,i}^* = \frac{H_i}{\log_2(P_{MAX}/K_i)}$, $\forall i$. That is, the AP transmits with the maximal power. □

When the uplink traffic exists, the interaction between P_i and b_i is more complicated. Prop. 3.4.2 gives a necessary condition for an optimal solution. Moreover, the proposition serves as the basis for designing an iterative algorithm to solve the optimization problem, which will be discussed later in subsection 3.4.3.

Proposition. 3.4.2. *If all wireless clients do not transmit data using P_{max} and P_{min}, $\mathbf{b}^*$ must satisfy $\mathbf{L}(\mathbf{b}^*) = \mathbf{R}(\mathbf{b}^*)$ in an optimal MG scheme.*

Proof. According to the KKT conditions, $\mathbf{b}^*$ is a local optimal solution only when (3.10)-(3.13) are satisfied for a set of lagrangian multipliers $\{u^* \leq 0, \mathbf{v}^* \leq \mathbf{0}, \boldsymbol{\sigma}^* \leq \mathbf{0}\}$:

$$\frac{s\lambda_i[L_i(\mathbf{b}^*) - R_i(\mathbf{b}^*)]}{(1-\sum_{j=1}^{c}\lambda_j b_j^*)^2} - u^*\lambda_i + v_i^* - \sigma_i^* = 0, \ \forall i. \tag{3.10}$$

$$u^*(\sum_{j=1}^{c}\lambda_j b_j^* - 1) = 0. \tag{3.11}$$

$$v_i^*(\frac{H_i}{\log_2(P_{max}/K_i)} - b_i^*) = 0, \ \forall i. \tag{3.12}$$

$$\sigma_i^*(b_i^* - \frac{H_i}{\log_2(P_{min}/K_i)}) = 0, \ \forall i, \tag{3.13}$$

where $\rho_i = \lambda_i b_i$, $\rho = \sum_{j=1}^{c} \rho_j$ and

$$L_i(\mathbf{b}) = [(1-\beta_i)P_i(b_i) + \beta_i P_R](1-\rho) + P_V(c-1) + \sum_{j=1}^{c} \rho_j[(1-\beta_j)P_j(b_j) + \beta_j P_R], \quad (3.14)$$

$$R_i(\mathbf{b}) = (1-\beta_i)P_i(b_i)(1-\rho)\ln(\frac{P_i(b_i)}{K_i}). \quad (3.15)$$

Since $\sum_{j=1}^{c} \lambda_j b_j^* \neq 1$, it is easy to see from (3.11) that $u^* = 0$. If $P_{min} < P_i^* < P_{max}$, we have $v_i^* = \sigma_i^* = 0$ from (3.12) and (3.13). Therefore, we obtain $L_i(\mathbf{b}^*) - R_i(\mathbf{b}^*) = 0\ \forall i$ from (3.10). □

3.4.2 The Phase Grouping Scheme

According to Lemma 3.2. in [104], we can model the downlink phase in the PG scheme as an M/G/1 queue q_0 with the arrival rate $\lambda_D = \sum_{i=1}^{c} \lambda_{d,i}$ and the mean transmission time $\overline{b_D} = (1/\lambda_D)\sum_{i=1}^{c} \lambda_{d,i} b_{d,i}$, where $b_{d,i} = \frac{H_i}{\log_2 P_{AP}/K_i}$ $(\forall i)$ and $P_{AP} \leq P_{MAX}$. In this case, the total average walk time is $s' = s_0 + \sum_{j=1}^{c} s_j$, and the expected cycle period is $\mathbf{E}[C] = \frac{s'}{1-\rho'}$, where the total workload is $\rho' = \sum_{j=1}^{c} \rho_{u,j} + \rho_D$, $\rho_{u,i} = \lambda_{u,i} b_{u,i}$ and $\rho_D = \lambda_D \overline{b_D}$. Furthermore, the expected receiving period is the busy period of q_0, $\mathbf{E}[Rx] = \rho_D \mathbf{E}[C]$. The expected transmission period of q_i is $\mathbf{E}[Tx_j] = \rho_{u,j}\mathbf{E}[C]$. Using (3.4), $P_{Tx,i} = K_i \times 2^{H_i/b_{u,i}}$ derived from (3.3) and the results in Table 3.2, the optimization problem for the PG scheme becomes

$$
\begin{aligned}
\min_{\mathbf{b_d},\mathbf{b_u}} \quad & (P_I - P_V)(cs_0 + s) + \frac{s_0 + s}{1 - \rho'}[\sum_{j=1}^{c} P_{Tx,j}\rho_{u,j} + \\
& c(P_R - P_V)\rho_D + P_V \sum_{j=1}^{c}(1 - \rho_{u,j})]. \\
s.t. \quad & \lambda_D \overline{b_D} + \sum_{j=1}^{c} \lambda_{u,j} b_{u,j} - 1 \leq 0. \\
& \frac{H_i}{\log_2(P_{max}/K_i)} - b_{u,i} \leq 0, \ \forall i. \\
& b_{u,i} - \frac{H_i}{\log_2(P_{min}/K_i)} \leq 0, \ \forall i. \\
& \frac{H_i}{\log_2 P_{MAX}/K_i} - b_{d,i} \leq 0, \ \forall i.
\end{aligned} \tag{3.16}
$$

We introduce the downlink ratio $\beta_i = \lambda_{d,i}/\lambda_i$ as before. Then we have $\lambda_D = \sum_{j=1}^{c} \beta_j \lambda_j$ and $\lambda_{u,i} = (1-\beta_i)\lambda_i$, where λ_i is the total traffic sent from and received by q_i. Similar to the MG scheme, we have two special cases: $\boldsymbol{\beta} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{1}$. There are therefore $2c$ decision variables (i.e., $b_{d,i}$ and $b_{u,i}, i = 1, \ldots, c$) involved in the optimization formulation. We can reduce the number of decision variables to c by considering the result in Prop. 3.4.3. As a result, the last constraint in (3.16) can be removed.

Proposition. 3.4.3. *In an optimal PG scheme, the AP transmits data with its maximal power P_{MAX}.*

Proof. Note that $\mathbf{E}[E_{PG}]$ increases monotonically with $\overline{b_D}$, because both ρ_D and ρ' increase with $\overline{b_D}$. Therefore, the optimal downlink power allocation is obtained by $\overline{b_D^*} = (1/\lambda_D)\sum_{j=1}^{c} \lambda_{d,j} b_{d,j}^*$, where $b_{d,i}^* = \frac{H_i}{\log_2 P_{MAX}/K_i}, \forall i$. That is, the AP uses the maximal power to transmit data. □

Similar to the MG scheme, we use the KKT conditions to obtain the necessary condition for an optimal PG scheme, which is stated in Prop.3.4.4.

Proposition. 3.4.4. *If all wireless clients do not transmit data using P_{max} and P_{min}, $\mathbf{b_u}^*$ must satisfy $\mathbf{L}(\mathbf{b_u^*}) = \mathbf{R}(\mathbf{b_u^*})$ in an optimal PG scheme.*

Proof. Let $\rho_D^* = \overline{b_D^*}\lambda_D$. $\mathbf{b_u}^*$ is a local optimal solution only when (3.17)-(3.20) are satisfied for a set of lagrangian multipliers $\{u^* \leq 0, \mathbf{v}^* \leq \mathbf{0}, \boldsymbol{\sigma}^* \leq \mathbf{0}\}$. Let $L_i(\mathbf{b_u}) = P_{Tx,i}(b_{u,i})(1-\rho) + \sum_{j=1}^{c} \rho_{u,j} P_{Tx,j}(b_{u,j}) + P_V(c - 1 + \rho_D)$ and $R_i(\mathbf{b_u}) = P_{Tx,i}(b_{u,i})(1-\rho)\ln(\frac{P_{Tx,i}(b_{u,i})}{K_i})$, $\forall i$. The KKT conditions are given by:

$$\frac{\mathbf{E}[C]\lambda_{u,i}[L_i(\mathbf{b_u^*}) - R_i(\mathbf{b_u^*})]}{(1 - \sum_{j=1}^{c} \lambda_j b_{u,j}^* - \rho_D^*)^2} - u^*\lambda_{u,i} + v_i^* - \sigma_i^* = 0, \ \forall i. \tag{3.17}$$

$$u^*(\rho_D^* + \sum_{j=1}^{c} \lambda_{u,j} b_{u,j}^* - 1) = 0, \tag{3.18}$$

$$v_i^*(\frac{H_i}{\log_2(P_{max}/K_i)} - b_{u,i}^*) = 0, \ \forall i. \tag{3.19}$$

$$\sigma_i^*(b_{u,i}^* - \frac{H_i}{\log_2(P_{min}/K_i)}) = 0, \ \forall i. \tag{3.20}$$

Obviously, $u^* = 0$ according to (3.18). If $P_{min} < P_i^* < P_{max}$, we have $v_i^* = \sigma_i^* = 0$ from (3.19) and (3.20). As a result, we have $L_i(\mathbf{b_u^*}) - R_i(\mathbf{b_u^*}) = 0, \ \forall i$ from (3.17). □

Proposition. 3.4.5. *In the pure downlink case, $\mathbf{E}[E_{PG}] > \mathbf{E}[E_{MG}]$ with the same downlink transmission rates.*

Proof. Proof: When $\boldsymbol{\beta} = \mathbf{1}$, we have $b_{u,i} = 0, \forall i$ and $\rho = \rho_D$. Using the same $b_{d,i}$, $\mathbf{E}[E_{PG}] - \mathbf{E}[E_{MG}]$ is given by

$$cP_I s_0 + \frac{\rho_D}{1-\rho_D}[(cP_R - P_V)s_0 + ((c-1)P_R - cP_V)s] > 0,$$

because $P_R \gg P_V$ in practice. In particular, $\mathbf{E}[E_{PG}^*] > \mathbf{E}[E_{MG}^*]$ with the same optimal transmission rates, because the AP always transmits data with P_{MAX} according to Prop.3.4.1 and Prop.3.4.3. □

Prop.3.4.5 shows that the MG scheme always consumes less energy per cycle than the PG scheme in the pure downlink case. In subsection 3.5.2, we show the advantage of the optimal MG schemes in a general case.

3.4.3 An Iterative Algorithm

An optimization problem is classified based on whether it is continuous or discrete, linear or nonlinear and constrained or unconstrained. According to the definitions in [22], the optimization problems (3.6) and (3.16) are continuous, constrained and nonlinear. The c-dimensional decision variable $\mathbf{b}$ or $\mathbf{b_u}$ are continuous, and the constraints are linear inequalities. These two problems are nonlinear, since their objective functions are nonlinear. Furthermore, the objective function of total energy consumption includes the reciprocal and the exponential functions of decision variable's elements, for example, $P_{Tx,i}(b_i) = K_i \times 2^{H_i/b_i}$ and $\frac{1}{1-\rho} = 1/(1 - \sum_{i=1}^{c} \lambda_i b_i)$ in (3.6). Therefore, we could not change these complicated objective functions into any polynomial forms or solve them by any analytical method applied in some linear optimization problems.

In order to solve these two optimization problems, we adopt the theory of Lagrange multipliers from the feasible direction viewpoint [22]. Using lagrange multipliers, Prop.3.4.2 and Prop.3.4.4 provide Karush-Kuhn-Tucker necessary conditions for the MG and PG schemes, respectively. Instead of strictly using the feasible direction method, we have designed the iterative algorithms based on these two propositions to quickly search the optimal solutions of $\mathbf{b}^*$ or $\mathbf{b_u}^*$. This algorithm uses the property of monotonic function to simplify the direction searching and applies Gauss-Siedel iteration to speed up the operation. In the following we use the MG scheme as an example to illustrate how the iterative

algorithms works, which is given in Figure 3.4.

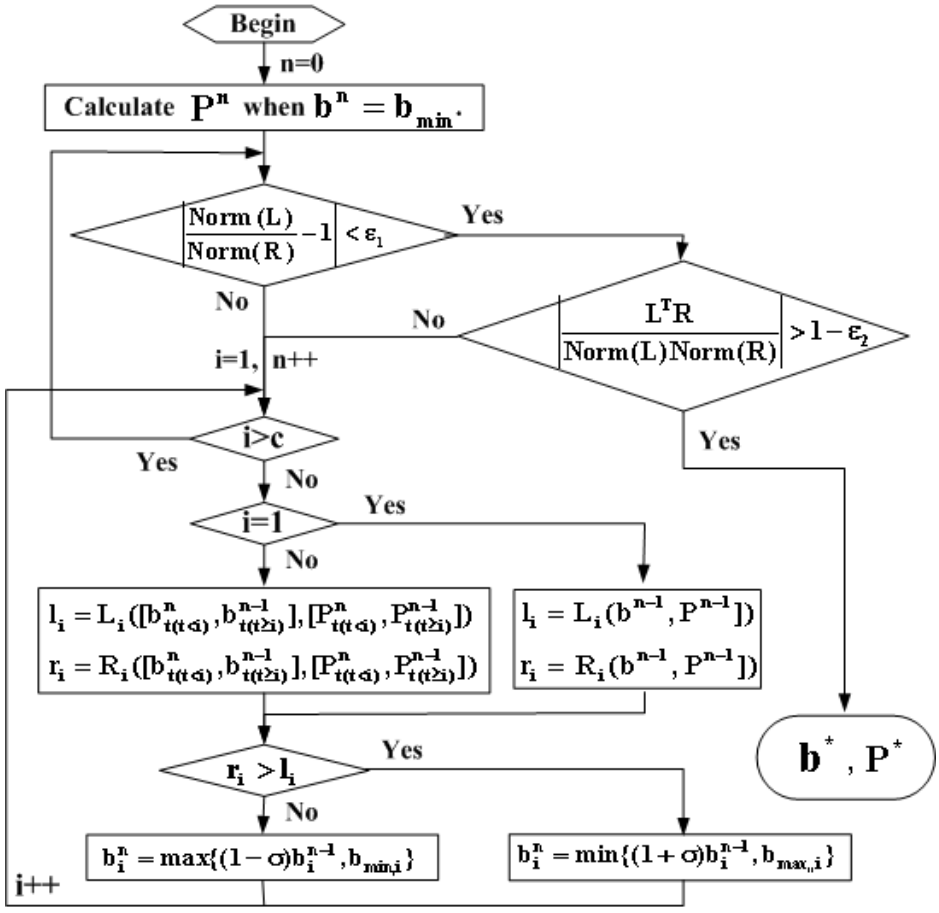
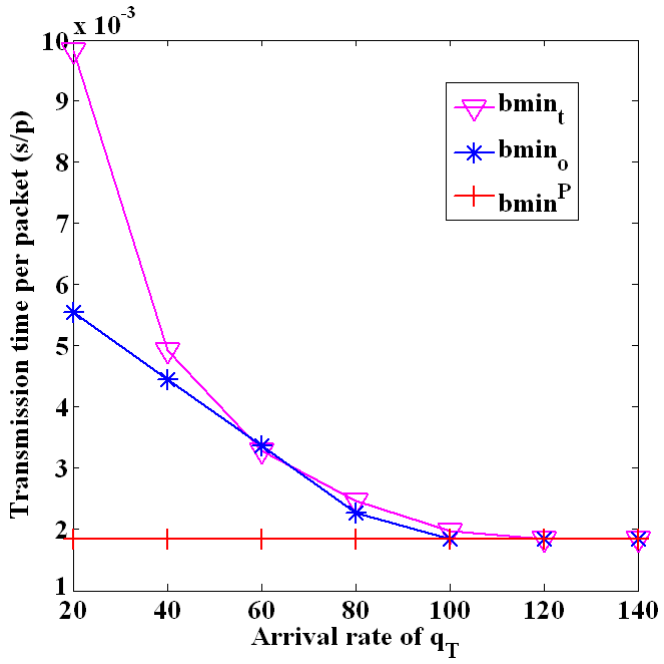


Figure 3.4. An iterative algorithm for the MG scheme.

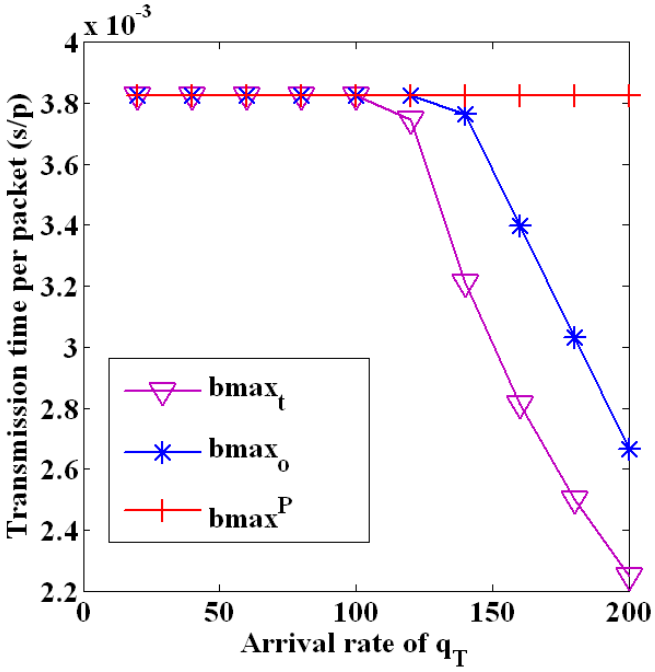
We first determine the feasible region of the mean service times, which is bounded by $\mathbf{b}_{min}$ and $\mathbf{b}_{max}$. According to the power constraints $[P_{min}, P_{max}]$, the range of the decision variables must be within $[b^P_{min,i}, b^P_{max,i}]$ $\forall i$ derived from (3.5). According to the stability constraint, we can obtain the minimum (maximum) average service time for any q_i, when other queues adopt $b^P_{max,j}$ $(b^P_{min,j})$, $j \neq i$. That is $b^\rho_{min,i} = (1 - \sum_{j\neq i} \lambda_j b^P_{max,j})/\lambda_i$ $(b^\rho_{max,i} = (1 - \sum_{j\neq i} \lambda_j b^P_{min,j})/\lambda_i)$. By combining the stability constraint with the power constraints, the feasible region is therefore given by $b_{min,i} = \max\{b^\rho_{min,i}, b^P_{min,i}\}$ and $b_{max,i} = \min\{b^\rho_{max,i}, b^P_{max,i}\}$.

Obviously, the traffic arrival rates and power limitations greatly influence the feasible region. First, we fix $[P_{min},\ P_{max}]$. When $\boldsymbol{\lambda}$ increases, $b_{min,i}$ is more likely to be determined by P_{max}, but $b_{max,i}$ is more likely to be determined by the stability constraint instead of P_{min}. The experiment result in Figure 3.5 validates this statement. Consider a tagged queue q_T and other identical queues. The arrival rate of q_T is increased while that of other queues are kept at 50p/s. As shown in Figure 3.5(a), when λ_{q_T} increases, the minimal service time of q_T ($b_{min,t}$) and that of other queues ($b_{min,o}$) converge to b^P_{min} which is determined by P_{max}. In Figure 3.5(b), when λ_{q_T} is small, the maximal service times of q_T ($b_{max,t}$) and that of other queues ($b_{max,o}$) is the same as b^P_{max} determined by the P_{min}. However, when λ_{q_T} increases, they go further away from b^P_{max}, since they are constrained by stability. Second, we fix the traffic arrival rates $\boldsymbol{\lambda}$. When P_{min} (P_{max}) reduces, $b_{min,i}$ and $b_{max,i}$ still have the same character as before.

Let $\mathbf{b}^n$ be the vector of the mean service times at the beginning of nth iteration. We set $\mathbf{b}^0 = \mathbf{b}_{min}$, and $\mathbf{P}^0_{Tx}$ can be computed from (3.5), when $n = 0$. From Prop.3.4.2, the ideal terminating condition is $\mathbf{L}(\mathbf{b}^n) = \mathbf{R}(\mathbf{b}^n)$. To obtain a close-enough condition, we exploit the fact that two vectors in the Euclidean space are the same if their norms are the same and the angle between them is zero. Therefore, given the accuracy tolerance parameters ε_1 and ε_2, the algorithm terminates when $|\frac{Norm(\mathbf{L})}{Norm(\mathbf{R})} - 1| < \varepsilon_1$ and $|\frac{\mathbf{L}^T\mathbf{R}}{Norm(\mathbf{L})Norm(\mathbf{R})}| > 1 - \varepsilon_2$. When ε_1 is near zero, the norms of $\mathbf{L}$ and $\mathbf{R}$ should be very close to each other. When ε_2 is near zero, $\cos\angle(\mathbf{L}, \mathbf{R})$ should be close to 1. Inside the iterative loop, R_i decreases monotonically with b_i (according to (3.15)). Therefore, in the nth iteration for q_i, if $R_i > L_i$, we increase b_i^{n-1} by $d \in (0, 1)$, the step size; otherwise, b_i^n is decreased by d. The middle results of $\mathbf{b}$ and $\mathbf{P}$ after each iteration must be recorded. Although the algorithm's complexity has not been proven, our ex-



(a) b_{min} in the feasible region.



(b) b_{max} in the feasible region.

Figure 3.5. Impacts of stability and power constraints on the feasible region.

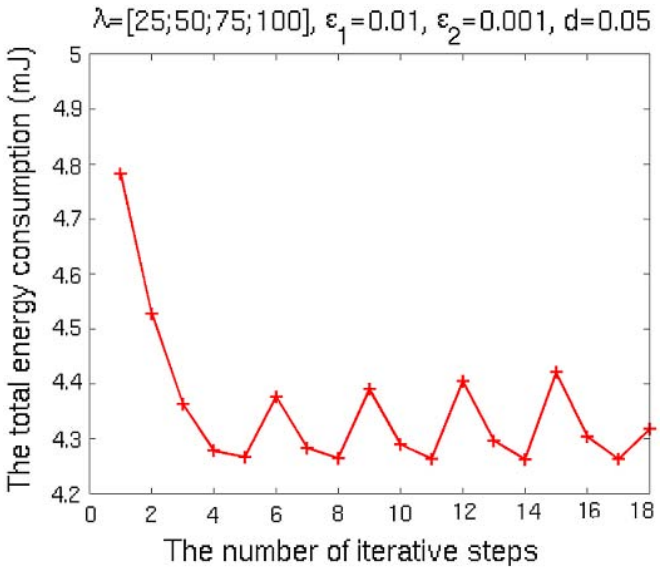
tensive experiment results show that the algorithm converges quite quickly. In most cases, $n \leq 50$. Particularly, it works well when the traffic arrival intensity is small (i.e., $\boldsymbol{\lambda}^T \mathbf{b}_{min} < 0.5$). We have also applied the Gauss-Siedel iteration to reduce the convergence time.

When $c = 2$, both objective functions are convex within the stability area, as shown in Figure 3.7. In this case the iterative algorithm gives a single optimal solution and then $\mathbf{b}^* = \mathbf{b}^n$, $\mathbf{P}^* = \mathbf{P}^n$. However, the objective functions are not convex for $c > 2$. As a result, the algorithm may return more than one local optimal solution, shown as Figure 3.6. For these cases we select the one, whose value of the objective function is smallest, to be the optimal solution. Thus, $\mathbf{b}^* = \mathbf{b}^t$ and $\mathbf{P}^* = \mathbf{P}^t$ where $t = \arg\min_{i \in \{0,1,\ldots,n\}} \mathbf{E}[E(\mathbf{b}^i)]$. As shown in Figure 3.6(a), for example, the iterative algorithm reaches several local minimums for the MG scheme, and the final optimal solution is obtained at the 14th iterative step (i.e., $\mathbf{P}^* = \mathbf{P}^{14}$).

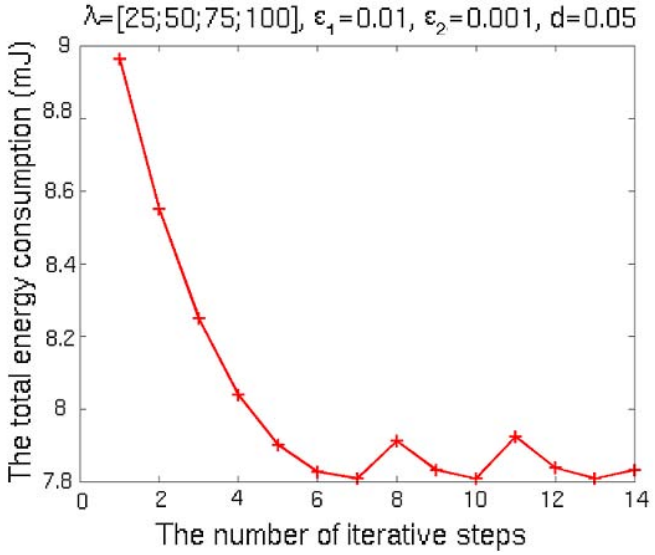
The PG scheme has a similar iterative algorithm as the MG scheme, but has different expressions of $\mathbf{L}(\mathbf{b})$ and $\mathbf{R}(\mathbf{b})$. The feasible region $[\mathbf{b}_{min}, \mathbf{b}_{max}]$ is given by $b_{min,i} = \max\{(1 - \rho_D^* - \sum_{j \neq i} \lambda_j b_{max,j}^P)/\lambda_i, b_{min,i}^P\}$ and $b_{max,i} = \min\{(1 - \rho_D^* - \sum_{j \neq i} \lambda_j b_{min,j}^P)/\lambda_i, b_{max,i}^P\}$. Moreover, the PG optimal solutions include $\mathbf{b_u}^*$ which determines $\mathbf{P}_{Tx}^*$ for the wireless nodes, as well as $\mathbf{P}_{AP}^* = P_{MAX} \times e_c$ for the AP. Figure 3.6(b) shows the steps executed by the iterative algorithm, where $\mathbf{P_{Tx}}^* = \mathbf{P}^{13}$.

3.5 Performance Evaluation

The experiment results presented in this section are based on $W = 2MHz$, $P_R = 2W$, $P_I = 1W$ and $P_V = 0.05W$. The power constraints are given by



(a) MG scheme for $\boldsymbol{\beta} = 0.5e_4$.



(b) PG scheme for $\boldsymbol{\beta} = 0.5e_4$.

Figure 3.6. Energy consumption at iterative steps.

$P_{MAX} = P_{max} = 10W$ and $P_{min} = 1W$. We set these parameters by making reference to the experiment results in [4, 10].

3.5.1 Evaluation: Model and Algorithm

We first evaluate the performance of the iterative algorithm proposed in the last section. We will show that the algorithm can compute power allocations that are very close to the optimal ones. For the purpose of estimating the optimal power allocations, we have employed an exhaustive search method which divides the range of each average service time $[b_{min,i}, b_{max,i}]$ $\forall i$ evenly into t parts. For all the possible t^c vectors of average service times, we first select those that satisfy the stability requirement and then find the one with the minimal energy consumption. Since the method is very time consuming, we have carried it out only for $c = 1, \ldots, 5$.

We first consider a case of 2 queues in Table 3.3 for both schemes. The results obtained from the iterative algorithm are marked by A ($d = 0.05$, $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.001$). The results obtained from the exhaustive search method are indicated by N. We compare the optimal vectors of the transmission power obtained from two methods by computing an differential rate $e_P = \frac{norm(\mathbf{P}^*_{\mathbf{A}} - \mathbf{P}^*_{\mathbf{N}})}{norm(\mathbf{P}^*_{\mathbf{N}})} \times 100\%$. The results show that the iterative algorithm yields optimal solutions that are very close to, if not the same as, the exact ones obtained by the exhaustive search. Besides the optimal results, the table also shows the number of iterative steps required by the iterative algorithm (denoted as f) and the loc^{th} iteration when the optimal solution is obtained. When $loc = m$, $\mathbf{b_u}$ is determined by $\mathbf{P}_{min}$ or $\mathbf{P}_{max}$. When $loc = m - 1$, the iteration stops after the first local solution is found. That is, the f^{th} step leads to a higher energy consumption. Therefore, for the MG and PG scheme, the single local solution is the global optimal, and the objective function is convex when $c = 2$.

The results of $c = 3, 4, 5$ are given in Table 3.4 and Table 3.5 for the MG

Table 3.3. Energy efficiency results of the MG and PG schemes for $c = 2$, $\mathbf{E}[\mathbf{F}] = [1024; 512]$bytes, $[N_1, N_2] = [0.02, 0.01]$W, $[g_{11}, g_{12}] = [6, 8]$ and $\alpha = 0.7$.

	$\boldsymbol{\lambda}(p/s)$	$\boldsymbol{\beta}$	$[b_1^*; b_2^*](ms/p)$	$\mathbf{P}^*(W)$	$e_P(\%)$	ρ^*	f	loc	$\mathbf{E}[E^*]$(mJ)
MG (N)	[30;60]	[0;0]	[7.6517;3.2117]	[1.000,1.000]	–	0.4223	–	–	0.7270
MG (A)	[30;60]	[0;0]	[7.6517;3.2117]	[1.000,1.000]	0	0.4223	16	16	0.7270
MG (N)	[60;100]	[0.3;0.3]	[5.3014;2.4305]	[2.5600;2.2518]	–	0.5611	–	–	1.6211
MG (A)	[60;100]	[0.3;0.3]	[5.3121;2.4330]	[2.5443;2.2442]	0.5100	0.5620	21	20	1.6211
MG (N)	[60;150]	[0.6;0.6]	[4.3852;2.0170]	[4.8520;4.4637]	–	0.5657	–	–	2.0389
MG (A)	[60;150]	[0.6;0.6]	[4.3922;2.0117]	[4.8234;4.5116]	0.8500	0.5653	14	13	2.0391
	$\boldsymbol{\lambda}(p/s)$	$\boldsymbol{\beta}$	$[b_{u,1}^*; b_{u,2}^*](ms/p)$	$\mathbf{P}^*(W)$	$e_P(\%)$	ρ^*	f	loc	$\mathbf{E}[E^*]$(mJ)
PG (N)	[30;60]	[0;0]	[7.6517;3.2117]	[1.000,1.000]	–	0.4223	–	–	1.2804
PG (A)	[30;60]	[0;0]	[7.6517;3.2117]	[1.000,1.000]	0	0.4223	16	16	1.2804
PG (N)	[60;100]	[0.3;0.3]	[6.0981;2.7828]	[1.7163;1.4759]	–	0.5673	–	–	2.5054
PG (A)	[60;100]	[0.3;0.3]	[6.0629;2.7769]	[1.7430;1.4851]	1.2500	0.5654	18	17	2.5054
PG (N)	[60;150]	[0.6;0.6]	[5.6201;2.5684]	[2.1521;1.8825]	–	0.5722	–	–	3.2436
PG (A)	[60;150]	[0.6;0.6]	[5.6057;2.5675]	[2.1682;1.8847]	0.5700	0.5718	19	18	3.2436

Table 3.4. Energy efficiency results for the MG scheme for $c > 2$.

	$\boldsymbol{\lambda}(packets/s)$	$\mathbf{b}^*(ms/packet)$	$\mathbf{P}^*(W)$	$e_P(\%)$	ρ^*	$\rho^{(0)}$	$\mathbf{E}[E^*]$(mJ)
N	[30;60;90]	[4.9960;2.2154;2.2927]	[3.0340;2.6753;2.7521]	–	0.4891	–	2.0997
A	[30;60;90]	[5.0087;2.2106;2.2940]	[3.0612;2.6156;2.7464]	1.3500	0.4894	0.3584	2.0997
N	[60;100;140]	[4.0096;1.8156;1.8551]	[7.0220;5.8684;6.2119]	–	0.6813	–	6.1787
A	[60;100;140]	[4.0441;1.8050;1.8731]	[6.6286;5.9663;6.0794]	3.8600	0.6854	0.6172	6.1784
N	[25;50; 75;100]	[4.6242;2.0594; 2.0594;2.1089]	[4.0075;3.4242; 3.4242;3.7462]	–	0.5839	–	4.2634
A	[25;50;75;100]	[4.6118;2.0354; 2.0354;2.1123]	[4.0455;3.5870; 3.5870;3.7234]	3.2000	0.5809	0.4621	4.2625
N	[50;100;100;150]	[3.6681;1.6189;	[10;10;10;10]	0	0.7592	0.7592	16.4317
A		1.6189;1.6800]					
N	[5;10;15;20;25]	[5.7982;2.5590;2.5590;	[1.9695;1.5870;1.5870;	0	0.2105	0.1248	1.7433
A		2.5590;2.6556]	1.5870;1.6969]				
N	[20;30;50;70;90]	[4.4648;2.0319;2.0319; 2.0319;2.1395]	[4.5420;3.6119;3.6119; 3.6119;3.5454]	–	0.5866	–	5.5509
A	[20;30;50;70;90]	[4.5658;2.0151;2.0151; 2.0151;2.0912]	[4.1912;3.7338;3.7338; 3.7338;3.8701]	5.1441	0.5818	0.4382	5.5472

Table 3.5. Energy efficiency results of the PG scheme for $c > 2$.

	$\boldsymbol{\lambda}(packets/s)$	$\mathbf{b_u}^*(ms/packet)$	$\mathbf{P}^*_{Tx}(W)$	$e_P(\%)$	ρ^*	$\rho^{(0)}$	$\mathbf{E}[E^*]$(mJ)
N	[30;60;90]	[6.0583;2.6284;2.7522]	[1.7466;1.4597;1.5245]	–	0.4728	–	3.6379
A	[30;60;90]	[5.9750;2.6370;2.7366]	[1.8121;1.4441;1.5503]	2.6600	0.4711	0.3584	3.6377
N	[60;100;140]	[4.4648;2.0319;2.0885]	[4.5420;3.6119;3.8897]	–	0.6903	–	10.1305
A	[60;100;140]	[4.5430;2.0051;2.0808]	[4.2666;3.8100;3.9462]	4.9200	0.6908	0.6172	10.1291
N	[25;50;75;100]	[5.1022;2.2796; 2.2796;2.3540]	[2.8848;2.3404; 2.3404;2.5100]	–	0.5550	–	7.8092
A	[25;50;75;100]	[5.1100;2.2553; 2.2553;2.3404]	[2.8708;2.4321; 2.4321;2.5605]	2.7700	0.5529	0.4621	7.8089
N	[50;100;100;150]	[3.6681;1.6189;	[10;10;10;10]	0	0.7592	0.7592	30.3315
A		1.6189;1.6800]					
N	[5;10;15;20;25]	[7.5121;2.9954;2.9954;	[1;1;1;1;1]	0	0.1929	0.1248	3.4167
A		2.9954;3.2117]					
N	[20;30;50;70;90]	[4.9960;2.1695;2.1695; 2.1695;2.2927]	[3.0866;2.8037;2.8037; 2.8037;2.7521]	–	0.5495	–	10.9766
A	[20;30;50;70;90]	[4.9157;2.1695;2.1695; 2.1695;2.2514]	[3.2548;2.8038;2.8038; 2.8038;2.9365]	3.9131	0.5469	0.4382	10.9749

scheme and the PG scheme, respectively. Note that $\boldsymbol{\beta} = k * e_c$. For example, the case of $k = 0.5$ represents a 50-50 mix of the uplink and downlink traffic. Once again the iterative algorithm yields quality solutions. Besides, if $\boldsymbol{\lambda}$ is small, then the initial workload $\boldsymbol{\rho}^{(0)} = \boldsymbol{\lambda}^T \mathbf{b}_{min}$ is very small. The final optimal solution is determined by $\mathbf{P}_{min}$ when the optimal workload $\boldsymbol{\lambda}^T \mathbf{b}_{max}$ is also small. For example, in the 5-queue system with a light workload, the optimal transmission power allocation for the PG scheme is [1;1;1;1;1]. However, if the traffic rates are very high, $\boldsymbol{\rho}^{(0)}$ is close to $\mathbf{1}$. Then the wireless clients should send packets as fast as possible, i.e., the packets are transmitted with the maximal power allowed. For example, the optimal power allocation for the MG scheme is $\mathbf{P}_{max}$ when $\boldsymbol{\lambda} = [50; 100; 100; 150]$.

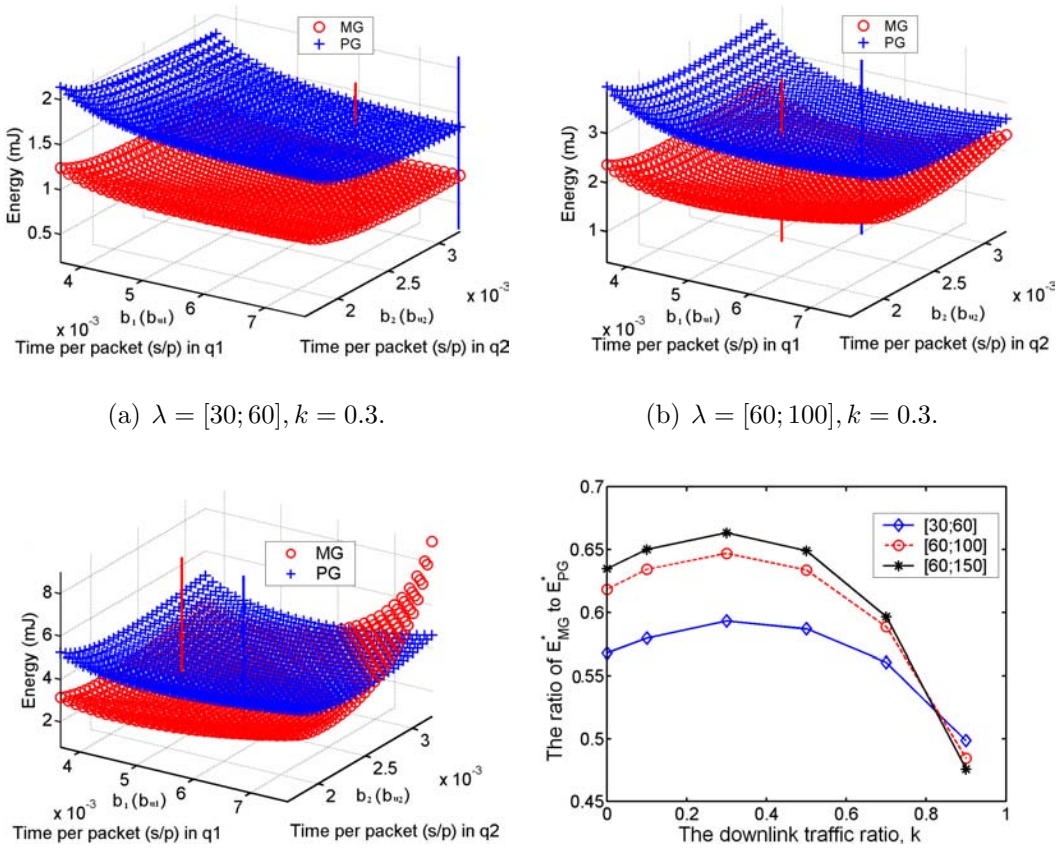
3.5.2 A Comparison of the MG and PG Schemes

Based on Table 3.3-3.5, $\mathbf{E}[E^*_{PG}] > \mathbf{E}[E^*_{MG}]$ and $\overline{E_{PG}}^* > \overline{E_{MG}}^*$ in all cases. Next, we present that the MG schemes is more energy efficient, since it consumes less energy than the PG scheme when the system stably operates under optimal transmission power allocations.

Energy Consumption

Figure 3.7 presents the energy consumption obtained from an exhaustive search method in 3 different workload scenarios: light (Fig 3.7(a)), moderate (Fig 3.7(b)) and heavy (Fig 3.7(c)). For both schemes, $\mathbf{E}[E]$ is shown to be a convex function of the average service time when $c = 2$. The vertical lines in these figures indicate the optimal solutions. Moreover, when the traffic arrival rate is too low or too high in reference to the power limitations (e.g., $\boldsymbol{\rho}^{(0)} < 0.2$

or $\boldsymbol{\rho}^{(0)} > 0.75$), the optimal service times could be determined by P_{min} or P_{max}. On the contrary, the optimal transmission times are determined by P_{min}. For example, $\mathbf{P}^* = [1; 1]$ in the PG scheme when $\boldsymbol{\lambda} = [30; 60]$ and $\boldsymbol{\beta} = 0.3 \times e_2$.



(a) $\lambda = [30; 60], k = 0.3$.

(b) $\lambda = [60; 100], k = 0.3$.

(c) $\lambda = [60; 150], k = 0.3$.

(d) E^*_{MG}/E^*_{PG} verses k.

Figure 3.7. Comparing energy efficiency for the PG and MG schemes for $c = 2$.

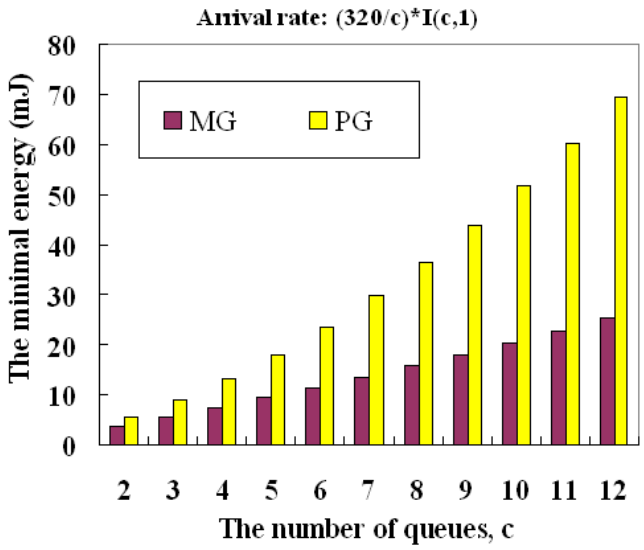
In most cases, $\mathbf{E}[E_{MG}] < \mathbf{E}[E_{PG}]$ within the feasible region, except when $\boldsymbol{\lambda}$ is high enough. For example, in Figure 3.7(c), $\mathbf{E}[E_{MG}] > \mathbf{E}[E_{PG}]$ when the elements of $\mathbf{b_u}$ are close to the upper boundary. The reason is that the MG scheme approaches much closer to its stability boundary which leads to a very long cycle period. In this case, the MG scheme consumes much more energy due to the long cycle. Furthermore, we repeat the experiments with different $\boldsymbol{\beta}$ by changing k

from 0 to 0.9, and the values of $\frac{\mathbf{E}[E^*_{MG}]}{\mathbf{E}[E^*_{PG}]}$ are given in Figure 3.7(d). As shown, the optimal MG scheme always has the lower energy consumption than the optimal PG scheme since $\frac{\mathbf{E}[E^*_{MG}]}{\mathbf{E}[E^*_{PG}]} < 1$. Here, they both satisfy the stability requirement. Therefore, the MG scheme is indeed more energy-efficient, especially when the proportion of the downlink traffic is large.

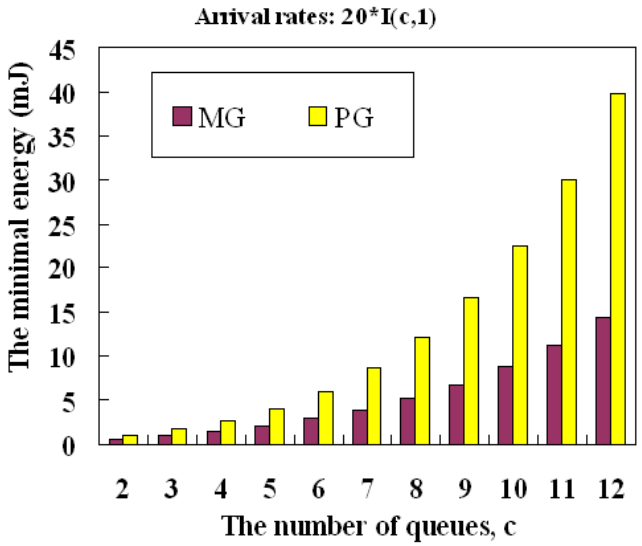
Moreover, we use the iterative algorithm to compare the energy efficiency of the MG and PG schemes for $c > 2$ in symmetric systems. Figure 3.8 compares the minimal energy consumption for two schemes with different traffic patterns. Again, they show that the optimal MG scheme can save more energy than the optimal PG scheme. Furthermore, the gap between the two schemes increases with the number of clients.

Energy Efficiency Metric

To probe further, we compare the two grouping schemes based on a bit-per-joule metric z, which is used to measure the energy efficiency. A possible method is to obtain their optimal power allocations which maximize z under the MG and PG schemes. Unfortunately, optimizing the grouping schemes directly based on z is extremely complicated. Therefore, we instead obtain optimal schemes which minimize the total energy consumption of all clients consumed in a cycle (i.e., the solutions to (3.6) and (3.16)). We then compare these optimal schemes based on z, and the z of an optimal scheme is defined as the ratio of the amount of data which arrives at all clients within one cycle period to the total energy consumption of all clients, $z = (\sum_{i=1}^{c} \lambda_i F_i)\mathbf{E}[C]/\mathbf{E}[E]$. We denote z_{MG} and z_{PG} for the MG and PG schemes, respectively. As we will show, the optimal MG scheme outperforms the PG optimal scheme (i.e., $z_{MG} > z_{PG}$). The following comparison results are based on symmetric systems in which the distributions of



(a) $\sum_{i=1}^{c} \lambda_i = 320$ for $k = 0.5$.



(b) $\lambda_i = 20\ \forall i$ for $k = 0.5$.

Figure 3.8. Minimal energy consumption verses c

inter-frame arrival times and packet sizes are the same for all queues.

Figure 3.9 presents the comparison results for the case of $k = 0.5$. The figure shows that the optimal MG scheme is more energy efficient than the optimal PG scheme because of $z_{MG}/z_{PG} > 1$ for all cases. Moreover, z_{MG}/z_{PG} increases with

c and l. That is, the energy efficiency gap between two optimal schemes increases with l and c. Furthermore, we repeat the experiments with different downlink-uplink traffic mix, and the results are shown in Table 3.6. The comparison results remain the same for these cases. Besides, the gap becomes more outstanding when there is a higher proportion of downlink traffic. That is, z_{MG}/z_{PG} increases with c for a given k, and z_{MG}/z_{PG} increases with k for a given c.

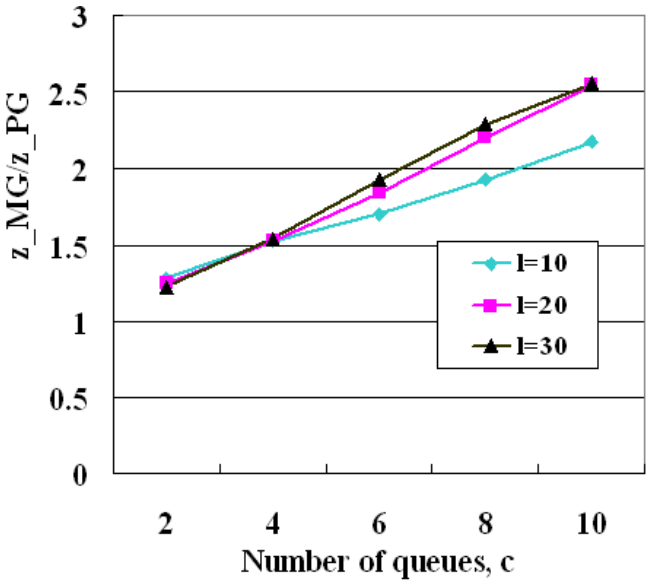


Figure 3.9. An energy efficiency comparison for $\boldsymbol{\lambda}$=$l \times e_c$ and $\boldsymbol{\beta}$=$0.5e_c$.

Table 3.6. The ratio z_{MG}/z_{PG} for $\boldsymbol{\beta}$=$k \times e_c$ and $\boldsymbol{\lambda}$=$20 \times e_c$.

c	2	4	6	8	10
k=0.2	1.26	1.42	1.53	1.67	1.80
k=0.5	1.25	1.52	1.84	2.20	2.55
k=0.8	1.28	1.76	2.22	2.80	3.44

Why is the MG Scheme More Energy Efficient?

In this subsection, we identify the factors that are responsible for the MG scheme's higher energy efficiency. To this end, we decompose the total energy

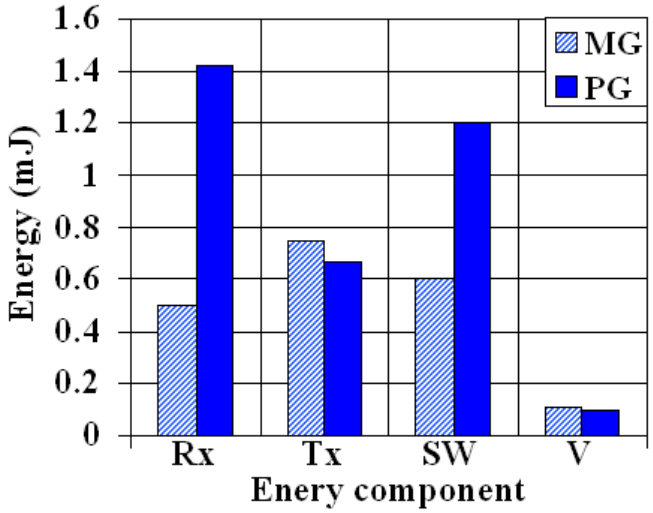
consumption into 4 components—energy due to receptions (Rx), transmissions (Tx), mode transitions (SW) and sleeping (V). We plot the results for both schemes in Figure 3.10(a) when $c = 3$. The figure clearly shows that the MG scheme is able to save a significant amount of energy on the reception and the mode transitions. Since the PG scheme broadcasts in the downlink phase, all wireless clients have to be in the active mode, thus consuming much more energy. Moreover, each client in the PG scheme performs more mode transitions. As for the energy consumed on the transmission and the sleeping modes, the MG and PG schemes consume almost the same amount of energy.

If we change the distribution of the traffic arrival rates while keeping the total traffic rate unchanged, it is interesting to observe that the energy consumptions of the 4 components will not change. For example, the experiments with $\boldsymbol{\lambda} = [30; 60; 90]$ and $\boldsymbol{\lambda} = [10; 50; 110]$ have the same results as that in Figure 3.10(a). In other words, the total arrival rate determines the energy consumption under the optimal schemes, when other parameters and conditions are unchanged.

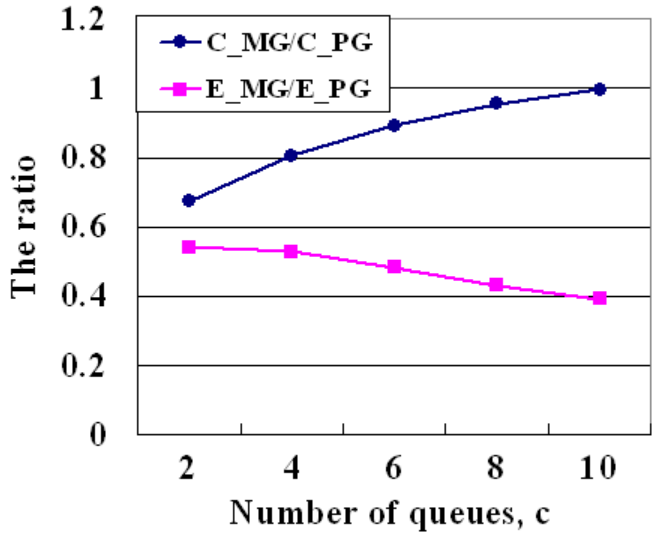
Since $z = (\sum_{i=1}^{c} \lambda_i \mathbf{E}[C^*])/\mathbf{E}[E^*]$, given $\boldsymbol{\lambda}$, z_{MG}/z_{PG} implies $\frac{\mathbf{E}[C^*_{MG}]/\mathbf{E}[C^*_{PG}]}{\mathbf{E}[E^*_{MG}]/\mathbf{E}[E^*_{PG}]} > 1$. All of our experiments support that $\mathbf{E}[C^*_{MG}]/\mathbf{E}[C^*_{PG}] > \mathbf{E}[E^*_{MG}]/\mathbf{E}[E^*_{PG}]$, for example the results given in Figure 3.10(b). Besides, the numerator $\mathbf{E}[C^*_{MG}]/\mathbf{E}[C^*_{PG}]$ increases faster than the denominator $\mathbf{E}[E^*_{MG}]/\mathbf{E}[E^*_{PG}]$ when c increases. This can explain why z_{MG}/z_{PG} increases with c, which was discussed earlier.

3.5.3 Delay Analysis

In this section, we study how delay performance is impacted by the optimal transmission power allocation. The delay is defined to be the duration from the packet arrival at one queue to its departure from the queue. We do not compare



(a) Comparisons based on the 4 components for $\boldsymbol{\lambda}$ =[60;60;60]and $k = 0.5$.



(b) Comparison of the energy consumptions per cycle and the cycle time for $\boldsymbol{\lambda}$=$20 * e_c$ and $k = 0.5$.

Figure 3.10. Comparing energy efficiency for the optimal MG and PG schemes.

our AP-centric TPC scheme with other energy efficiency schemes in the term of delay, since they are designed under different constraints. We use experiments to compare the packet delay in the optimal MG scheme with the one in the optimal PG scheme. Moreover, we compare the packet delay of each scheme with and

without the optimal transmission power allocation.

Average Delay Under Optimal Power Allocations

These experiments yield average packet delay under different optimal power allocations. Firstly, we caculate the average delay of uplink and downlink packets in each queue individually. For example in queue i, the average delay of uplink packets is the sum of the average service time $b_{u,i}$ and the average waiting time [48] $\frac{(1+\rho_{u,i})(\sigma^2_{C_i}/E[C]+E[C])}{2}$, where C_i is the cycle time seen by queue i. When omitting the variance of C_i, the the average delay of uplink packets for queue i is $T_{U,i} = (1+\rho_{u,i})E[C]/2 + b_{u,i}$. Therefore, we obtain the average delay of uplink packets denoted by $\mathbf{T_U} = \{T_{U,1}, \ldots, T_{U,c}\}$ in the MG and PG schemes. We also obtain the average delay of downlink packets denoted $\mathbf{T_D} = \{T_{D,1}, \ldots, T_{D,c}\}$, where $T_{D,i} = (1+\lambda_{d,i}b_{d,i})E[C]/2 + b_{d,i}$ in the MG scheme and $T_{D,i} = (1+\lambda_D\overline{b_D})E[C]/2 + b_{d,i}$ in the PG scheme.

Table 3.7 shows the total energy consumed per polling cycle ($\overline{E}$), the average delay of the downlink packets and uplink packets ($\mathbf{T_D}$ and $\mathbf{T_U}$) as well as the delay variances (σ_{TD} and σ_{TU}) in the asymmetric four-client systems which use optimal power allocations. Here, $\sigma_{TD} = \sum_{i=1}^{c}(T_{D,i} - \sum_{i=1}^{c} T_{D,i}/c)^2/c$ and $\sigma_{TU} = \sum_{i=1}^{c}(T_{U,i} - \sum_{i=1}^{c} T_{U,i}/c)^2/c$. Three traffic workloads of $\boldsymbol{\lambda} = [10; 15; 20; 25] \times x$, $x = 1, 3, 5$, $\boldsymbol{\beta} = 0.5e_4$ are considered, and the packet size is 512 bytes. Clearly, the optimal MG scheme consumes less average energy per cycle than the corresponding optimal PG scheme. For all optimal PG schemes, the average delay of uplink packets are longer than that for the downlink packets. Additionally, their downlink delay increases with x. The small values of delay variances show that the four clients with different arrival rates have similar delay. In other words, the optimal power allocations do not adversely affect the fairness of delay in most

cases. But σ_{TU} increases a lot when $x = 5$ in both optimal schemes, which means that the unfairness of uplink packet on delay among wireless clients increases with the traffic workload is large.

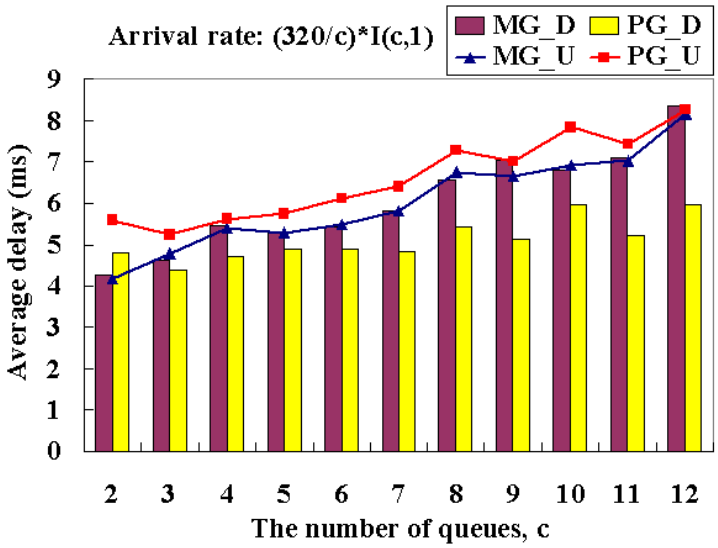
In order to simplify the delay analysis, the following simulations are performed in a symmetric network with $\boldsymbol{\lambda} = ke_c$ and $\boldsymbol{\beta} = 0.5e_c$, $c = 2, \ldots, 12$. The intensity weighted mean of delay [102] is considered for two kinds of packets in the network, which is called as delay in brief. One is the delay of downlink packets, $\overline{T_D} = \sum_{i=1}^{c} \frac{\lambda_{d,i} b_{d,i}}{\rho_d} T_{D,i}$. The other is the delay of uplink packets, $\overline{T_U} = \sum_{i=1}^{c} \frac{\lambda_{u,i} b_{u,i}}{\rho_u} T_{U,i}$. The results of delay under the optimal power allocations are shown in Figure 3.11. $\overline{T_D}$ is close to $\overline{T_U}$ in all optimal MG schemes, while $\overline{T_D}$ is less than $\overline{T_U}$ in all optimal PG schemes. In most cases, the optimal PG schemes have lower delay for the downlink packets than the optimal MG schemes, while the optimal MG schemes have the lower delay for the uplink packets than the optimal PG schemes. Therefore, the optimal PG scheme may provide the better delay performance when β increases, and the optimal MG scheme may provide the better delay performance when β decreases.

Energy Consumption Verses Delay

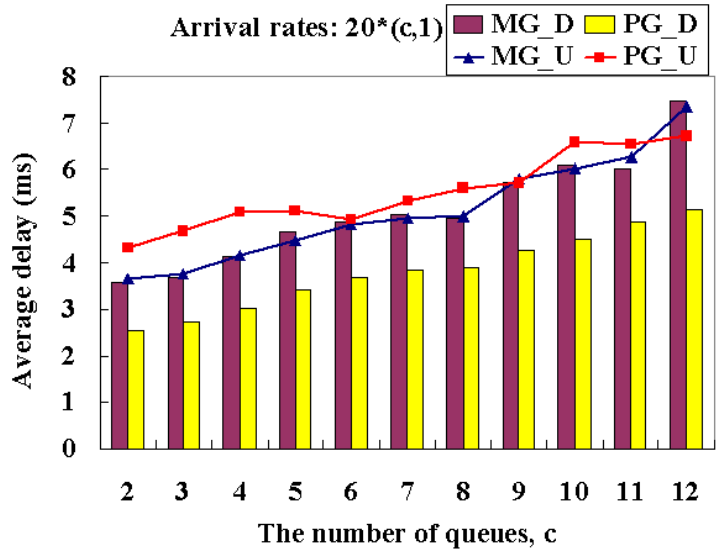
Next, we investigate the relationship between the energy consumption and the delay performance. The symmetric system with $\boldsymbol{\lambda} = 80e_c$ and $\boldsymbol{\beta} = 0.5e_c$ is considered on two grouping schemes. We may adopt a transmission power allocation instead of the optimal one, which randomly selects the packet transmission times from the feasible region $[\mathbf{b}_{min}, \mathbf{b}_{max}]$. For the random power allocation, we define two quantities in reference to the energy consumption $\overline{E^*}$ and the downlink (uplink) delay $\overline{T_D^*}$ ($\overline{T_U^*}$) obtained under the optimal power allocation. Define the *energy inflation ratio* (r_e) of a random power allocation by $\frac{\overline{E} - \overline{E^*}}{\overline{E^*}}$, where $\overline{E}$ is

Table 3.7. The average delay for $\boldsymbol{\lambda} = [10; 15; 20; 25] \times x$ and $\boldsymbol{\beta} = 0.5e_4$.

Scheduling	x	$\overline{E}(mJ)$	$\mathbf{T_U}(ms)$	σ_{TU}	$\mathbf{T_D}(ms)$	σ_{TD}
MG	1	1.3564	[4.0246;3.8502;4.1127;4.1467]	0.0132	[3.6882;4.1600;4.0260;4.1960]	0.0402
MG	3	3.5190	[4.7928;4.4650;4.9875;4.6139]	0.0383	[4.5765;5.0436;4.4353;4.7767]	0.0523
MG	5	9.2480	[5.0329;4.3240;4.8556;5.2596]	0.1192	[4.3194;4.7842;4.7272;4.9750]	0.0571
PG	1	2.6617	[5.3914;5.1583;4.9916;4.9916]	0.0265	[2.8538;2.8861;2.7985;3.3033]	0.0402
PG	3	6.6091	[5.3249;5.0583;5.6859;5.0535]	0.0668	[3.8474;4.1120;3.9285;4.1517]	0.0163
PG	5	15.4546	[5.0577;5.8118;6.4648;5.6484]	0.2511	[4.8643;5.2513;5.0549;4.7520]	0.0361



(a) $\sum_{i=1}^{c} \lambda_i = 320$ for $k = 0.5$.



(b) $\lambda_i = 20$ for $k = 0.5$.

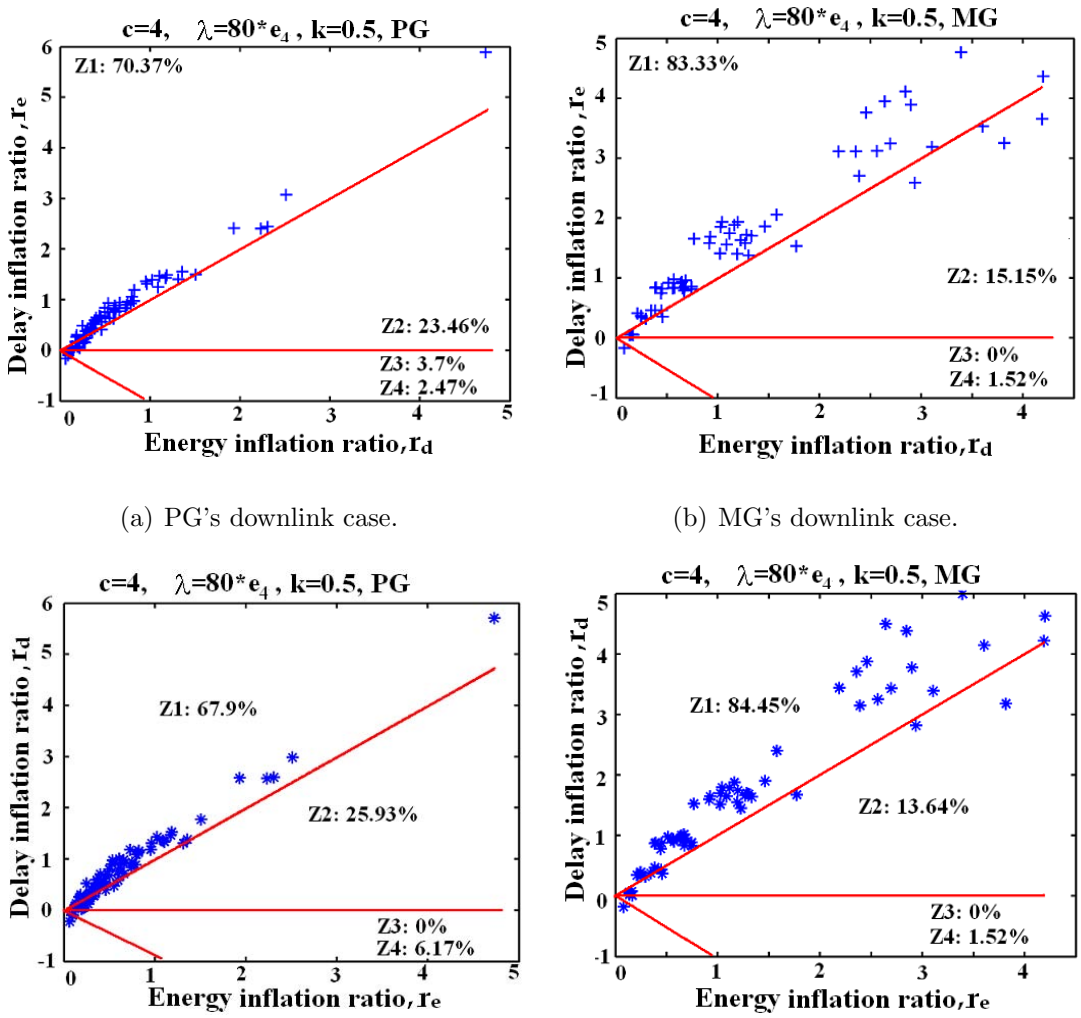
Figure 3.11. Delay comparison under optimal schemes.

the energy consumption obtained under the random power allocation. Similarly, we define the *delay inflation ratio* (r_d) of a random power allocation by $\frac{\overline{T_D}-\overline{T_D^*}}{\overline{T_D^*}}$ ($\frac{\overline{T_U}-\overline{T_U^*}}{\overline{T_U^*}}$) for the downlink (uplink) packets.

The results of 60 random allocations depicted in Figure 3.12 show that the optimal MG and PG schemes definitely achieve the minimal energy consumption and generally do not degrade delay. All of the energy inflation ratios are positive (i.e., $r_e \geq 0$), and most of the delay inflation ratios are positive. However, neither the optimal MG scheme nor the optimal PG scheme can achieve the best delay performance because of some negative delay inflation ratios.

Next, we divide the random allocations into 4 regions in Figure 3.12: Z_1 ($r_d > r_e > 0$), Z_2 ($r_e > r_d > 0$), Z_3 ($-r_e < r_d < 0$) and Z_4 ($r_d < -r_e < 0$). The random allocations in Z_1 and Z_2 have the longer delay than the optimal allocation since $r_d > 0$. Therefore, the optimal allocation outperforms the random allocations in these 2 regions on saving energy and reducing delay. It means that the more random schemes locate in Z1 and Z2, the better performance the optimal scheme has. On the other hand, the random allocations in Z_3 and Z_4 have better delay performance than the optimal allocation since $r_d < 0$. However, the balance between delay and energy consumption may still be satisfied in Z_3, because the energy inflation ratio is always higher than the delay inflation ratio in this region (i.e., $r_e > |r_d|$). It is easy to find that most of the random allocations fall into Z_1 and Z_2 (at least 94%). Only a small percentage (at most 6%) of the random allocations belong to Z_4. As a result, we can conclude that both schemes can achieve optimal power allocation without the expense of a higher average delay most of the time.

Finally, we compare the optimal scheme and random schemes in the symmetric system of $\boldsymbol{\lambda} = 20e_c$ and $\boldsymbol{\beta} = 0.5e_c$, when the number of queue c is varied. The



(a) PG's downlink case.

(b) MG's downlink case.

(c) PG's uplink case.

(d) MG's uplink case.

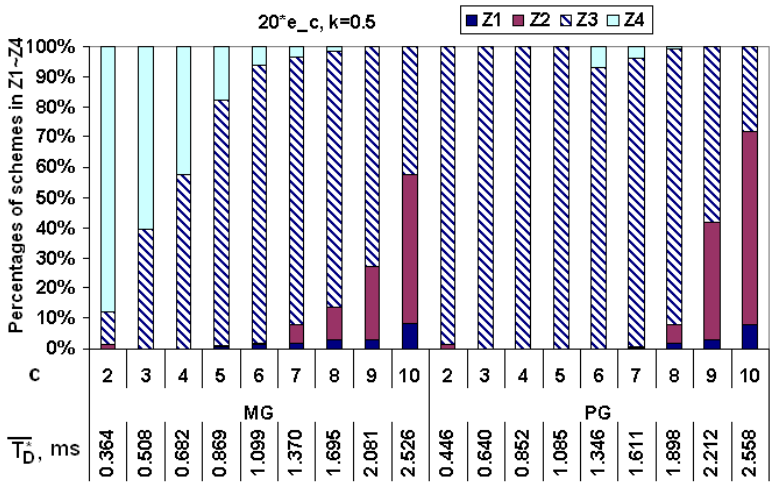
Figure 3.12. Energy inflation ratio and delay inflation ratio for a random power allocation for $c = 4$.

average delay $\overline{T_U^*}$ and $\overline{T_D^*}$ (marked below the x axis) increases with c in both optimal PG and MG schemes. In Figure 3.13(a), the total percentage of Z_1, Z_2 and Z_3 increases with c. Therefore both optimal schemes have more positive influence on the delay performance when the scale of network increases. However, when $c \leq 4$, more than 40% MG schemes locate in Z_4, which indicates the optimal MG schemes save a little energy by degrading much downlink delay of the light traffic load. In these cases, the optimal PG schemes are better in achieving the tradeoff between the energy efficiency and the downlink delay, since almost all of PG schemes locate in $Z3$. Moreover, Figure 3.13(b) shows that both optimal schemes has a similar impact on the average uplink delay.

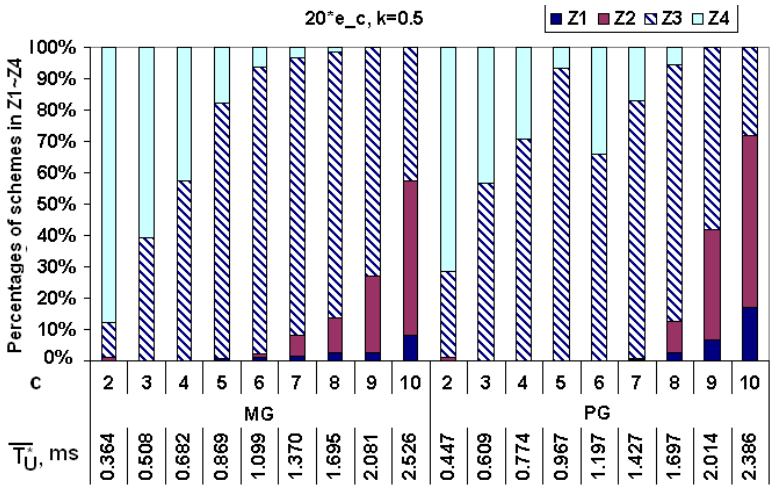
3.6 Effect of an Uncooperative Client

We have so far performed extensive experiments when all users are assumed to adopt the optimal power allocations. In this section we relax this assumption and study the effect of an uncooperative user on itself and others. The numerical results show that wireless client has not an incentive to change the optimal transmission power, since this uncooperative action does not bring any benefit.

As an example, we discuss the MG scheme for a symmetric system with $\boldsymbol{\lambda} = 60e_4$ and $\boldsymbol{\beta} = 0.5e_4$. The iteration algorithm gives the optimal power allocation $\mathbf{P}^* = P^*e_4$. In the following, all parameters or results for the optimal schemes are indicated by superscript $*$. Moreover, we consider a tagged queue q_T as the single uncooperative client, and his transmission power is gP^*, where g is a change parameter. When $g \neq 1$, q_T changes its optimal transmission power. We label the results for q_T with subscript t, the results of other clients using P^* with subscript o and the results of the whole system with subscript s.



(a) The average delay of AP.



(b) The average uplink delay.

Figure 3.13. Delay comparison for different c, $\boldsymbol{\lambda} = 20 \times e_c$ and $k = 0.5$.

We first evaluate how the uncooperative user affects energy consumption. Denote the energy consumption per cycle by ξ and that with the optimal power allocation $\mathbf{P}^*$ as ξ^*. We define the *energy reduction ratio* (ERR) under a random scheme as $(\xi^* - \xi)/\xi^*$. The case of $g = 1$ corresponds to the optimal power allocation; therefore, the ERR for $g = 1$ is 0. Since ξ^* is the minimum value for the whole system, the system ERR is always negative when $g \neq 1$. Figure 3.14 shows the ERR of q_T when q_T uses a higher (i.e., $g > 1$) or lower (i.e., $g < 1$) power. When q_T uses a higher power, it will consume more energy while other queues may receive a little benefit on energy saving. As the figure shows, the total energy consumption still increases. On the other hand, if q_T uses a lower power, its energy consumption per cycle reduces. But in most cases the total energy consumption is still higher than ξ^*, because other queues consumes more energy.

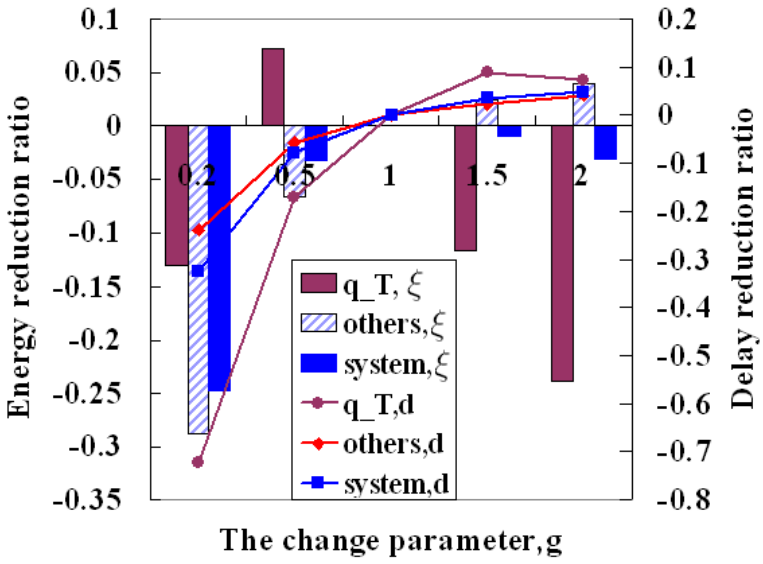


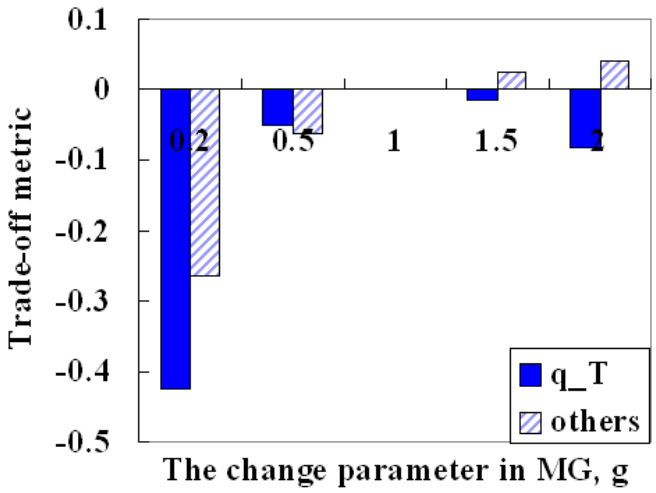
Figure 3.14. Energy reduction ratio and delay reduction ratio in the presence of an uncooperative user.

We now evaluate how the uncooperative user affects the average delay of all

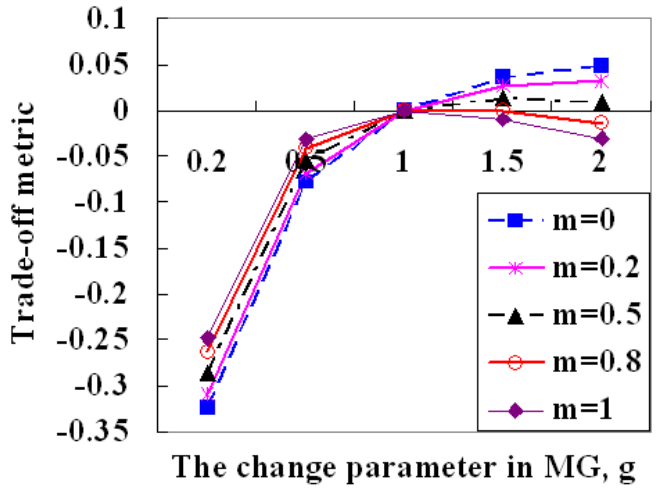
of downlink and uplink packets, which is denoted as d. The *delay reduction ratio* (DRR) is similarly defined as $(d^* - d)/d^*$ and d^* is the average packet delay for an optimal MG scheme. As shown in Figure 3.14, when q_T uses a higher power ($g > 1$), the average packet delay in each queue decreases; therefore, the average delay for the whole system also decreases. On the other hand, when q_T uses a lower power ($g < 1$), the delay performance will degrade in each queue as well as the entire system. Moreover, it is possible to improve the delay performance at the cost of higher energy consumption. That is, there is a tradeoff between the energy conservation and the delay reduction, to be discussed next.

Lastly we evaluate the effect of the uncooperative user on both delay and energy consumption. To this end, we introduce a *delay-energy tradeoff metric* $\varphi = (1 - m)\frac{d^* - d}{d^*} + m\frac{\xi^* - \xi}{\xi^*}$, where m is a tradeoff parameter. The higher φ is, the better delay-energy tradeoff is. Note that the special cases of $m = 1$ and $m = 0$ correspond to the ERR and DRR, respectively. Using $m = 0.5$, we compute the trade-off metrics which are shown in Figure 3.15(a). The figure shows that q_T cannot improve its delay-energy tradeoff as long as it changes its optimal transmission power, i.e., $\varphi_t < 0$ when $g \neq 1$. However, when q_T uses higher transmission power (i.e., $g > 1$), other queues will benefit a little since $\varphi_o > 0$. For the whole system, our optimal solution achieves the highest φ_s when $m > 0.5$, as shown in Figure 3.15(b). Even when the delay-energy tradeoff metric is dominated by the delay performance, i.e., $m \leq 0.5$, the nonoptimal solutions achieve only a little higher φ_s than the optimal solution. These analytical results can be obtained and analyzed similarly for $c \neq 4$ and asymmetric systems.

In conclusion, the optimal transmission power allocation minimizes energy consumption and lets all clients performs well together. The uncooperative client q_T cannot get any benefit by changing its optimal transmission power. The



(a) φ_t and φ_o for $m = 0.5$.



(b) Impact of m on φ_s.

Figure 3.15. Delay-energy tradeoff metric in the presence of an uncooperative user.

total energy consumption of all clients increases when $g \neq 1$, since q_T does not adopt the optimal transmission power. When q_T reduces its transmission power, nobody gets benefit in terms of energy saving, delay performance and delay-energy tradeoff. When q_T increases its transmission power, other clients consumes less energy, shortens packet delay and improves delay-energy tradeoff, while q_T

consumes more energy and degrades delay-energy tradeoff.

Chapter 4

Improving Energy Efficiency Using Centralized PSM

The *power saving mode* (PSM) specified in the IEEE 802.11 standard is unable to achieve energy efficiency in many practical scenarios. It is suboptimal when an infrastructure wireless network serves two or more wireless clients which have different traffic patterns or QoS requirements. In this chapter, we propose a new AP-centric PSM scheme called Centralized PSM (C-PSM) to improve energy efficiency of *all* PSM-enabled clients under heterogeneous traffic scenarios. Our overall approach is to alleviate two inefficient factors of the standard PSM scheme. Then C-PSM optimally chooses the PSM parameters to reduce channel contentions and avoid the clients to wake up too frequently.

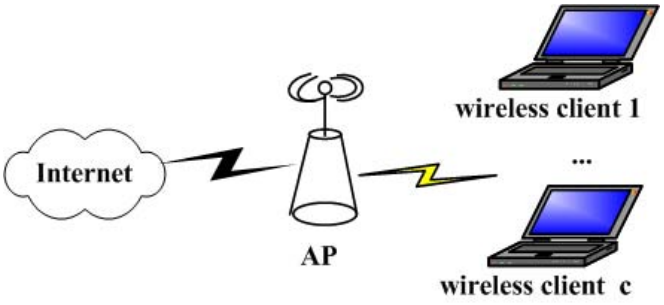
We first describe the system model and develop a MATLAB simulator which implements the IEEE 802.11 *Distributed Coordination Function* (DCF) with PSM. Based on the simulation results, we can identify two root causes responsible for the standard PSM's ineffectiveness: frequent unnecessary wake-ups and

intense channel contentions (the worst cases incur channel collisions). After that, we introduce C-PSM which is designed to reduce unnecessary wake-ups and channel contentions. In C-PSM, the AP selects optimal *beacon interval* (BI) and the *listen interval* (LI) based on the traffic patterns of all PSM-enabled clients. Here, the LI parameter of each client is an integer (i.e., $\gamma = 1, 2, \dots$) which determines client's wake-up frequency by the granularity of BI. The AP could also reduce the channel collisions by adjusting the client's minimal *contention window* (CW). Moreover, a first-wake-up schedule could be adopted to further alleviate the channel contention by minimizing the number of awaken clients during one BI.

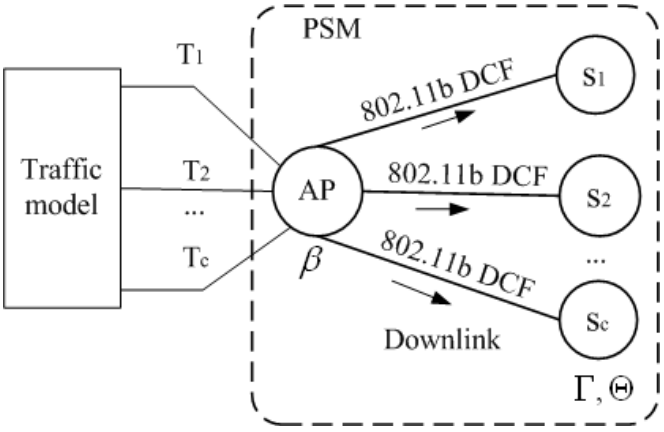
Moreover, we have defined four performance indexes to compare different PSM-based schemes. The evaluation results show that C-PSM reduces more power consumption (up to 83.73%), obtains higher energy efficiency (up to 320%) and induces shorter AP buffering delay (up to 90%) than the standard PSM scheme (with default parameters).

4.1 System Model

We study and improve the performance of PSM in an infrastructure wireless network shown in Figure 4.1(a). The network consists of an AP and c PSM-enabled wireless clients s_j, $j = 1, \dots, c$, which run on IEEE 802.11b with a transmission rate of 11Mps. We have chosen IEEE 802.11b over IEEE 802.11a/g for its less complicated protocols, since the rate adaptation algorithms in IEEE 802.11a/g's will make the analysis much more difficult. We purposely ignore the rate adaptation component in order to focus on the energy-saving sources of PSM in the MAC layer. We only consider the downlink traffic (i.e., those from the AP to the clients), because PSM is effective only for the downlink-dominant



(a) The network topology.



(b) The system model.

Figure 4.1. The network topology and system model.

communication patterns. Furthermore, we model the traffic arriving at the AP by traffic sources each of which generates data frames to a corresponding client. Therefore, the traffic sources can model various traffic characteristics which are affected by applications, upper-layer protocols and network path properties.

Figure 4.1(b) shows the system model. There are two random variables in our model: the inter-frame arrival time and the frame size. Let T_j be the inter-frame arrival time for s_j's traffic source and the mean of T_j be δ_j. Let the vector of the client's mean inter-frame arrival times be $\Delta = [\delta_1, \ldots, \delta_c]$. Our current study allows T_j to take on four types of distributions: deterministic (DET), uniform (UNI), exponential (EXP) and Pareto (PAR). On the other hand, the frame size

distribution is either deterministic or uniform. Besides the BI (β), the AP in C-PSM can configure the LIs for all clients ($\Gamma = [\gamma_1, \ldots, \gamma_c]$) and the minimal CWs for all clients ($\Theta = [\theta_1, \ldots, \theta_c]$).

4.2 Simulator Design

Due to the complexity of analyzing wireless networks, simulation has been a major tool to study their performance [107]. Popular simulators include ns-2 [141], J-SIM [171], OPNET [133], OMNeT++ [159], CSIM [79], GloMosim [109] (its commercial version is QualNet [160]) and MATLAB [169]. In general, simulator could capture finer details than analytical models. Unfortunately, many publicly available simulators have not implemented most operations (or components) of infrastructure networks and PSM, such as the beacon frames and PS-Poll frames. Even for the de facto simulator ns-2 [1] in the academic community, the PSM module [105] only supports a single client. For other simulators, such as J-SIM, the 802.11 PSM scheme has not been implemented or released to the public.

We have attempted to implement the full version of PSM in ns-2 and J-SIM, but met many obstacles which cannot be resolved without a major overhaul to the code structure. This prompted us to design and implement a new simulator based on MATLAB which provides an easy-to-use language and other supporting facilities to model the MAC sublayer accurately and effectively. Since we focus on the energy consumption due to the MAC sublayer, we abstract the PHY layer as an error-free channel and the propagation delay as zero. The effect of the PHY layer could be studied in future work by integrating the existing MATLAB codes for the IEEE 802.11b PHY, e.g., [33]. There are also other MATLAB-

based IEEE802.11 simulators. For example, a simple simulator of mobile wireless networks [34] is designed for IEEE 802.11a by MATLAB. The simulator in [136] implements IEEE 802.11 DCF with or without RTS/CTS in ad-hoc networks using MATLAB.

4.2.1 A MATLAB-based Simulator

We have developed a MATLAB-based simulator for the IEEE 802.11b DCF with PSM. This simulator implements their details including the transmissions of all frames, the usage of backoff algorithm, as well as the adjustment CWs. At the same time, we have excluded other nonessential elements (e.g., authentication and (de)association) and the RTS/CTS mechanism which is often turned off to increase communication performance.

The simulator divides the time into fixed slots. The time slots are used in the channel contention and backoff algorithm in DCF. Their length, denoted by t_s, is medium-dependent, e.g., 20μs in IEEE 802.11b and 9μs in IEEE 802.11a/g. The time unit of our simulator is then designed to be a fraction of t_s, for example $t_u = 0.5t_s$ in this thesis. The operations and state transitions which are associated with the AP, clients and channel are updated by time units. Table B.1 in Appendix B lists all state variables in this simulator.

Figure 4.2 shows a high-level structure of the simulator. The simulator inputs include the configurable parameters (e.g., β, Γ and Θ), traffic matrix $\mathbf{A}$ and t_{sim}. The matrix $\mathbf{A}$ contains the arrival times of data frame for the clients which are pre-generated according to the specified arrival distribution. The matrix $\mathbf{A}$ has c rows and m_p columns, and $\mathbf{A}(i, j)$ gives the arrival time at the AP of the jth data frame addressed for s_i. Usually, m_p is the number of data frames sent by the

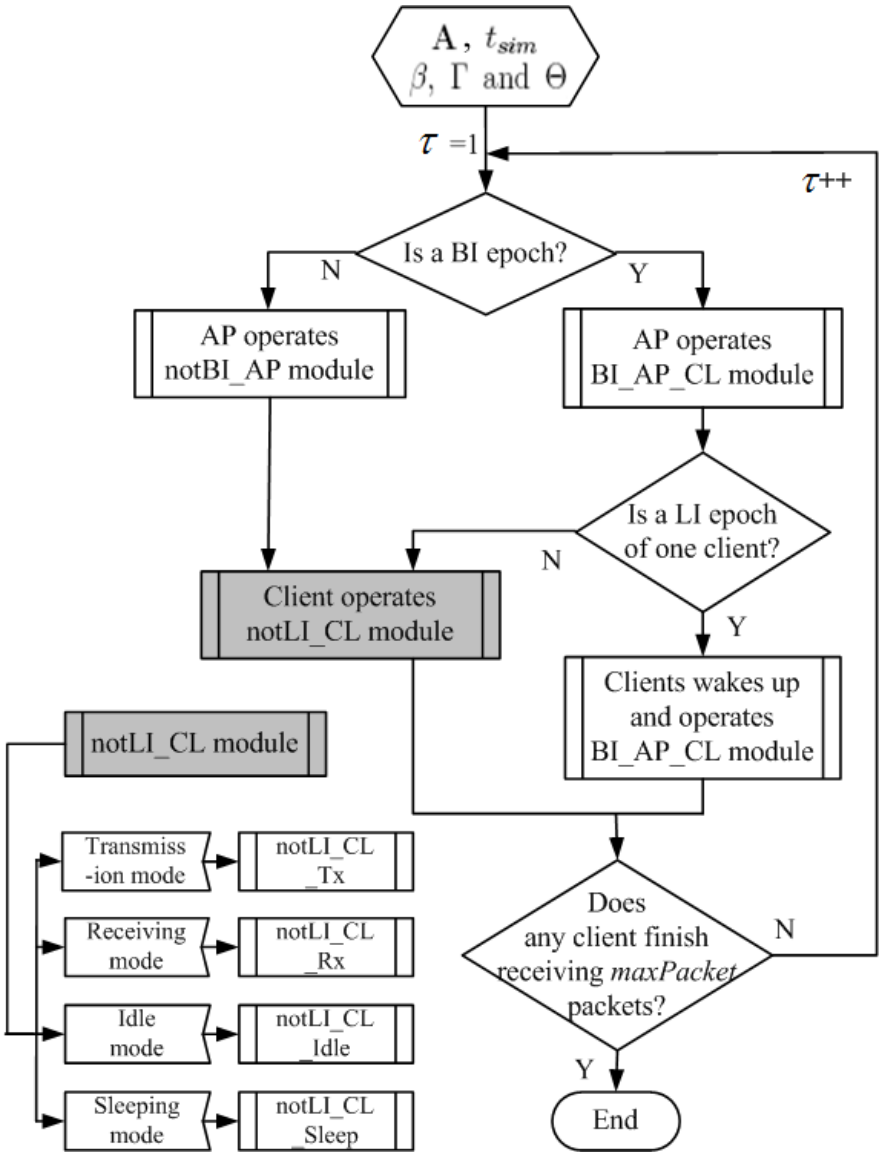


Figure 4.2. A high-level program structure of the PSM simulator.

highest-rate traffic flow within t_{sim}. Here t_{sim} is the input parameter which helps determine the simulation period. Instead of using t_{sim} directly, a simulation session will terminate when any client finishes receiving m_p frames, and m_p is determined by

$$m_p = \max_{i=1,\ldots,c}\{V_i\},\ V_i = \max_j\{A_{i,j} \leq t_{sim}\}. \tag{4.1}$$

At the same time, we have designed three main modules (`BI_AP_CL`, `notBI_AP`

and `notLI_CL`) to keep track of the states and operations of the AP, clients and channel. We label the time unit of our simulator by an index τ, starting from 1. One time slot consists of two time units where $t_u = 10\mu s$ and $t_s = 20\mu s$. Based on the AP's timing of sending the beacon frames, a time unit is regarded as a *BI epoch* when τ is a multiple of $\frac{\beta}{t_u}$. Based on the client's timing of waking up, a time unit is regarded as an *LI epoch* of s_j when τ is a multiple of $\frac{\beta * \gamma_j}{t_u}$. Since the LI of each client is a multiple of BI, an LI epoch must be a BI epoch.

We have designed the `BI_AP_CL` module for the AP's activities in BI epochs and for clients' activities in their LI epochs. Since the AP and clients interact with each other directly in those epochs, implementing their activities in the same module is a logical choice. On the other hand, their activities in other units are independent; therefore, we have implemented separate modules for them: Module `notBI_AP` for the AP when it is not in BI epochs and module `notLI_CL` for the clients when they are not in LI epochs. We will summarize the services performed by these three main modules in Subsection 4.2.2. For details, see Appendix B.

Simulation parameters	Values
Number of clients	1 to 20
Data transmission rate (DTR)	11 Mbps
Basic transmission rate (BTR)	2 Mbps
Data frame size (DFS)	512 bytes
Beacon frame size (BFS)	28 bytes
PS-Poll frame size (PFS)	14 bytes
ACK frame size (AFS)	14 bytes
Transmission power	1.4 W
Reception power	0.9 W
Idle power	0.7 W
Sleeping power	0.060 W
Wake-up energy consumption, e_w	0.003 J
Slot time, t_s	20μs
SIFS	10μs
DIFS	50μs

Table 4.1. The parameters used in the simulation experiments.

The simulation parameters in Table 4.1 are used for the experiments conducted in this thesis [1]. We will consider a single client and as many as 20 clients. The transmission rate of data frames is 11 Mbps while the basic transmission rate of control frames is fixed to 2Mbps [2]. This set of power and energy consumption parameters for LUCENT IEEE 802.11 WAVELAN PC card has been widely used [58, 119] and evaluated in [58]. And other parameters have been widely utilized in many wireless simulators and applications. Moreover, the detail discussion of power consumption models is given in Section 4.6.

Finally, the simulator produces detailed trace files which record frame exchanges, channel collisions and clients' mode transitions. By carefully analyzing the trace files, we have validated the correctness of the PSM simulator. After observing the time of convergence from several preliminary experiments, we have let each experiment run for at least 20 seconds in simulation time. We have also repeated each simulation setting for 20 times and reported their average values. All the results reported in this thesis fall within a 95% confidence level. Therefore, we have obtained the following metrics from the trace files:

1. P: the total power consumed by the clients in watt (W).
2. T: the total throughput of clients in bits per second (bps).
3. $R_{T/P}$: $\frac{T}{P}$, the total energy efficiency metric in bits per joule (bpJ).
4. $R_{c/t}$: $\frac{N_c}{N_t}$, where N_t is the total number of transmission attempts by all clients, and N_c is the total number of channel collisions experienced by all clients.

[1] The support of the long preamble is mandatory, and is the default setting on most devices [65]. We therefore use a long preamble, $192\mu s$. Then the transmission time of one frame equals to $192 + \frac{size*8}{transmission\ rate}\mu$s.

[2] The basic transmission rate is 1Mbps or 2Mbps [65] which depends on the coding scheme. We fix it as 2Mbps.

5. $R_{u/w}$: $\frac{N_u}{N_w}$, where N_w is the total number of wake-ups by all clients, and N_u is the total number of unnecessary wake-ups by all clients.

6. $R_{bB/B,k}$: $\frac{N_{bB,k}}{N_B}$, where N_B is the total number of BIs, and $N_{bB,k}$ is the total number of BIs in which k clients are involved in channel contention, $k \geq 2$.

7. d_j: the average buffering delay of s_j's frames at the AP.

4.2.2 Three main modules in the simulator

1. Module `BI_AP_CL` manages the operations of the AP in each BI epoch and the operations of each client s_j ($\forall j$) in its LI epoch. When the channel is not busy, the module allows the AP to send a beacon frame. The beacon frame includes Traffic Indication Map, denoted by **TIM** $= (TIM_1, \ldots, TIM_j)$. If the AP has buffered frames for s_j, $TIM_j = 1$; otherwise, $TIM_j = 0$. The module also lets the clients which are in their LI epoch enter the receiving mode for the beacon reception. If no client wakes up to receive the beacon frame, this unheard beacon will be marked. Moreover, the module updates the related important states. For example, τ could be updated by increasing $beacon\_units - 1$ where $beacon\_units = \frac{BFS}{t_u * BTR}$ is the transmission time of a beacon frame by units. When the channel is busy, the module will dispose the beacon frame that experiences collision.

2. Module `notBI_AP` manages the operations of the AP when the current unit is not a BI epoch. When the AP is sending a beacon or data frame, the module need not do anything. When the AP is idle, the module determines whether it is ready to send a data frame according to the 802.11 DCF. The AP randomly selects the receiving client among all awakened clients after receiving their PS-Poll frames. If the channel is available, the AP will transmit the data frame.

The module performs this transmission and updates the relevant states, such as recording the channel state to be busy, updating the power mode of the selected client and resetting the CW of the AP as the default value (i.e., 31). Moreover, τ is updated by increasing $data\_units - 1$ where $data\_units = \frac{DFS}{t_u * DTR}$ is the transmission time of a data frame by units. If the AP suffers from a collision due to the busy channel, the module will give up the data transmission (assume that this data frame should be retransmitted later) and update the AP's backoff timer after doubling its current CW.

3. Module `notLI_CL` manages the operations of the clients in their non-LI epochs. It consists of four submodules according to the different power modes of clients: `notLI_CL_Tx`, `notLI_CL_Rx`, `notLI_CL_Idle` and `notLI_CL_Sleep`. The mode transitions of s_j $(\forall j)$ in a non-LI epoch are implemented by these four submodules, illustrated by the continuous lines in Figure 4.3.

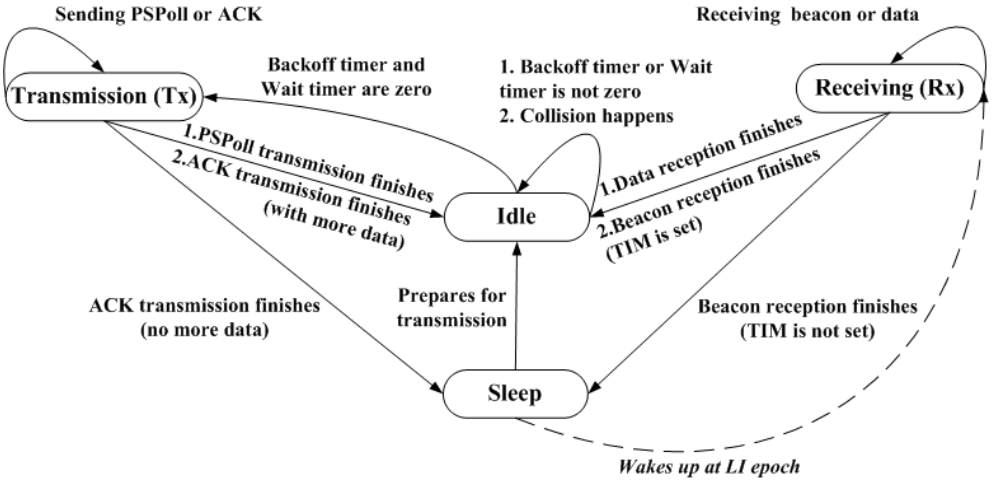


Figure 4.3. The state diagram of clients.

1. When s_j is in the transmission mode (Tx), submodule `notLI_CL_Tx` only takes actions and updates the important states when a transmission is finished. When s_j finishes sending a PS-Poll frame, it will enter the idle mode

to prepare for receiving a data frame. When s_j finishes sending an ACK, s_j will create a PS-Poll frame and enter the idle mode if it has more buffered frames in the AP. If s_j does not have any more buffered frame, it will go to sleep.

2. When s_j is in the receiving mode (Rx), submodule `notLI_CL_Rx` only takes actions and updates the important states when a reception is finished. When s_j finishes receiving a data frame, it will create an ACK and enter the idle mode. When s_j finishes receiving a beacon frame, it will create a PS-Poll frame if $TIM_j = 1$ and enter the idle mode. When s_j does not have any buffered data frame in the AP (i.e., $TIM_j = 0$), it will go to sleep.

3. When s_j is in the idle mode, submodule `notLI_CL_Idle` determines whether s_j is ready to send a frame according to the 802.11 DCF. If the backoff timer or the waiting timer [3] is not zero, the submodule will decrease one of the timers by 1 while s_j stays in the idle mode. If both timers are zero, the submodule will perform the transmission of a PS-Poll or ACK frame in s_j. If the transmission begins when the channel is available, s_j will enter the transmission mode and the related important states will be updated. For example, the CW of s_j will be reset as θ_j and τ will be updated by increasing $pspoll\_units - 1$ ($ack\_units - 1$) where $pspoll\_units = \frac{PFS}{t_u * BTR}$ ($ack\_units = \frac{AFS}{t_u * BTR}$) is the transmission time of a PS-Poll (ACK) frame by units. If the channel collision occurs, the submodule will dispose this transmission and update s_j's backoff timer after doubling its CW. Moreover, when an ACK is collided, the submodule will retransmit its corresponding

[3] The waiting timer counts the interframe spaces such as SIFS, DIFS and EIFS. If the previous frame is received or the backoff timer returns to zero, the client must wait for one interframe space in the idle channel before sending one frame.

data frame.

4. When s_j is in the sleep mode, submodule `notLI_CL_Sleep` takes actions only when the mode transition of client happens. When s_j wants to send a frame, the submodule lets s_j go to the idle mode for the transmission preparation and updates the related important states. In particular, if all clients are sleeping at a non-BI epoch and do not have any transmission attempt, τ will be updated by increasing $[\frac{\beta}{t-u} - mod(\tau, \frac{\beta}{t_u})] - 1$, which is helpful to speed up the simulation.

4.3 Analysis of the IEEE 802.11 PSM

We first discuss the system stability in our model. Since the standard PSM is applicable to save energy when the system workload is not heavy, we study the PSM performance under light traffic. In particular, we show the great effects of BI and LIs on the energy consumption and performance metrics. That is why we can improve energy efficiency by adjusting BI and LIs.

4.3.1 A Stability Analysis

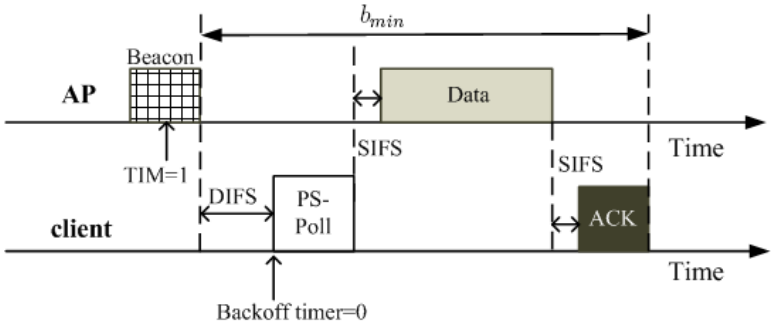


Figure 4.4. Shortest service time of receiving one data frame.

We first consider the condition of system stability in an extreme case when the PSM-enabled client is too busy to go to sleep and the channel contention is totally avoided. In this case, the shortest service time for retrieving one data frame include the transmission time of three fames and the additional overhead of IEEE 802.11 DCF, shown in Figure 4.4. After receiving one beacon frame, a PSM-enabled client should wait for a short period (we set this short period to DIFS) before sending a PS-Poll frame when the AP has buffered its data frames. The client then transmits a PS-Poll frame after waiting for a random backoff timer (i.e., it changes from 0 to the current CW. Here we adopt its minimal value). Moreover, the AP immediately responds to the PS-Poll frame by transmitting the data frame at the end of a SIFS [65]. And the client finishes receiving this frame and sends an ACK frame after another SIFS has elapsed.

According to Table 4.1, the shortest service time for one data frame is given by

$$b_{min} = \frac{DFS}{DTR} + \frac{PFS + AFS}{BTR} + DIFS + 2SIFS \approx 1.13ms.$$

According to the stability condition, the total arrival rate of frames in a wireless network, $\lambda = \sum_{j=1}^{c} \frac{1}{\delta_j}$, should satisfy $\lambda b_{min} < 1$. Moreover, we define the utilization criterion $\rho = \lambda b_{min}$ to measure the system workload.

However, the real service time b for receiving one data frame in our model must be larger than b_{min}, because the PSM introduces sleeping periods (similar to vacation), one-PS-Poll-one-packet scheme introduces idle periods (due to channel contention), and even channel collision occurs the longer backoff overhead. In general, the difference between b and b_{min} depends on the PSM setting and channel condition. Similarly, we have the condition of system stability as $\frac{1}{\lambda} > b$.

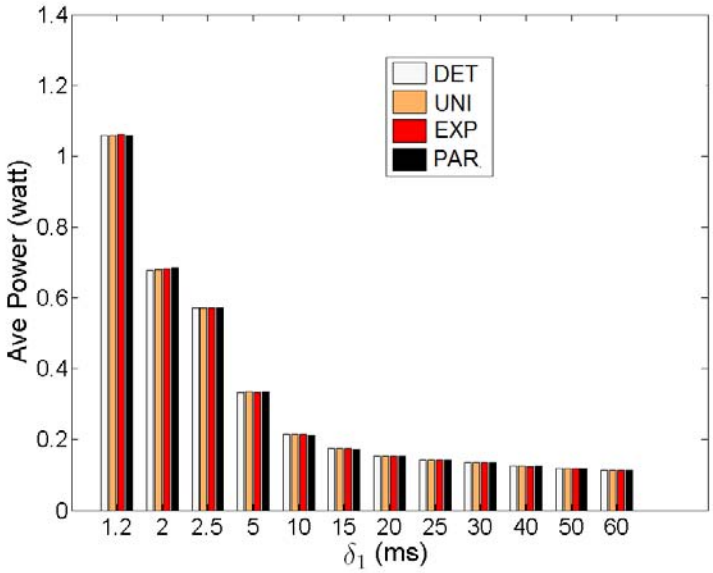
Since $b_{min} \leq b$, we get the necessary condition of system stability:

$$1/\sum_{j=1}^{c} \frac{1}{\delta_j} > b_{min}. \tag{4.2}$$

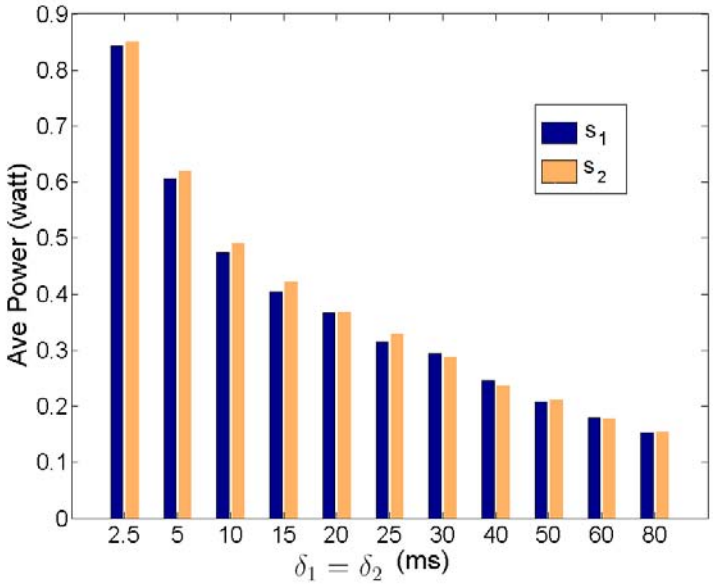
For example, when $c = 1$, $\delta_1 > 1.13$ms is the necessary stability condition of a single client system. When $c = 2$, the necessary stability condition is $\delta_1 = \delta_2 > 2.26$ms for two symmetric clients.

Next, we study the energy saving due to the standard PSM scheme (i.e., $\beta = 100$ms and $\gamma_j = 1, \forall j$) under different system workload. In order to guarantee the system stability, we set the minimal inter-frame arrival time to 1.2ms when $c = 1$ and 2.5ms when $c = 2$. Figure 4.5(a) shows that the average power of s_1 reduces with δ_1 in the one-client system under different traffic distributions. When $\delta_1 \leq 2$ms, the system workload is heavy, since $\rho = \frac{b_{min}}{\delta_1} \geq 0.565$ and the client's average power is larger than 0.7W, the idle power consumption. That is, the standard PSM scheme could not save much energy when the traffic workload is heavy. The client mostly spends high power in the active modes to deal with heavy traffic instead of going into the sleep mode. When $\delta_1 \geq 2.5$ms, the client's average power is always less than 0.6W and reduced with δ_1. The standard PSM scheme is more effective to save more energy when the system workload reduces.

Figure 4.5(b) shows the average power of two symmetric PSM-enabled clients under the EXP distribution of inter-frame arrival times. When $\delta_1(\delta_2)$ increases, the average power of s_1 (s_2) reduces. When $\delta_1 = \delta_2 \leq 5$ms, the average power of each client is larger than or close to 0.7W, the idle power consumption. In these cases, the client seldom sleeps and the system workload is heavy since $\rho = \frac{2b_{min}}{\delta_1} \geq 0.452$. When $\delta_1 = \delta_2 \geq 10$ms, the system workload reduces with δ_1 (δ_2) and the standard PSM scheme is effective to save energy, because the power consumption of each client is close to or less than 0.5W. Similar simulation results



(a) Power of the single client for $c = 1$ and $\gamma_1 = 1$.



(b) Power of two symmetric clients under the EXP traffic distribution for $c = 2$ and $\gamma_1 = \gamma_2 = 1$.

Figure 4.5. Power verses inter-frame arrival time for $\beta = 100$ms.

are obtained under other traffic distributions.

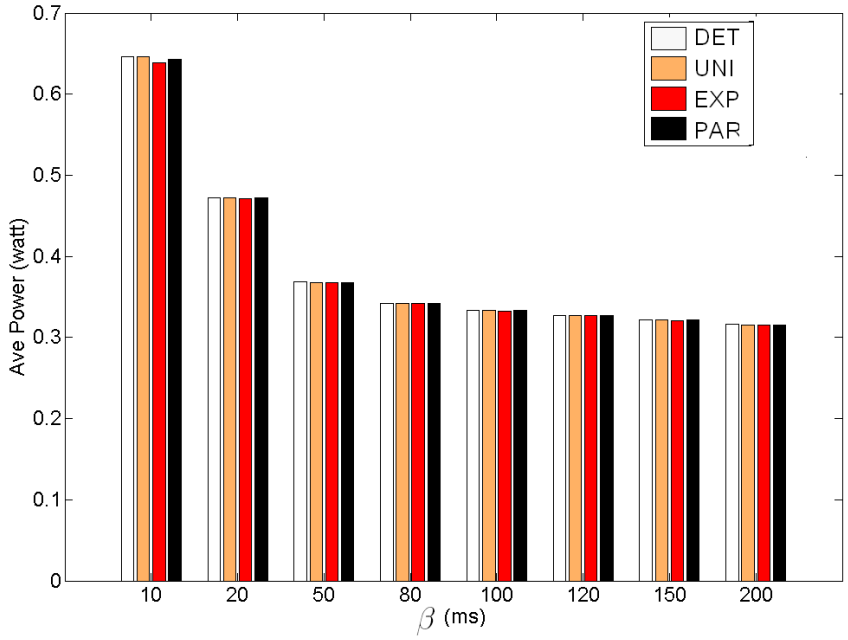
The above simulation results show that the standard PSM scheme is effective to save energy when the workload of this system is light. Nedevschi et al. [129] have also shown that sleeping is valuable to save energy depending (primarily) on the utilization of the network (i.e., the practical sleeping algorithms are valuable when the network utilization is less than 30%) and the power profile of wireless devices. Therefore, we consider the performance of PSM under light traffic only when the Δs that satisfy $\rho \leq 30\%$. Moreover, the effects of the power profile of wireless devices on the PSM performance are discussed in Section 4.6.

4.3.2 Effects of the Beacon Interval

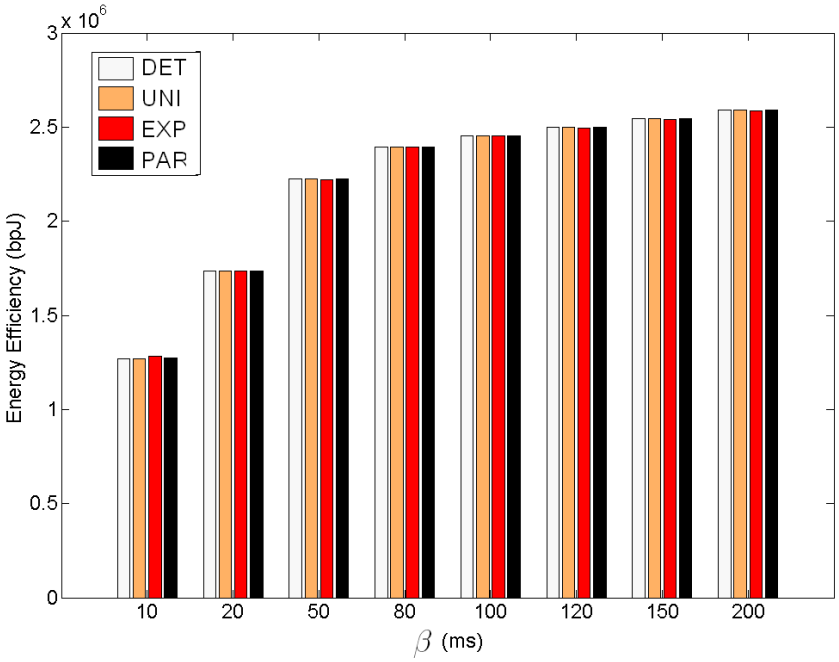
Previous studies have shown that the default value of BI (i.e., $\beta = 100$ms) is not optimal in many cases. Shown in Figure 4.6, the average power of the single client reduces with β when the traffic is light with $\delta_1 = 5$ms and $\rho = 22.6\% < 30\%$. The energy efficiency of this client increases with β, which is defined as the ratio of throughput to average power. Therefore, we could increase β to improve energy efficiency only when the single client is stable (i.e., the number of buffered frames is limited). But such an one-client system is far away from a general infrastructure network which includes two or more PSM-enabled client. It omits the important issues of clients' interaction such as channel contentions.

We use a two-client system as an example to study the effects of β in the range of $10 \sim 200$ms, and the other parameters are set by default, i.e., $\Gamma = [1; 1]$ and $\Theta = [31; 31]$. Each client here wakes up in every BI epoch to compete for channel resource and the AP's service by sending a PS-Poll frame.

The results in Figures 4.7 show the effects of β for two symmetric clients

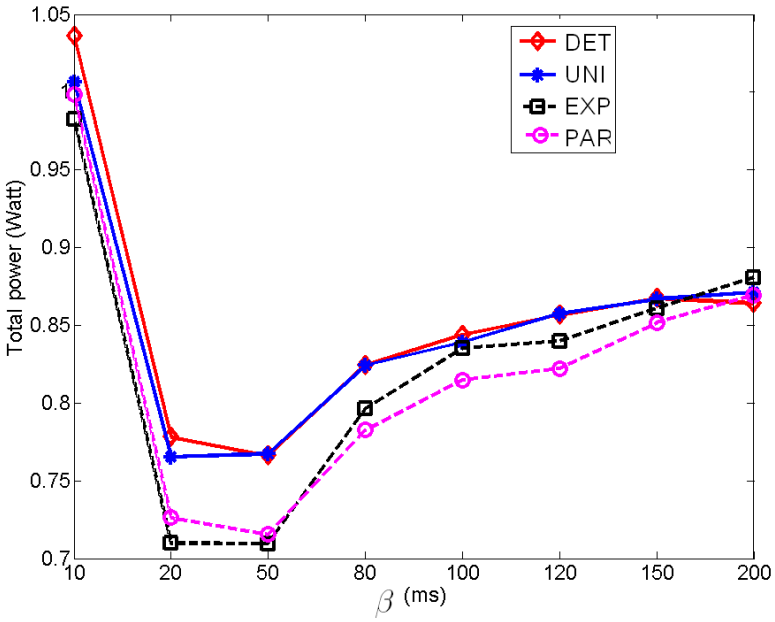


(a) Average power.

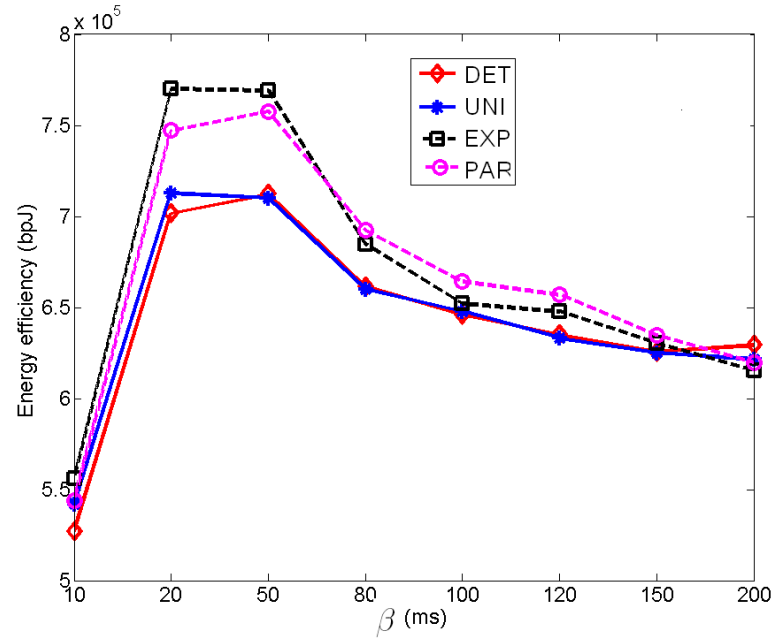


(b) Energy efficiency metric.

Figure 4.6. The effects of BI for $c = 1$, $\delta_1 = 5$ms and $\gamma_1 = 1$.



(a) P verses β for $\gamma_1 = \gamma_2 = 1$.



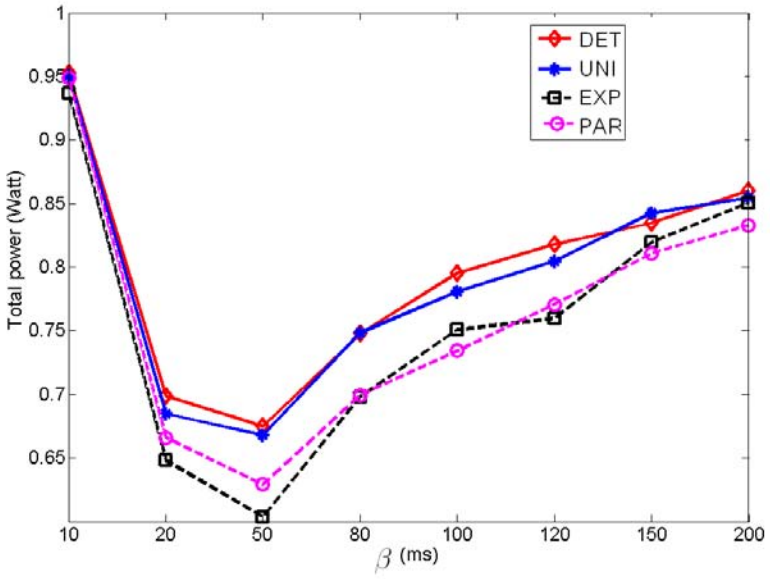
(b) $R_{T/P}$ verses β for $\gamma_1 = \gamma_2 = 1$.

Figure 4.7. The effects of BI for $c = 2$ and $\delta_1 = \delta_2 = 15$ms.

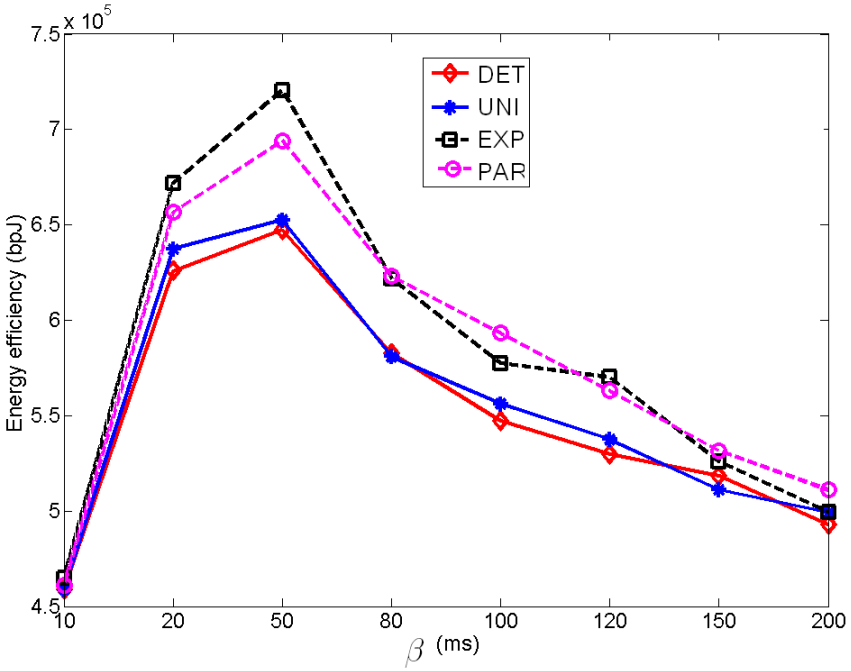
when $\delta_1 = \delta_2 = 15$ms and $\rho \approx 15\% < 30\%$. As shown in Figure 4.7(a), P is high when β is too small, because much energy is wasted on clients' frequent wake-ups. When β is too large, many frames have been accumulated at the AP. Therefore, energy is wasted on channel contentions, since two clients simultaneously wake up to compete for retrieving these frames. The optimal β in this example is 50ms, instead of the default value of 100 ms. Moreover, since we only consider downlink traffic, T equals to AP's throughput and is almost equal to a constant when the system is stable. $R_{T/P}$ is inversely proportional to P. Then Figure 4.7(b) behaves similarly as Figure 4.7(a).

In most cases, the PSM-enabled clients are not symmetric. We should study the effects of BI in the asymmetric network. When $\delta_1 = 15$ms and $\delta_2 = 25$ms ($\rho \approx 12\% < 30\%$), the curves of P in Figure 4.8(a) are approximately concave and the curves of $R_{T/P}$ in Figure 4.8(b) are approximately convex. Similar to the above symmetric system, the best BI under all traffic distributions is 50ms, since it achieves the lowest P and the highest $R_{T/P}$. In brief, s_1 with a smaller inter-frame arrival time ($\delta_1 < \delta_2$) is called as a fast client, while s_2 is called a slow client. Next, we use the case of the EXP traffic as an example to study the effects of β on each client.

The clients' average power and energy efficiency are sensitive to β and Δ. When $\beta = 10$ms, the average power of each client is high in Figure 4.9(a) because of a lot of unnecessary wake-ups. As illustrated in Figure 4.10(b), the ratio of unnecessary wake-ups to total wake-ups in the fast client s_1 (denoted as $R_{u/w,1}$) is 54% while the $R_{u/w,2}$ in the slow client s_2 is higher to be 68%. When $\beta > 20$ms, the average power of the fast (slow) client roughly increases (decreases) with β. In Figure 4.9(b), the fast client achieves the highest energy efficiency and lowest average power when $\beta = 50$ms and $\gamma_1 = 1$; whereas the slow client achieves its

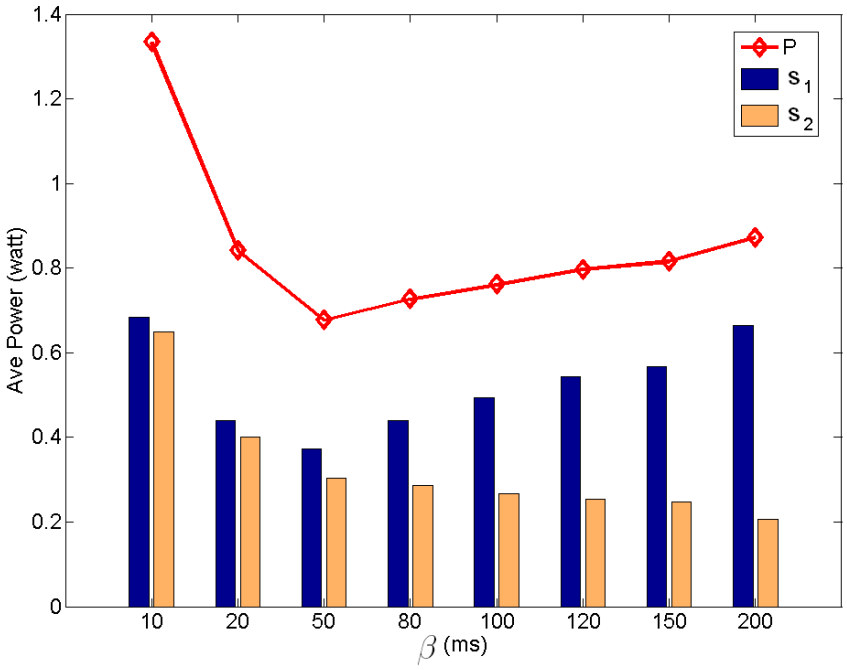


(a) P verses β for $\gamma_1 = \gamma_2 = 1$.

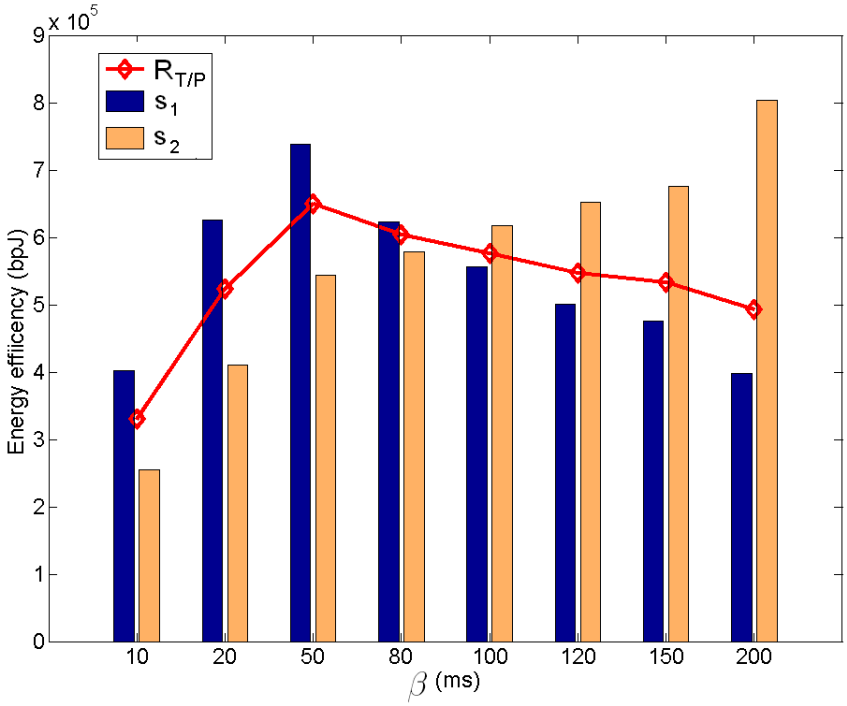


(b) $R_{T/P}$ verses β for $\gamma_1 = \gamma_2 = 1$.

Figure 4.8. The effects of BI for $c = 2$ and $\Delta = [15; 25]$ms.



(a) Average power verses β for $\gamma_1 = \gamma_2 = 1$.



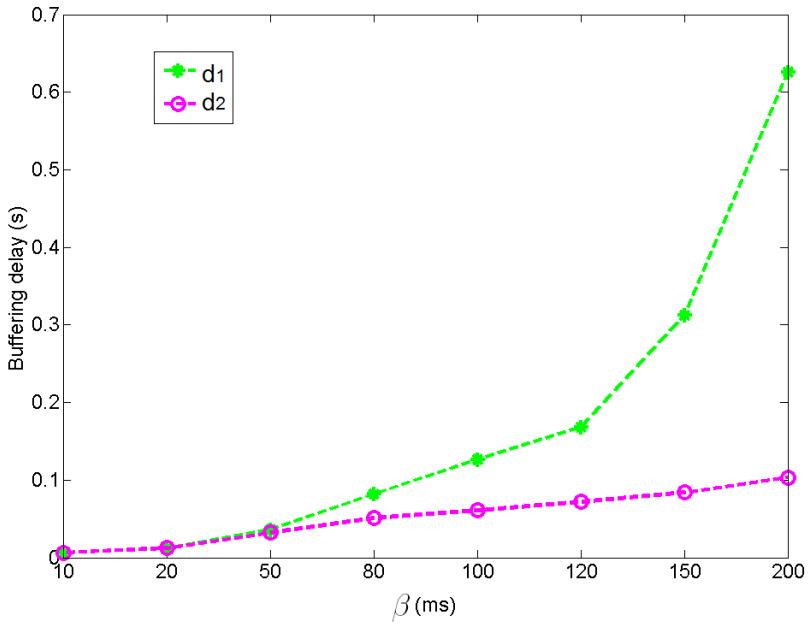
(b) Energy efficiency metric verses β for $\gamma_1 = \gamma_2 = 1$.

Figure 4.9. The effects of BI on power and energy efficiency under the EXP traffic distribution for $c = 2$ and $\Delta = [15; 25]$ms.

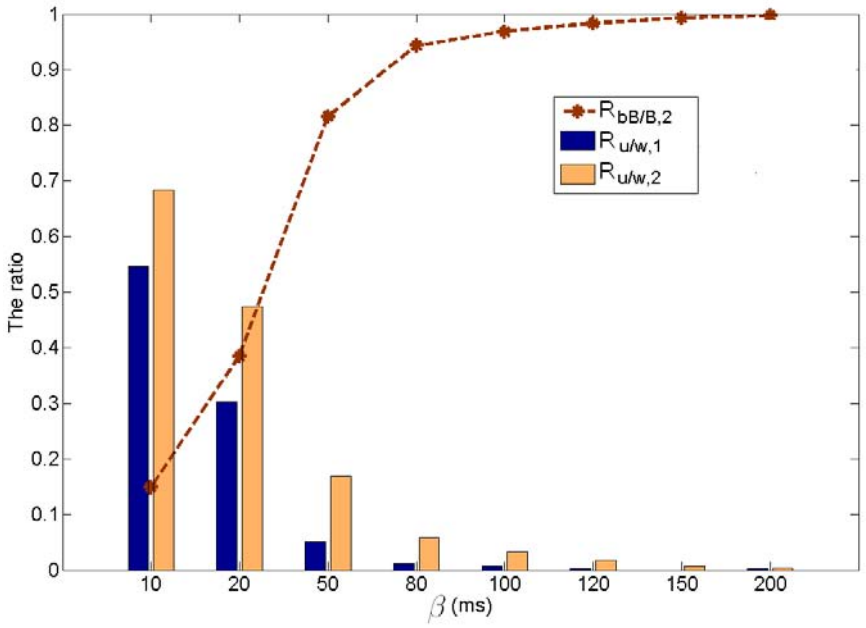
best metrics when $\beta = 200$ms and $\gamma_2 = 1$. That is, the fast (slow) client should wake up at a short (long) LI period (i.e., $\beta \times \gamma_j$) to improve its own energy efficiency. There must exist a trade-off of BI between two clients to achieve the lowest P and the highest $R_{T/P}$. The red line presents that the most energy-efficient case for both clients in these experiments is the case of $\beta = 50$ms and $\gamma_1 = \gamma_2 = 1$.

Moreover, Figure 4.10(a) shows that the mean AP buffering delay of each client roughly increases with β. The fast client's d_1 increases faster than the slow client's d_2. When β increases, the number of buffered frames for s_j increases. Two clients will spend more energy and time to deal with these frames. They also have a higher probability to be involved in the channel contention. As illustrated in Figure 4.10(b), the ratio of the BIs when two or more awaken clients compete for channel to the total BIs, $R_{bB/B,2}$, is larger than 0.9 when $\beta \geq 80$ms. Therefore, the smaller BI is helpful to save energy by reducing the AP buffering delay especially on the fast client. However, the value of BI should not be too small. Otherwise, the frequent wake-ups result in a high $R_{u/w,j}$ in s_j and consume high power consumption.

In summary, we may adjust β to improve energy efficiency of all PSM-enabled clients when the unnecessary wake-ups and the channel contentions are reduced. However, when $\gamma_j \equiv 1$, the adjustment of β is not enough to improve the energy efficiency of all asymmetric clients. For example, the best case in Figure 4.10(b) still suffers from a high ratio of simultaneous wake-ups, i.e., $R_{bB/B,2} = 82\%$ when $\beta = 50$ms. Next, we study the effects of Γ on the asymmetric clients.



(a) AP buffering delay verses BI for $\gamma_1 = \gamma_2 = 1$.



(b) Ratios of unnecessary wake-ups and simultaneous wake-ups verses BI for $\gamma_1 = \gamma_2 = 1$.

Figure 4.10. The effects of BI on performance metrics under the EXP traffic distribution for $c = 2$ and $\Delta = [15; 25]$ms.

4.3.3 Effects of the Listen Intervals

Most papers assume that each PSM-enabled client wakes up in every BI epoch (i.e., $\gamma_j \equiv 1$, $\forall j$). The assumption is widely accepted when the studies of PSM consider a single PSM-enabled client or symmetric PSM-enabled clients. In these studies, the optimization of β is used to improve communication performance and save energy. However, this assumption is not suitable when the network has two or more asymmetric PSM-enabled clients. The clients could adopt different LIs to wake up every a multiple of BI when $\gamma_j \geq 1$, $\forall j$.

The choice of $\Gamma = [\gamma_1, \ldots, \gamma_c]$ can influence the clients' energy consumption and communication performance. Lei et al. [113] obtained optimal Γ to maximize energy efficiency based on the M/G/1 and D/G/1 queueing models. However, some assumptions used in these models are not suitable for the standard PSM scheme. In our simulation study, we have fixed β and considered different Γs for the two clients.

Firstly, we fix $\beta = 50$ms (which is the best β in Subsection 4.3.2) to discuss the simulation results in the two-client system when $\Delta = [15; 25]$ms. Table 4.2 for the EXP distribution clearly shows that the default case of $[1; 1]$ is not optimal for the six performance metrics. The best value for each metric is underlined. Overall, the case of $[1; 2]$ achieves the best performance than the rest. It has reduced the overhead of collisions with the lowest $R_{c/t}$, avoided frequent unnecessary wake-ups with the low $R_{u/w}$ (which is less than 0.05) and greatly reduced the energy consumed on channel contentions with the lowest $R_{bB/B,2}$. Moreover, the case of $[1; 2]$ achieves the shortest d_1 in the fast client at the cost of degrading the d_2 in the slow client. Its total performance metrics are best, since the fast client has a greater influence to the whole system than the slow one. It worths noting that

$\frac{\gamma_1}{\gamma_2}$ is closest to $\frac{\delta_1}{\delta_2}$ for $\Gamma = [1;2]$. We have observed similar results for the other three inter-frame arrival time distributions.

Table 4.2. Simulation results of different Γs under the EXP traffic distribution for $c = 2$, $\Delta = [15;25]$ms and $\beta = 50$ms.

Γ	$R_{c/t}$	$R_{u/w}$	$R_{bB/B,2}$	P(W)	$R_{T/P}(10^5$bpJ)	d_1(ms)	d_2(ms)
$[1;1]$	1.54%	11.51%	81.37%	0.6109	7.1578	37.4	32.3
$[1;2]$	1.04%	4.97%	42.32%	0.5487	7.9674	29.8	60.0
$[2;1]$	1.07%	12.24%	46.79%	0.6032	7.2316	81.0	28.0
$[2;2]$	1.25%	1.67%	49.16%	0.7470	5.8260	125.4	61.3

On the other hand, the case of $[2;2]$ results in the lowest $R_{T/P}$ and the highest P. The $R_{bB/B,2}$ is close to $\frac{1}{\gamma_j} = 0.5$, which means that two clients always compete for receiving buffered frames when they wake up. Since d_j $(\forall j)$ is longest, two clients have to spend the longest time on receiving data or waiting for an available channel. The buffered data frames for the fast client could not be retrieved within one LI period since $d_1 = 125.4ms > \gamma_1\beta = 100$ms. Therefore, both clients cannot sleep often, because many data frames have been buffered in the AP during the long LI period. As a result, they consume much energy in $\Gamma = [2;2]$.

Table 4.3. Simulation results of different Γs under the EXP traffic distribution for $c = 2$, $\Delta = [15;25]$ms and $\beta = 100$ms.

Γ	$R_{c/t}$	$R_{u/w}$	$R_{bB/B,2}$	P(W)	$R_{T/P}(10^5$bpJ)	d_1(ms)	d_2(ms)
$[1;1]$	1.36%	2.12%	97.18%	0.7420	5.8768	134.4	60.5
$[1;2]$	0.80%	0.38%	48.84%	0.6135	7.1238	61.2	118.8
$[2;1]$	0.81%	2.60%	49.89%	0.7719	5.6435	338.5	50.7
$[2;2]$	0.85%	0.17%	50.00%	0.8443	5.0763	624.4	102.6

When β increases to the default value 100ms, the case of [1;2] is still the best scheme among four different Γs. After comparing Table 4.2 and Table 4.3, we find that the metrics of the different Γs have similar trends when β is fixed. For the same Γ, the clients consume more power in the increase of β and then

degrade energy efficiency. For example, in the case of $[1;1]$, $P = 0.5487$W when $\beta = 50$ms, while $P = 0.6135$W when $\beta = 100$ms. The reason is that the $R_{bB/B,2}$, d_1 and d_2 increase with β. And then, the clients have higher probabilities to be involved in the channel contention which costs more energy and time.

We therefore conclude that the PSM performance could be improved by using optimal β and Γ. The joint optimization of β and Γ should reduce unnecessary wake-ups to save energy consumed on client's mode transitions and alleviate channel contention to reduce energy consumed on client's idle mode. A faster client could adopt a smaller γ and wakes up more frequently than a slower client. Moreover, we may give a higher priority to the slower client to improve the access fairness among asymmetric clients. We can also obtain optimal CWs to mitigate the buffering delay resulted from the PSM scheme.

4.4 Designing Centralized PSM

This section presents Centralized PSM (C-PSM). The most important feature is that the AP manages the PSM operations by determining optimal parameters of the AP and all PSM-enabled clients. Therefore, this scheme is regarded to be AP-centric.

Using the main algorithm, the AP firstly searches for an optimal BI (β^*) and for an optimal vector of LIs (Γ^*) to reduce the mentioned PSM inefficient factors. Then it searches an optimal vector of CWs (Θ^*) to balance client's access probability which is to mitigate the unfairness among clients. Next, AP optionally obtains an optimal first-wake-up schedule (FWS). FWS is an optimal sequence of the first wake-up times for clients, denoted as $\mathbf{r}^*$. It can minimize the maximal number of waking clients in a BI epoch. We will show that FWS can

further decrease the probability of simultaneous wakeups if two or more elements of Γ^* are the same or have the same common factor.

We firstly describe the procedure of C-PSM scheme as following. The AP sends a beacon frame every β^* and the PSM-enabled client s_j wakes up every $\gamma_j^* \times \beta^*$ to listen to the beacon frame, $j = 1, \ldots, c$. When the TIM of the beacon shows that the AP buffers frames for s_j, this client will contest to retrieve a data frame by sending a PS-Poll frame. The AP randomly selects one awaken client to serve after receiving one or more PS-Poll frames. When the AP sends one data frame to the selected client, other awaken clients stay in idle mode and wait for the next chance to retrieve their data frames. When the selected client finishes receiving a data frame, it will send out an ACK frame. If the more data bit is not set in the received data frame, this client will go to sleep till its next LI epoch. Otherwise, it will contest the channel with other awaken clients by sending another PS-Poll. Moreover, the 802.11 DCF is adopted by the AP and all wireless clients. The minimal CW value of the AP is 31 by default, and the one of s_j is θ_j^*, $\forall j$. A backoff algorithm is used to double the current CW when a collision occurs to PS-Poll, ACK or data frames. If FWS is not adopted (i.e., $r_j^* \equiv 0$), all PSM-enables clients first wake up to receive a beacon at the beginning of the simulation; otherwise, the PSM-enabled client s_j first wakes up at the time of $r_j^* \beta^*$, $\forall j$.

Note that the AP and wireless clients using C-PSM follow the standard PSM scheme with the optimal parameters. C-PSM is therefore compatible with the IEEE 802.11 standard. Its overhead is that the AP should determine the optimal parameters for all wireless devices. For example, the AP notifies γ_j^* and Θ_j^* (and r_j^* if necessary) to each client. Next, we describe how to obtain these optimal parameters in C-PSM by designing the main algorithm and FWS.

4.4.1 The Main Algorithm

The AP uses a main algorithm to determine and deploy optimal PSM settings on itself and all clients. The AP first searches for the optimal β (denoted by β^*) and optimal Γ (denoted by Γ^*) based on the client's frame arrival patterns. These optimal settings are expected to bring significant improvement to energy efficiency, because the intervals are selected to reduce the number of unnecessary wake-ups and channel contentions. The AP also obtains the optimal Θ (denoted by Θ^*) to ensure that any client will not be denied channel access for too long.

After analyzing the simulation results obtained in section 4.3, we draw the following principles for β^*, Γ^* and Θ^*.

1. β^* should not be too small. A frequent transmission of beacon frames increases the workload of AP and aggravates the channel contention. Anderson et al. [127] suggested that the value of β is at least 10ms in a practical environment. We recommend the minimal β^* to be 10ms in this thesis (i.e., $\beta^* \geq$ 10ms).

2. In order to reduce energy consumption due to unnecessary wake-ups, we reduce the unnecessary wake-ups of clients. Then the optimal LI period of s_j (i.e., $\gamma_j^* \times \beta^*$, $\forall j$) should be large enough during which at least one data frame arrives at the AP with a high probability. In general, the client s_j should adopt a larger γ_j^* when its δ_j is larger, $\forall j$.

3. In order to reduce energy consumption due to channel contention, we should reduce the number of awaken clients during one BI. The reduction of clients' simultaneous wake-ups decreases the occurrences of channel contention to save clients' energy consumed on idle mode.

4. We should not set the optimal LI period of s_j (i.e., $\gamma_j^* \times \beta^*$, $\forall j$) to be too large. Otherwise, AP will accumulate many data frames, and s_j will spend much energy in active to retrieve them. The worst case happens when s_j does not retrieve all buffered frames within one BI. And then s_j and other newly waking clients will be involved in the channel contention and spend a great deal of energy in idle mode.

5. In order to balance the performance among clients, we may assign a smaller θ_j^* to s_j which uses a larger γ_j^*. That is, the PSM-enabled client, which seldom wakes up, has a higher probability of winning the channel and retrieving its data frames from the AP.

Next, we design the main algorithm according to these principles. The inputs to the main algorithm include Δ, β_{min}, ϵ_β and ϵ_Θ. β_{min} is a lower bound of β^*, and ϵ_β and ϵ_Θ are the step sizes for searching β^* and Θ^*, respectively. The algorithm executes the following three consecutive steps to yield β^*, Γ^* and Θ^*.

Step 1 (Determining the β^* and Γ^* candidates) is to obtain a number of β^* and Γ^* candidates for the second step.

We first estimate the optimal LI periods of clients and denoted them as $[L_1; \ldots; L_c]$ (ms). We further let $L_j = \alpha_j \times \delta_j$, where $\alpha_j \geq 1$ is an integer scaling factor. To reduce unnecessary wake-up, the probability that an awaken client finds an empty buffer at the AP (denoted by Pr_0) should be less than a given threshold $0 < \xi \leq 1$. The choice of this threshold reflects the tradeoff between the number of unnecessary wake-ups and the period of channel contention. If ξ is too large, the LI will be short, and a lot of unnecessary wake-ups will occur. If ξ is too small, the frames buffered during the long LI may cause channel contention. After running a number of empirical simulations, we have

selected $\xi = 0.05$. Then, α_j is the smallest integer which satisfies $Pr_0 \leq 0.05$. Table 4.4 shows examples of determining α_j under the four inter-frame arrival time distributions. For example, under the EXP traffic, L_j is three times of δ_j, since $Pr_0 \leq 0.05$ when $\alpha \geq 3$.

To determine the β^* candidates, following the guideline in [127], we set β_{min} to 10ms. To determine the upper bound of β, we note that BI should not be larger than any client's LI. Therefore, the upper bound of β is given by $\min_{\forall j} L_j$. We then select $n+1$ β^* candidates uniformly within the range of $[\beta_{min}, \min_{\forall j} L_j]$, where $n = \lfloor (\min_{\forall j} L_j - \beta_{min})/\epsilon_\beta \rfloor$. Moreover, for each β^* candidate β_i, we consider three Γ^* candidates:

1. $\Gamma_{i,1} = [\lceil L_1/\beta_i \rceil; \cdots; \lceil L_c/\beta_i \rceil]$,
2. $\Gamma_{i,2} = [\langle L_1/\beta_i \rangle; \cdots; \langle L_c/\beta_i \rangle]$, where $\langle x \rangle$ gives the round-off value of a real number x, and
3. $\Gamma_{i,3} = [\lfloor L_1/\beta_i \rfloor; \cdots; \lfloor L_c/\beta_i \rfloor]$.

Table 4.4. Empty probability verses scaling factor under four traffic distributions.

Distribution	Pr_0				
	$\alpha = 1$	$\alpha = 2$	$\alpha = \mathbf{3}$	$\alpha = 4$	$\alpha = 5$
DET	0	0	0	0	0
UNI	0.5	0	0	0	0
EXP	0.3679	0.1353	**0.0498**	0.0183	0.0067
PAR (k=1/3)	0.2963	0.0787	0.0315	0.0156	0.0089

Step 2 (Determining β^* and Γ^*) is to obtain the best β^* and Γ^* candidates from the set identified from step 1. The criterion is based on minimizing the number of simultaneous wake-ups. There are two sub-steps to achieve the goal.

In the first sub-step, we search for the best Γ for each β^* candidate obtained in step 1. That is, for a given β_i obtained in step 1, we select the best Γ among

$\Gamma_{i,1}$, $\Gamma_{i,2}$ and $\Gamma_{i,3}$ that minimizes the number of simultaneous wake-ups. Since γ_js are integer valued, we can compute the least common multiple (LCM) for all the elements of each Γ^* candidate. Note that the LCM gives the minimal number of BIs for which two or more clients wake up simultaneously. Therefore, a larger LCM means a lower number of simultaneous wake-ups. We therefore choose the best Γ based on the largest LCM and denote this choice by Γ_i^*.

In the second sub-step, given $(\beta_i,\ \Gamma_i^*)$, $i = 1, \ldots, n+1$, we select the best Γ from the Γ_i^*s that will again minimize simultaneous wake-up. The criterion is based on the largest spread of their elements which is measured by the ratio of the standard deviation to the mean of the elements in Γ_i^*. Therefore, Γ^* is given by the Γ_i^* that gives the highest ratio, and β^* is the corresponding β_i.

Step 3 (Determining Θ^*) seeks to determine Θ^* according to the Γ^* obtained in the last step. The motivation is to mitigate the possible unfairness in the frame buffering delay experienced by the clients. The approach is to assign a smaller θ_j^* to the client with a larger γ_j^*. In this way, the client that wakes up less frequently will have a higher priority to retrieve its frames during channel contention. To do so, we use $\theta_j^* = 31 + \epsilon_\Theta(\max_{\forall j}(\gamma_j^*) - \gamma_j^*)$. Therefore, the client with the largest γ_j is assigned with the default value ($\theta_j^* = 31$), and others with smaller γ_j will backoff beyond the default.

4.4.2 The First-wake-up Schedule

An optional step is to schedule the first wake-up times of the clients, so that the maximal number of waking clients at one BI epoch is minimized. The optimization problem is given by

$$\min_{\mathbf{r}} \ : \ \max_{u=0,1,2,\ldots} N(\nu, \mathbf{r}, \Gamma^*). \tag{4.3}$$

The vector $\mathbf{r}$ presents the first wake-up times by the granularity of BI. The client s_j first wakes up at the r_jth BI epoch where $r_j \in [0, LCM(\Gamma^*) - 1]$ is an integer. The function $N(\nu, \mathbf{r}, \Gamma^*)$ is the number of waking clients at the νth BI epoch when the clients wake up according to $\mathbf{r}$ and Γ^*. And the first-wake-up schedule (FWS) adopts the optimal solution denoted as $\mathbf{r}^*$.

FWS could further decreases the probability of simultaneous wake-ups when several clients adopt the same γ_j^* or the elements of Γ^* have the same common factor. A simplest but typical example is about the two-client network whose $\Gamma^* = [2; 2]$. Without using FWS, two clients always wake up simultaneously and contest the channel source when their buffered frames exist. After using FWS, the AP assigns the first wake-up time for each client one by one (i.e., $\mathbf{r}^* = [0; 1]$), then two clients will never wake up in the same BI epoch to contest the channel. Therefore, FWS further improves energy efficiency by alleviating channel contentions and reducing collisions.

We have developed an algorithm to solve the optimization problem (4.3) which can be decomposed into a series of wake-up scheduling problems (WSPs) in [116]. Next, we introduce the WSP problem and then design our algorithm based on the stepwise solving method for WSP.

When a new PSM-enabled client s_j joins in an infrastructure network (with a set of m clients, denoted as $\mathcal{S}$) in the νth BI epoch, WSP is formulated to minimize the maximal number of waking clients in the following BI epoches. The optimization problem of WSP is given by

$$\min_{k_j(\nu)} \quad : \quad \max_{u=1,2,\ldots} \{N(\nu + u)\} \tag{4.4}$$

where $N(\nu+u)$ is the number of waking clients at the $(\nu+u)$th BI epoch, and the wake-up counter $k_j(\nu)$ records the remaining BIs that the client s_j will wake up.

Given the LI parameter of each client γ_i, $0 \leq k_i(\nu) \leq \gamma_i - 1$, $i \in \mathcal{S} \cup j$. Moreover, $N(\nu + u)$ equals to $\sum_{i \in \mathcal{S} \cup j} w_i(\nu + u)$ where the wake-up indicator $w_i(\nu + u)$ is 1 if s_i wakes up at the $(\nu + u)$th BI epoch; otherwise, it is 0. Then the stepwise solving method [116] is designed to calculate $k_j^*(\nu)$, the optimal wake-up counter of s_j. And $k_j^*(\nu)$ is a function $f(j, \gamma_j, m, w_i(\nu), k_i(\nu), \gamma_i)$, $i = 1, \ldots, m$.

It is easy to see that $k_j^*(\nu)$ is the optimal first wake-up time of s_j when $\nu = 0$, i.e., $r_j^* = k_j^*(0)$. Therefore, our algorithm obtains $k_j^*(0)$ client by client at $\nu = 0$ based on Γ^* and determines the FWS as $\mathbf{r}^* = [k_1^*(0), \ldots, k_c^*(0)]$. Using FWS, the client s_j first wakes up optimally at the r_j^*th BI epoch, $\forall j = 1, \ldots, c$. For example, if $r_j^* = 1$, s_j will miss the first beacon frame at the beginning of simulation but wake up for the first time after β^*. The detail steps of our algorithm for obtaining $\mathbf{r}^*$ is given below:

1 Initialize the following variables: $\nu = 0$, $\mathcal{S} = \{s_1\}$, $m = 1$, $w_1(\nu) = 1$, $k_1(\nu) = 0$, $r_1^* = 0$ and $j = 2$.

2 If $j > c$, return $\mathbf{r}^*$ and exit; else go to item 3.

3 Find the optimal wake up time of s_j, where

$$k_j^*(\nu) = f(j, \gamma_j^*, m, w_i(\nu), k_i(\nu), \gamma_i^*), i = 1, \ldots, m$$

4 Update variables: $r_j^* = k_j^*(\nu)$, $m = m + 1$ and $\mathcal{S} = \mathcal{S} \cup s_j$.

5 If $r_j^* = 0$, then $w_j(\nu) = 1$ else $w_j(\nu) = 0$.

6 Increase j by 1 and go back to item 2.

4.5 Performance Evaluation

In this section, we evaluate the performance of C-PSM by comparing with the IEEE 802.11 PSM scheme using default parameters (refered to as S-PSM). However, we do not compare C-PSM with other user-centric/AP-centric PSM schemes, because the study scope and the design objectives are different. C-PSM improves energy efficiency for all clients, whereas the user-centric PSM schemes consider only a single client. On the other hand, C-PSM is standard-compliant, but most AP-centric schemes are not compatible with the PSM scheme.

We use the four performance metrics (power saving, throughput, energy efficiency and frame buffering delay) to evaluate the performance of PSM-based schemes. The simulation results show that C-PSM has the advantages in saving power and improving the energy efficiency.

4.5.1 Evaluation Methodology

We define four indices in reference to S-PSM's performance to measure the improvement of one PSM-based scheme over S-PSM. S-PSM's quantities are labeled by superscript S whereas the given scheme's quantities do not have this superscript. They can be compared based on power saving (index η_P), throughput (index η_T), energy efficiency (index $\eta_{T/P}$) and frame buffering delay (index η_D). For easy comparison, we have defined them as below, such that a positive value indicates improvement over S-PSM. For example, compared with S-PSM, the given scheme saves more energy when $\eta_P > 0$, provides higher throughput when $\eta_T > 0$, achieves better energy efficiency when $\eta_{R_{T/P}} > 0$ and reduces the data frame's delay when $\eta_D > 0$.

$$
\begin{aligned}
\eta_P &= (P^S - P)/P^S \times 100\%, \\
\eta_T &= (T - T^S)/T^S \times 100\%, \\
\eta_{T/P} &= (R_{T/P} - R^S_{T/P})/R^S_{T/P} \times 100\%, \\
\eta_D &= \frac{1}{c} \times \sum_{j=1}^{c} (d^S_j - d_j)/d^S_j \times 100\%.
\end{aligned}
$$

Moreover, the larger index means the more significant improvement of the given scheme in terms of the corresponding metric. For example, for two PSM schemes A and B, if the η_P of A is larger than the η_P of B, then A will save more energy than B. Therefore, S-PSM serves as a benchmark. All PSM-based schemes can compare with each other based on the evaluation methodology.

4.5.2 Two Clients

We first evaluate C-PSM in a two-client system. To evaluate the effectiveness of different components of C-PSM, we consider three different versions. The first one is a "full version" which includes the FWS discussed in the last section. The other two (Scheme-1 and Scheme-2), however, exclude the FWS (i.e., $r_j = 0, \forall j$) and adopt the default congestion window size (i.e., $\theta_j = 31, \forall j$). The difference between Scheme-1 and Scheme-2 is that Scheme-2 adopts the default Γ value ($\gamma_j = 1, \forall j$). We also consider the S-PSM in the evaluation. To sum up, we will compare these four schemes in this subsection:

1. C-PSM : β^*, Γ^*, Θ^*,$\mathbf{r}^*$;
2. Scheme-1: β^*, Γ^*, $\theta_j = 31$, $r_j = 0$;
3. Scheme-2: β^*, $\gamma_j = 1$, $\theta_j = 31$, $r_j = 0$;

4. S-PSM : $\beta = 100$ms, $\gamma_j = 1$, $\theta_j = 31$, $r_j = 0$.

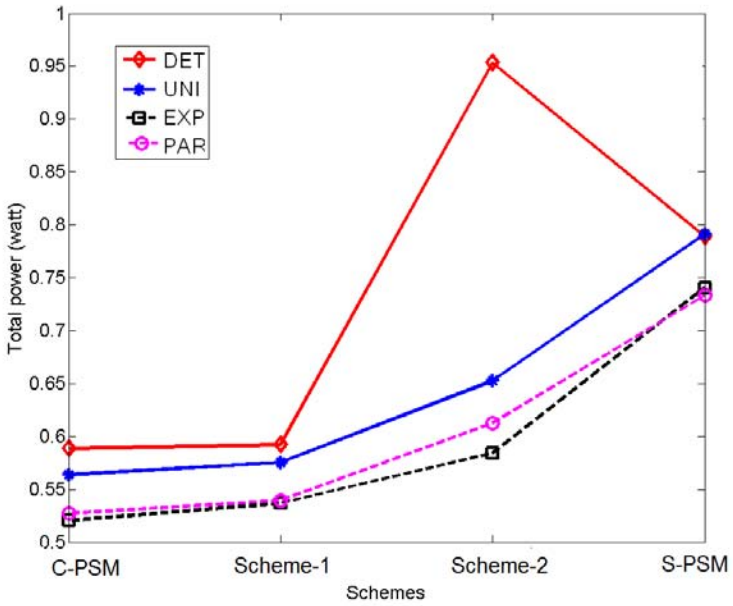
Given Δ, Table 4.5 shows the optimal PSM parameters obtained by the AP in C-PSM under different traffic distributions.

Table 4.5. Optimal parameters in C-PSM for $c = 2$ ($\epsilon_\beta = 2$ms and $\epsilon_\Theta = 8$).

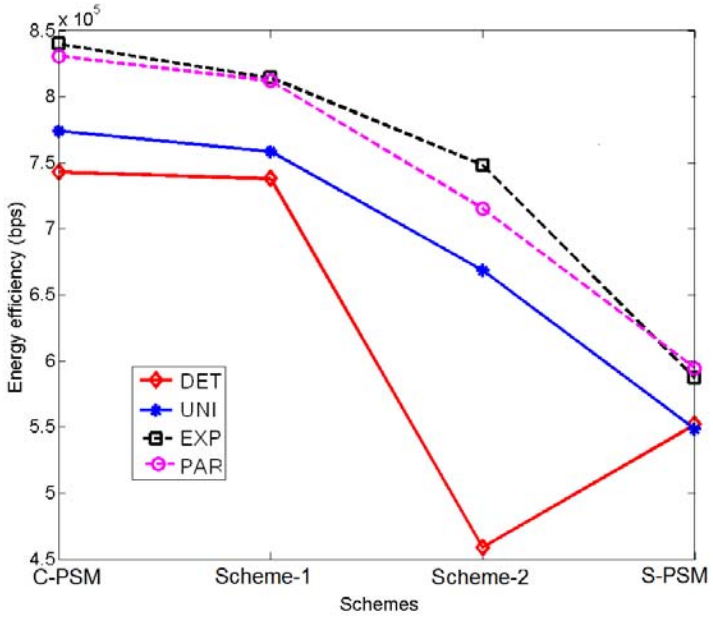
Δ(ms)	ρ	distribution	β^*(ms)	Γ^*	Θ^*	$\mathbf{r}^*$
$[15; 25]$	12.05%	DET	10	$[2; 3]$	$[39; 31]$	$[0; 0]$
		UNI	26	$[1; 2]$	$[39; 31]$	$[0; 0]$
		EXP,PAR	38	$[1; 2]$	$[39; 31]$	$[0; 0]$
$[15; 15]$	15.07%	DET	10	$[2; 2]$	$[31; 31]$	$[0; 1]$
		UNI	10	$[3; 3]$	$[31; 31]$	$[0; 1]$
		EXP,PAR	10	$[5; 5]$	$[31; 31]$	$[0; 1]$

Figure 4.11 depicts that C-PSM outperforms S-PSM on saving energy and improving energy efficiency under different distributions when $\Delta = [15; 25]$ms. C-PSM achieves lowest P and highest $R_{T/P}$ among four schemes. Note that $r_j^* = 0, \forall i$ for this scenario when γ_j $(\forall j)$ are relative prime. In this case, all clients wake up at the beginning of simulation. Therefore, C-PSM and Scheme-1 differ only in the adoption of Θ^*. And Scheme-1 performs a little worse than C-PSM, since it does not adopt Θ^*. Comparing with Scheme-1, Scheme-2 increases power and decreases energy efficiency, since it does not use Γ^*. After adopting β^*, Scheme-2 still outperforms S-PSM except under the DET distribution of traffic. Scheme-2 is the worst under the DET distribution of traffic, because two clients wake up much frequently (every $\beta^* = 10$ms) and both waste much energy on the unnecessary wake-ups.

As shown in Table 4.6, C-PSM's improvement over S-PSM can be very significant. For example, C-PSM reduces power consumption by 29.37%, improves energy efficiency by 43.01% and reduces average buffering delay by 54.8% under exponential traffic distribution. C-PSM's η_P and $\eta_{T/P}$ are higher than Scheme-1's



(a) Total power P.



(b) Total energy efficiency $R_{T/P}$.

Figure 4.11. A comparison of four PSM schemes for $c = 2$ and $\Delta = [15; 25]$ms.

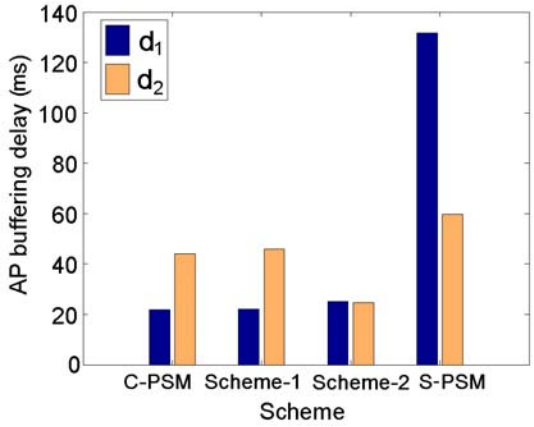
Table 4.6. Comparing the performance of C-PSM, Scheme-1 and Scheme-2 for $\Delta = [15; 25]$ms.

index, %	scheme	DET	UNI	EXP	PAR
η_P	C-PSM	25.41	28.75	29.73	28.13
	Scheme-1	24.91	27.28	27.53	26.47
	Scheme-2	-20.82	17.52	21.10	16.48
$\eta_{T/P}$	C-PSM	34.63	41.18	43.01	39.76
	Scheme-1	33.71	38.33	38.65	36.57
	Scheme-2	-16.88	21.95	27.38	20.30
η_D	C-PSM	82.33	68.79	54.80	54.18
	Scheme-1	82.08	68.15	53.07	53.56
	Scheme-2	94.54	79.79	69.88	68.75
η_T	C-PSM	0.41	0.59	0.50	0.45
	Scheme-1	0.40	0.60	0.48	0.42
	Scheme-2	0.43	0.58	0.50	0.47

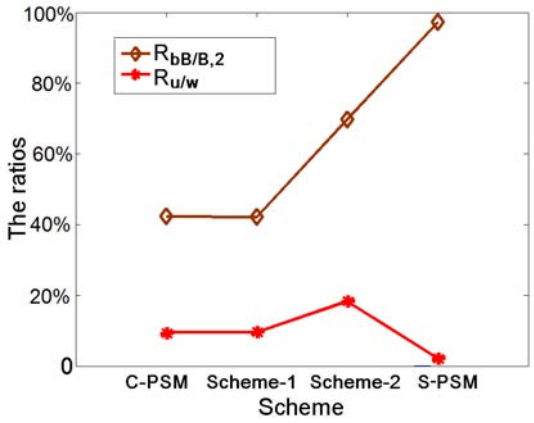
and Scheme-2's. Therefore, employing β^*, Γ^* and Θ^* gives the best performance in power saving and energy efficiency.

The results in Table 4.6 also show that the performance benefit of adopting both β^* and Γ^* is significant, because Scheme-1 outperforms Scheme-2 by a large margin in η_P, $\eta_{T/P}$ and η_D. Without using Γ^*, Scheme-2's performance is even worse than S-PSM for a small number of cases. On the other hand, the performance improvement due to Θ^* is minor, because C-PSM outperforms Scheme-1 by only a small percentage for most of the cases (the largest deviation is recorded in $\eta_{T/P}$ for EXP).

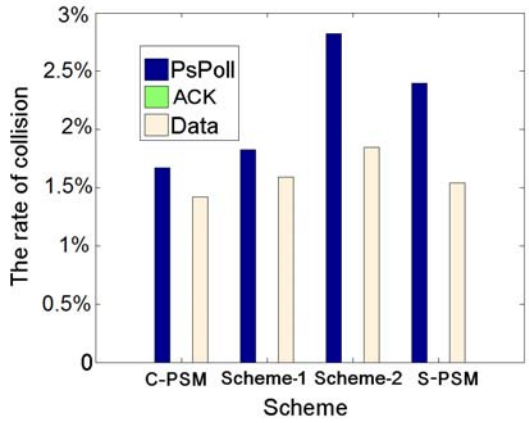
We next probe into the factors responsible for C-PSM's good performance. As an example, we examine the exponential traffic distribution. First, Figure 4.12(a) shows that C-PSM's AP buffering delay is shorter than S-PSM's. The clients using C-PSM can consequently return to the sleep mode earlier, because the frames with shorter buffering delay can be retrieved faster. C-PSM also improves the fairness of clients, since it greatly decreases the delay difference of two clients



(a) The AP buffering delay.



(b) The ratios of unnecessary wake-ups and simultaneous wake-ups



(c) The collision ratios.

Figure 4.12. Four schemes under the EXP traffic distribution for $\Delta = [15; 25]$ms.

compared with S-PSM. C-PSM speeds up the retrieval of frames in the fast client, for example, the delay of s_1 is one sixth of that in S-PSM. Moreover, Scheme-1's delay is only slightly higher than C-PSM's, suggesting that Θ^* does not have noticeable impact on the delay. On the other hand, Scheme-2 has shorter delay of the slow client s_2 than C-PSM. But its benefit is too small to affect the performance.

Second, the metric $R_{bB/B,2}$ in Figure 4.12(b) shows that C-PSM and Scheme-1 are successful in reducing the channel contention. Scheme-2 and S-PSM, on the other hand, suffer from heavy channel contentions. The clients using Scheme-2 are involved in channel contentions at 68% of BIs, i.e., $R_{bB/B,2} = 68\%$ while the $R_{bB/B,2}$ in C-PSM is only 40%. In particular, the contentions occur in almost every BI for S-PSM since $R_{bB/B,2} \approx 100\%$. The metric $R_{u/w}$ in Figure 4.12(b) also shows that C-PSM and Scheme-1 can reduce unnecessary wake-ups, as compared with Scheme-2. The clients in Scheme-2 frequently wake up every β^*, but almost 20% of the wake-ups are unnecessary. However, S-PSM's $R_{u/w}$ is very low, because the two clients simultaneously wake up to compete for channel in almost all epoches.

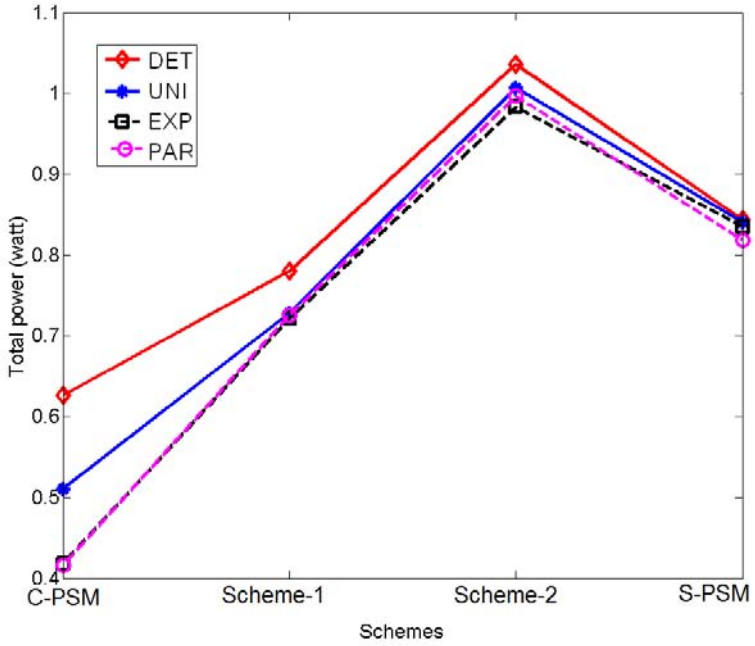
Third, Figure 4.12(c) illustrates that the collision ratios of PS-Poll and data frames in C-PSM are less than those in Scheme-1. It means that Θ^* is useful in reducing channel collisions. That is why C-PSM performs better than Scheme-1 slightly. We also find that the collision ratios in Scheme-2 are higher than those in C-PSM and Scheme-1. For example, the PS-Poll collision ratio in Scheme-2 is highest and nearly 1.5 times of that in C-PSM. Therefore, the clients using Scheme-2 are less energy-efficient since they have to spend more energy to deal with the collisions. Additionally, the collision ratios of ACK are much near to zero. The reason is that one awaken client always returns ACK after it has

finished receiving a data frame and a SIFS has elapsed. The channel is rarely occupied by the other client or the AP within such a short SIFS. Therefore, ACK rarely suffers from collisions especially when $c = 2$. If the number of clients increases, the probability of ACK collisions will increase, since the channel contention is more intensive.

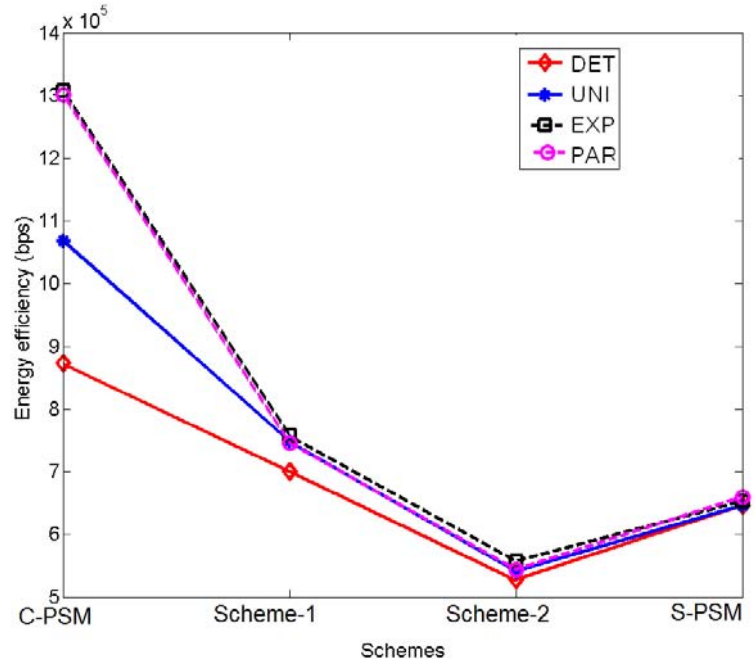
C-PSM is also applicable to the two-client system when the clients are symmetric. Shown in Table 4.5, C-PSM uses β^*, Γ^*, the default CW values and the nonzero vector $\mathbf{r}^*$ when $t = [15; 15]$ms. FWS is adopted, since γ_1^* is equal to γ_2^* and larger than 1. $\mathbf{r}^* = [0; 1]$ determines that s_1 firstly wakes up at the beginning of simulation while s_2 delays one β^* to wakes up for the first time.

C-PSM achieves the lowest P and the highest $R_{T/P}$ show in Figure 4.13. It performs best among these four schemes after adopting all optimal parameters. Especially, the simulation results validate that the ratio of simultaneous wake-ups and the total collision ratios are all zero in C-PSM. The reason is that two clients never wake up in the same BI epoch after using FWS. Without using FWS, Scheme-1 still outperforms S-PSM. But it has lower total power and higher total energy efficiency than C-PSM, which means that FWS is helpful to improve energy efficiency. Without using Γ^*, Scheme-2 is worse than S-PSM, since two clients wake up in all BI epochs ($\beta^* = 10$ms) and consume much energy on the frequent wake-ups.

Table 4.7 also shows that C-PSM outperforms other three schemes, since its indices are highest under all traffic distributions. The values of η_T are near zero while η_P, $\eta_{R_{T/P}}$ and η_D are up to 49.87%, 100.21% and 88.47% individually. That is, C-PSM largely reduces power consumption, increases energy efficiency and shortens the AP buffering delay while keeping the throughput. Without FWS, Scheme-1 using β^* and Γ^* only gets a little advantage over the standard scheme,



(a) Total power P.



(b) Total energy efficiency $R_{T/P}$.

Figure 4.13. A comparison of four PSM schemes for $c = 2$ and $\Delta = [15; 15]$ms.

Table 4.7. Comparing the performance of C-PSM, Scheme-1 and Scheme-2 for $\Delta = [15; 15]$ms.

index, %	scheme	DET	UNI	EXP	PAR
η_P	Our PSM	25.71	39.30	49.87	49.10
	Scheme-1	7.50	13.34	13.56	11.30
	Scheme-2	-22.88	-19.76	-17.72	-21.80
$\eta_{R_{T/P}}$	Our PSM	34.91	65.34	100.21	97.14
	Scheme-1	8.35	15.78	16.00	13.04
	Scheme-2	-18.44	-16.16	-14.64	-17.56
η_D	Our PSM	88.47	80.46	70.93	69.97
	Scheme-1	84.15	74.26	56.39	55.50
	Scheme-2	93.74	91.49	92.46	92.13
η_T	Our PSM	0.23	0.36	0.37	0.35
	Scheme-1	0.23	0.33	0.27	0.27
	Scheme-2	0.23	0.40	0.49	0.41

because η_P and $\eta_{R_{T/P}}$ are about 0.1. Therefore, FWS is significant to improve the PSM performance when the clients are symmetric. In contrast, Scheme-2 obtains the negative performance indices. It greatly increases the power consumption and reduces energy efficiency by using a small β^* and letting two clients frequently wake up in every BI epoch.

4.5.3 More Than Two Clients

We have applied C-PSM to the network including more than two PSM-enabled clients. Firstly, we evaluate C-PSM when the number of clients increases from 3 to 6 and the level of traffic is light where $\rho \in (10\%, 30\%)$. In these networks, two or more clients are symmetric (i.e., at least two clients have the same mean of inter-frame arrival times). According to each Δ, the main algorithm obtains the optimal parameters of C-PSM, shown in Table 4.8. C-PSM must adopt FWS, since some elements of Γ^* are not relative prime numbers or even the same. For example, four clients in C-PSM adopts $\Gamma^* = [1; 1; 2; 2]$ where γ_3 and γ_4 have a

greatest common divisor 2, and then FWS is $\mathbf{r}^* = [0;0;0;1]$. Such that s_1, s_2 and s_3 wake up at the beginning of simulation while s_4 defers the first wake-up time for one β^*.

Table 4.8. Optimal parameters in C-PSM for $c = 3 \sim 6$ ($\epsilon_\beta = 2$ms and $\epsilon_\Theta = 8$).

Δ, ms	ρ,%	distribution	β^*, ms	Γ^*	Θ^*	$\mathbf{r}^*$
[20;30;30]	13.18	DET UNI EXP,PAR	16 30 46	[1;2;2]	[39;31;31]	[0;0;1]
[20;20; 30;30]	18.83	DET UNI EXP,PAR	16 30 46	[1;1;2;2]	[39;39; 31;31]	[0;0;0;1]
[20;20;30; 30;30]	22.6	DET UNI EXP,PAR	14 28 40	[1;1;2;2;3]	[47;47;39; 39;31]	[0;0;0;1;0]
[20;20;20; 30;30;30]	28.25%	DET UNI EXP,PAR	14 28 40	[1;1;2;2;3;3]	[47;47;39; 39;31;31]	[0;0;0;1;0;1]

The positive indices in Table 4.9 show that C-PSM outperforms S-PSM in terms of power saving, energy efficiency and AP buffering delay while keeping or slightly increasing throughput when the number of clients changes from 3 to 6. For example, C-PSM in the three-client network reduces power consumption by 36.78%, improves the energy efficiency by 59.11% and shortens the average buffering delay by 52.16% under the EXP distribution of traffic while the total throughput is kept, since its η_T is near zero. On the other hand, the indices of C-PSM without FWS (in the column denoted as '-FWS') are less than the ones of C-PSM under all traffic distributions. For example, without using FWS, the $\eta_{T/P}$ of C-PSM at most decreases by 22% under the EXP distribution of traffic. Therefore, the FWS is much helpful to improve energy efficiency when the symmetric clients exist.

Table 4.9. Simulation results of C-PSM under different Δs for $c = 3, 4, 5, 6$.

Index	Δ(ms)	DET		UNI		EXP		PAR	
		C-PSM	-FWS	C-PSM	-FWS	C-PSM	-FWS	C-PSM	-FWS
η_P	[20;30;30]	36.38	29.00	39.08	30.08	36.78	26.43	36.31	27.33
	[20;20;30;30]	30.87	27.56	29.78	25.78	26.01	24.39	25.38	22.30
	[20;20;30;30;30]	30.50	26.38	25.66	21.64	25.41	20.74	23.97	20.51
	[20;20;20;30;30;30]	24.31	21.83	20.61	17.26	19.34	17.30	18.87	17.06
$\eta_{R_{T/P}}$	[20;30;30]	59.86	43.22	65.92	44.56	59.11	36.73	58.00	38.41
	[20;20;30;30]	48.09	41.35	45.67	37.84	36.72	33.84	35.89	30.50
	[20;20;30;30;30]	58.21	49.34	46.38	38.86	41.61	33.20	37.56	31.41
	[20;20;20;30;30;30]	46.17	41.44	38.32	32.66	30.71	27.52	30.81	27.89
η_D	[20;30;30]	84.00	81.80	68.69	64.62	52.16	45.23	51.98	46.37
	[20;20;30;30]	86.78	85.68	73.06	71.46	58.04	56.32	59.52	57.44
	[20;20;30;30;30]	87.78	86.42	76.50	74.79	63.97	60.82	64.07	61.27
	[20;20;20;30;30;30]	88.38	87.42	78.86	77.13	66.56	65.18	67.72	66.05
η_T	[20;30;30]	1.71	1.69	1.08	1.07	0.60	0.59	0.63	0.58
	[20;20;30;30]	2.37	2.40	2.29	2.31	1.16	1.20	1.40	1.40
	[20;20;30;30;30]	9.96	9.95	8.82	8.82	5.63	5.57	4.58	4.45
	[20;20;20;30;30;30]	10.63	10.57	9.82	9.77	6.43	6.47	6.13	6.08

5example, in the six-client network where ρ is as high as

Next, we compare the simulation results of C-PSM, C-PSM without FWS and S-PSM to explain the above findings. As an example, we study these three schemes under the EXP traffic.

Table 4.10. Comparing the performance of C-PSM, Scheme-1 and Scheme-2 under the EXP traffic distribution for $\Delta = [20; 30; 30]$ms.

metrics	C-PSM	C-PSM without FWS	S-PSM
P (Watt)	0.8242	0.9591	1.3037
T (10^5bps)	4.7538	4.7536	4.7257
$R_{T/P}$ (10^5bpJ)	5.7675	4.9563	3.6248
d_1(ms)	34.4	36.3	234.2
d_2(ms)	55.8	63.2	84.5
d_3(ms)	54.9	64.7	87.4
$R_{c/t}$	1.44%	1.78%	2.14%
$R_{u/w}$	10.06%	10.66%	4.86%
$R_{bB/B,2}$	83.83%	10.47%	7.64%
$R_{bB/B,3}$	**0**	39.05%	92.29%

Shown in Table 4.10, C-PSM saves energy by shortening the period of channel contention. The d_j ($\forall j$) of C-PSM are smallest than other two schemes. That is, each client can receive its buffered frames most quickly and then enter to sleep instead of spending much energy and time on idle mode during channel contention. C-PSM also saves energy by reducing channel contentions. It prevents all clients from simultaneously waking up to the channel, since $R_{bB/B,3}$ is zero while $R_{bB/B,3}$ is as high as 92.29% in S-PSM. Moreover, C-PSM consumes a small amount of energy on unnecessary wake-ups, since $R_{u/w}$ is near 10% and reduces collisions ($R_{c/t}$) by about one-third. From what has been discussed above, C-PSM outperforms S-PSM.

C-PSM without FWS outperforms S-PSM but degrades the performance of C-PSM. The total power increases and the total energy efficiency decreases. With-

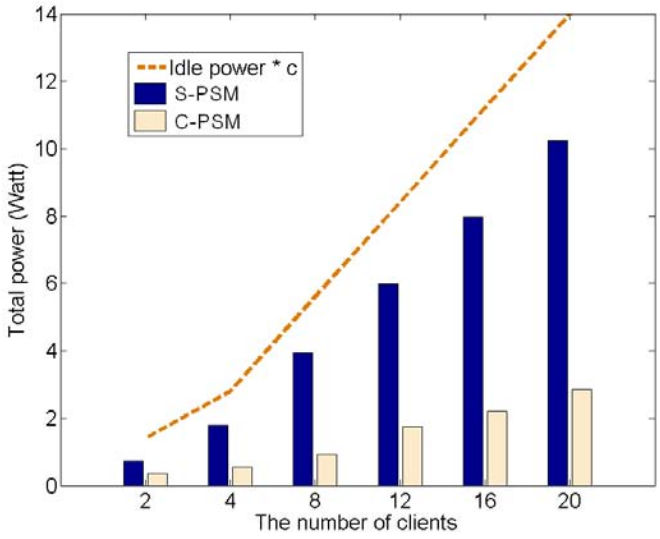


Figure 4.14. Total power verses c under the EXP traffic distribution for $\delta_j = 10c$ms, $(j = 1, \ldots, c)$.

out using FWS, three awaken clients compete for receiving data in 39.05% of BIs while C-PSM can totally avoid such a situation after using FWS. On the other hand, the $R_{bB/B,2}$ in the C-PSM without FWS is smaller than the one in C-PSM. It is helpful to save energy but does not dominate the energy consumption of all clients. The reason is that more energy is consumed on channel contention when all of three clients wake up simultaneously. In a global view, FWS is helpful to improve energy efficiency, because the number of clients which wake up at the same beacon epoch has been minimized.

Furthermore, we find that C-PSM is applicable for a large scale network and saves more energy when the number of clients increases. These two sets of simulations evaluate the performance of C-PSM when the number of PSM-enabled clients increases up to 20.

In the first set of simulations, we let $\delta_j = 10c$(ms), $j = 1, \ldots, c$. The total

amount of traffic of all symmetric clients will not change with c, since the total arrival rate of frames λ always equals 100 frames per second. The traffic workload is light since $\rho = 11.3\% < 30\%$. Under light traffic, S-PSM can save energy, since the average power of each client is always lower than its idle power. For example in Figure 4.14, P^S is much less than the c times of client's idle power under the EXP distribution of traffic. C-PSM scheme can further reduce energy consumption, since the power difference between P^S and P^* is always larger than zero. Moreover, this power difference increases with c. Therefore, C-PSM saves more energy when the number of clients increases.

Table 4.11. The indexes of C-PSM verses the number of clients for $\delta_j = 10c$ms $(j = 1, \ldots, c)$.

Index,%	T_j	c=2	c=4	c=8	c=12	c=16	c=20
η_P	DET	22.30	63.75	79.20	71.54	75.52	72.98
	UNI	45.37	72.52	78.82	70.58	71.74	72.22
	EXP	51.09	70.33	76.07	70.98	72.27	72.14
	PAR	50.76	70.43	76.68	69.77	70.91	70.33
$\eta_{R_{T/P}}$	DET	29.07	177.76	396.05	263.65	325.31	286.14
	UNI	83.50	265.28	384.96	251.95	270.96	281.65
	EXP	105.04	238.69	327.07	257.23	277.64	281.02
	PAR	103.89	239.39	338.50	241.89	260.80	255.76
η_D	DET	92.35	87.63	85.78	87.47	90.18	89.21
	UNI	73.94	70.37	84.07	85.57	87.28	88.11
	EXP	64.09	70.89	82.68	85.02	87.04	88.43
	PAR	62.93	69.48	82.25	85.78	87.10	88.19
η_T	DET	0.29	0.70	3.17	3.51	4.10	4.32
	UNI	0.24	0.40	2.72	3.55	4.83	6.02
	EXP	0.29	0.48	2.19	3.68	4.73	6.14
	PAR	0.40	0.37	2.27	3.36	4.94	5.55

Table 4.11 also shows that C-PSM achieves the significant improvements of four main metrics in a large network. For example, the indices of η_P, $\eta_{T/P}$ and η_D in are as high as 76.07%, 327.07% and 82.68% individually when $c = 8$ under the EXP distribution of traffic. Moreover, C-PSM scheme improves clients'

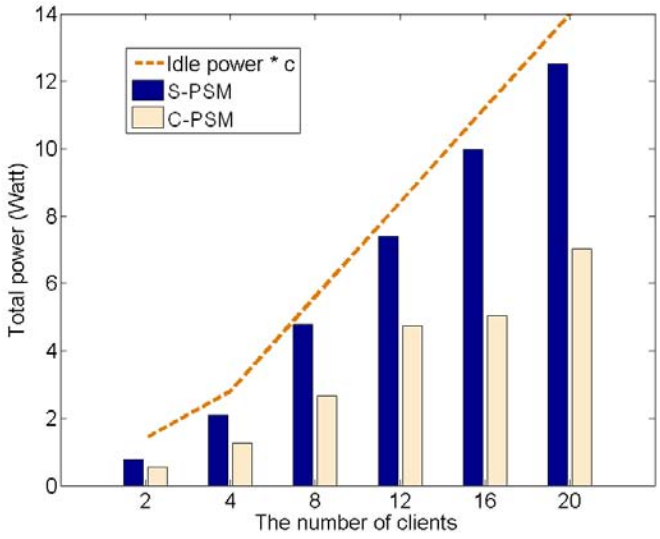


Figure 4.15. Total power verses c under the EXP traffic distribution for $\delta_j = 10 + 5j$ms $(j = 1, \ldots, c)$.

throughput in a large network, since η_T slightly increases with the number of clients.

In the second set of simulations, we let $\delta_j = 10 + 5j(\text{ms})$, $j = 1, \ldots, c$. The total frame arrival rate increases with the number of clients where ρ increases from 13.18% to 49.51%. When $c \geq 8$, the traffic is not light any more since $\rho \geq 32\%$.

We firstly find that C-PSM has a wider applicability than S-PSM. When the network supports many clients whose workload is not light, C-PSM is effective to save energy. But S-PSM cannot save much energy, since the clients are too busy to sleep. For example in Figure 4.15 under the EXP distribution of traffic, P^S is much close to the c times of client's idle power when $c \geq 8$. However, P^* is much less than the c times of client's idle power even when c increases to 20. That is, C-PSM is still effective to save energy even when the network workload

is as high as $\rho \approx 50\%$. Moreover, C-PSM saves more energy when the number of clients increases, since the power difference $P^S - P^*$ increases with c. Although not shown here, the similar simulation results are obtained under other traffic distributions.

Table 4.12. The indexes of C-PSM verses the number of clients for $\delta_j = 10+5j$ms $(j = 1, \ldots, c)$.

Index,%	T_j	c=2	c=4	c=8	c=12	c=16	c=20
η_P	DET	1.01	29.04	42.17	44.31	51.49	52.71
	UNI	21.93	37.81	44.29	37.34	51.98	50.97
	EXP	28.98	39.73	44.20	36.01	49.44	43.89
	PAR	25.64	37.59	44.19	35.65	48.85	43.49
$\eta_{R_{T/P}}$	DET	0.55	55.54	132.72	174.52	248.45	286.48
	UNI	28.79	76.14	138.29	137.83	245.38	255.72
	EXP	41.51	75.73	124.41	115.23	201.97	172.88
	PAR	35.07	71.90	126.80	114.65	203.60	173.49
η_D	DET	96.60	94.98	95.59	93.39	94.00	91.36
	UNI	76.05	79.90	85.27	85.56	86.88	85.72
	EXP	66.12	69.74	77.41	77.21	79.52	76.21
	PAR	65.05	69.82	77.69	77.10	79.66	76.30
η_T	DET	0.45	10.38	34.58	52.88	69.04	82.76
	UNI	0.55	9.54	32.75	49.03	65.84	74.41
	EXP	0.49	5.91	25.23	37.74	52.68	53.12
	PAR	0.43	7.29	26.58	38.13	55.28	54.56

Table 4.12 also shows that C-PSM outperforms S-PSM on saving power, improving energy efficiency, shortening delay and increasing throughput. When the traffic is not light (i.e., $c \geq 8$), C-PSM improves energy efficiency a lot, since it not only reduces power consumption but also increases throughput greatly. For example, compared with S-PSM, C-PSM saves 49.44% of power, increases 52.68% of throughput and then finally achieves 201.97% higher energy efficiency in the sixteen-client system under the EXP distribution of traffic.

4.6 Effects of Power Consumption Model on C-PSM

The power profile of wireless devices has a great impact on the performance of energy-saving scheme using sleeping [129]. This profile includes the client's power consumption in the transmission/reception/idle/sleeping mode and the cost of the mode transition (when the client wakes up from the sleeping mode to active mode). The energy consumed on the client's wake-up is the product of wake-up power[4] and wake-up time. The set of the above parameters are defined as a power consumption model in this thesis.

We compare several power consumption models and explain why model $\mathcal{A}$ is selected in our simulator. Next, we study the effects of these power consumption models on the performance of C-PSM. The improvements of C-PSM over S-PSM mainly depend on the client's wake-up energy consumption (e_w) and the ratio of idle power to sleeping power denoted as $R_{I/S}$.

4.6.1 Power Consumption Models

Table 4.13 lists several power consumption models discussed in this section. They have been widely used in the previous PSM studies.

We adopt model $\mathcal{A}$ in the simulator, which is typical and widely used. Model $\mathcal{A}$ is comparable to the hardware characteristics of many popular wireless interface cards. The ratio of transmission power to reception power in model $\mathcal{A}$ is approximately 160% which is similar to this ratio in ORiNOCO 11a/b/g Combo-

[4]During the mode transition, the client's power consumption is near or higher than the transmission power [163]. It could be estimated as two times of the idle power [94] (e.g., module $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{E}$).

Table 4.13. Five power consumption models.

Item	model $\mathcal{A}$ [58, 119]	model $\mathcal{B}$ [94, 157, 118]	model $\mathcal{C}$ [105, 13]	model $\mathcal{D}$ [91]	model **E** [32]
Transmission power	1.4W	1.65W	0.75W	1.3W	0.85W
Reception power	0.9W	1.4W	0.75W	0.95W	0.85W
Idle power	0.7W	1.15W	0.75W	0.79W	0.85W
Sleeping power	0.06W	0.045W	0.05W	0.17W	0.005W
$R_{I/S}$	1167%	2556%	1500%	468%	17000%
Wake-up power	0.7*2W	1.15*2W	0.75W	0.51W	0.85*2W
Wake-up time	2ms	2ms	2ms	13ms	2ms
e_w	0.003J	0.005J	0.0015J	0.0066J	0.0034J

Card [142], ORiNOCO 11a/b/g PCI card [143], CISCO AIRONET 802.11A/B/G Wireless Cardbus adapter [49], CISCO AIRONET 350 Series Wireless LAN Client Adapters [50] and Aironet's PC4800 PCMCIA NIC [57]. In model $\mathcal{A}$, the reception power is near to the idle power, just like in the CISCO AIRONET 802.11A/B/G Wireless Cardbus adapter [49] and Aironet's PC4800 PCMCIA NIC [57]. Moreover, the sleeping power is about an order of magnitude lower than the idle power in model $\mathcal{A}$, and $R_{I/S} \approx 1200\%$ is also common in many popular wireless network interface cards.

Model $\mathcal{B}$ is also widely applied in the PSM studies [94, 157, 118][5]. But the relationships of the power parameters in model $\mathcal{B}$ are quite different from the ones in many popular products. We just study model $\mathcal{B}$ by comparing it with model $\mathcal{A}$. Here, we conservatively estimate e_w as $\frac{2}{1000} \times 1.15W \times 2 = 0.005J$, which is near two times of the e_w in model $\mathcal{A}$. And the $R_{I/S}$ in model $\mathcal{B}$ is approximately two times of the $R_{I/S}$ in model $\mathcal{A}$.

Model $\mathcal{C}$ [105, 13] models the power usage as 0.75W while the client is active (i.e., sending data, receiving data and idle) and 0.05W while the client is

[5]Other papers [108, 134] uses model $\mathcal{B}$ but sets the power consumption of one client as 1.5W when it actively sends or receives frames.

asleep. However, 802.11 NICs consume somewhat more power while sending and receiving data than while idle in reality. Therefore, we should not adopt this rough model in the simulator. According to the measurement results of Enterasys RoamAbout NIC [105], we obtain $e_w = 0.0015J$, which is about a half of the e_w in model $\mathcal{A}$.

Model $\mathcal{D}$ is obtained by the measurements of Aironet 350 NIC [91] where the sleeping power is as high as 0.17W and $R_{I/S}$ is lowest. On the other hand, model $\mathcal{E}$ is obtained by the specifications of SDIO Wireless LAN Card [32] where the sleeping power is very low, 0.005W. We compare model $\mathcal{A}$ with these two models to study the effects of the sleeping power on C-PSM.

4.6.2 Performance of C-PSM in Different Power Consumption Models

We study the performance of C-PSM in five different power consumption models. As an example, we only consider the two-client system ($\Delta = [15; 25]$, $\rho < 30\%$) under different distribution of light traffic. As mentioned before, C-PSM could improve the energy efficiency of clients by reducing the channel contention. Firstly, the clients save energy, since they consume much low power in the sleeping mode, instead of consuming high power in the idle mode. If the idle power is low or even near to the sleeping power (i.e., $R_{I/S}$ is small), the clients will benefit a little on saving energy from C-PSM. Secondly, in order to reduce the simultaneous wake-ups of clients, C-PSM generally adopts a small BI (i.e., $\beta < 100$ms), and then the number of clients' wake-ups increases. If e_w increases, the energy saving due to C-PSM will also reduce.

We have used Table 4.14 to show that the advantages of C-PSM decrease with

e_w and increase with $R_{I/S}$. By comparing model $\mathcal{A}$ and model $\mathcal{C}$, we find that η_P and $\eta_{R_{T/P}}$ decrease with e_w when $R_{I/S}$ is similar. By comparing model $\mathcal{A}$ and model $\mathcal{E}$, we find that η_P and $\eta_{R_{T/P}}$ increase with the $R_{I/S}$ when e_w is similar. By comparing model $\mathcal{A}$ and model $\mathcal{D}$, we also find that the C-PSM achieves more improvements when e_w decreases and $R_{I/S}$ increases. Moreover, C-PSM achieves the greatest improvement over S-PSM in model $\mathcal{C}$, since its η_P and $\eta_{R_{T/P}}$ are highest under all traffic distributions. Comparing model $\mathcal{C}$ with other models, the $R_{I/S}$ is middle and e_w is lowest.

Table 4.14. Indexes of C-PSM in different power consumption models for $c = 2$ and $\Delta = [15; 25]$.

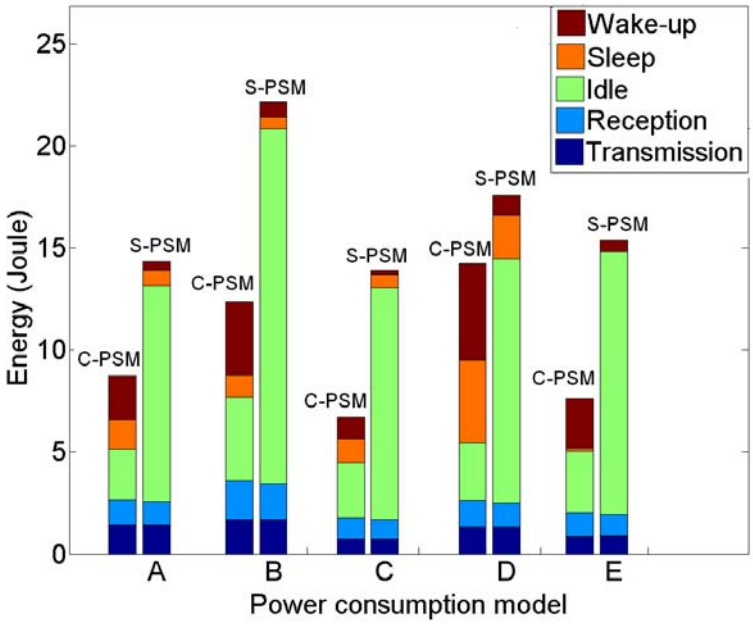
Index	T_j	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\mathcal{D}$	$\mathcal{E}$
η_P	DET	26.16	30.86	46.46	-4.86	38.44
	UNI	27.72	32.32	42.12	5.65	38.73
	EXP	28.99	33.72	39.63	12.69	39.58
	PAR	29.56	34.37	40.31	13.02	40.30
$\eta_{R_{T/P}}$	DET	36.25	45.52	87.91	-4.05	63.44
	UNI	39.01	48.46	73.58	6.48	63.98
	EXP	41.29	51.37	66.17	14.91	66.04
	PAR	42.30	52.71	67.93	15.24	67.91

Next, we analyze the energy consumption on different modes to explain the above findings. As an example, we only discuss the simulation results under the EXP distribution of traffic. Shown in Figure 4.16(a), the fast client s_1 using C-PSM (since $\delta_1 < \delta_2$) consumes least total energy in model $\mathcal{C}$. The ratio of s_1's total energy using C-PSM to s_1's total energy using S-PSM is near 50%, which is highest among five power consumption models. Therefore, s_1 using C-PSM achieves the largest improvements in model $\mathcal{C}$. The reason is that the $R_{I/S}$ is not large and the e_w is lowest in model $\mathcal{C}$. Using C-PSM, s_1 greatly reduces the energy consumed on the idle mode while slightly increases the energy consumed on the sleeping mode and the wake-up duration. On the other hand, s_1 using C-

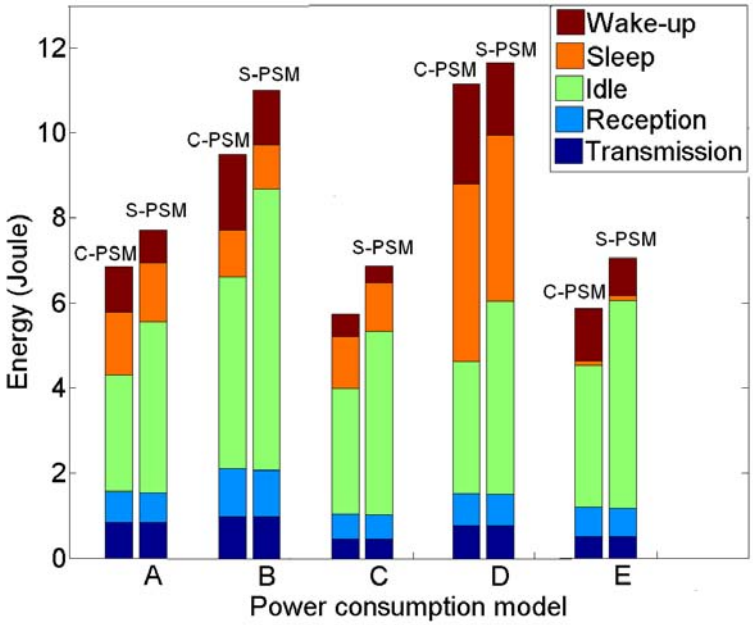
PSM in other power consumption models reduces the energy consumption on idle while largely increasing the energy consumption on other modes. For example, in model $\mathcal{A}$, model $\mathcal{B}$ and model $\mathcal{E}$, C-PSM obviously increase s_1's energy consumed on its wake-ups, since the e_w is very high. In model $\mathcal{D}$, C-PSM increases s_1's energy consumed on the sleeping mode, since its $R_{I/S}$ is very low.

Shown in Figure 4.16(b), the slow client s_2 using C-PSM consumes least total energy in model $\mathcal{C}$. Therefore, both clients using C-PSM achieve the lowest average power consumption in model $\mathcal{C}$. But the effects of C-PSM on the slow client are small, since the energy consumption of s_2 in C-PSM is similar to the one in S-PSM under all power consumption models. The main source of s_2's energy saving in C-PSM is the reduction of energy in the idle mode. Here, the inefficiency of C-PSM is the increase of s_2's energy on its wake-ups. For example in model $\mathcal{D}$, C-PSM cannot save much energy, since the e_w is so high that s_2 consumes 1.5 times of energy on its wake-ups compared with S-PSM.

Finally, we should point out that C-PSM is available to improve energy efficiency of clients if $R_{I/S}$ is large and e_w is small. Although we have not provided their criteria in quantity, our simulation results have shown that C-PSM can obviously outperforms S-PSM when $R_{I/S} > 1000\%$ and $e_w < 0.006J$. Among these five power consumption models, model $\mathcal{D}$ ($R_{I/S} = 468\%$ and $e_w = 0.0066J$) does not satisfy the criteria. Just in model $\mathcal{D}$, C-PSM is worse than S-PSM under the DET traffic (since $\eta_P < 0$ and $\eta_{R_{T/P}} < 0$) and only achieves a little improvement under other three distributions of traffic (since $\eta_P < 14\%$ and $\eta_{R_{T/P}} < 16\%$).



(a) s_1's energy consumption using C-PSM and S-PSM.



(b) s_2's energy consumption using C-PSM and S-PSM.

Figure 4.16. Clients' energy consumption under the EXP traffic distribution in different power consumption models for $c = 2$ and $\Delta = [15; 25]$ms.

Chapter 5

Conclusions and Future Work

In this thesis, we have considered the problem of improving the energy efficiency for all clients in an IEEE 802.11 infrastructure network. We have proposed and evaluated two access point (AP)-centric schemes: one based on the power saving mode (PSM) and the other on the transmission power control (TPC). Both PSM and TCP are the main instruments for saving clients' energy. The PSM allows an idle client to enter into the low-energy sleeping mode, whereas the TPC enables a client to reduce the transmission power by operating in a lower rate. However, many previous works concentrated on using them to reduce a single client's energy consumption and did not consider the impact of the interaction among multiple clients on their energy performance. Our main approach is to exploit the AP's knowledge about the clients' traffic patterns to optimize the TPC and PSM for optimizing all-client energy efficiency.

Firstly, we have optimized energy efficiency for *all* wireless clients in a polling-based wireless network. The energy-efficient design has been deployed in the AP which employs both the PSM and TPC to conserve energy as much as possible. Since the decrease in one's transmission power could adversely affect other

clients, we have formulated an optimization problem to minimize the energy consumption during a polling cycle and, at the same time, to ensure that all clients are stable, i.e., their queue lengths would not go unboundedly. The resulted stability-constrained optimization formulation enables the AP to compute the optimal power allocations for two polling schedules: phase grouping and mobile grouping. The experiment results have shown that the mobile grouping schedule is much more energy efficient, because it decreases the number of mode transitions and allows a client to sleep for a longer time. Using simulation, we have also investigated the impact of the optimal power allocations on the queueing delay. The results show that the optimal power allocation does not degrade the delay for over 90% of the time. Even when the average delay becomes longer, the tradeoff is still considered beneficial as a whole. Furthermore, the simulation results also show that an uncooperative user should not benefit from the delay and energy performance when it changes the optimal transmission power. That is, there are incentives for users to adopt the optimal power allocations computed by the iterative algorithm.

There are a few directions to extend this work. For example, a logical next step is to design a polling schedule that would optimize the network's energy efficiency. The iterative algorithm devised in this thesis serves a good starting point toward that goal. It would also be useful to investigate the impact of the TPC on the throughput of the individual clients and of the entire network. Additionally, game theory is a good tool to analyze the complicated relationship among wireless clients which compete for the same channel resource. We may study whether an equilibrium state for all wireless clients exists in terms of one performance metric, for example, delay, throughput and average power consumption.

Secondly, we have shown that the frequent unnecessary wake-ups of client, the

simultaneous wake-ups of clients and channel collisions erode energy efficiency and delay. Therefore, the avoidance or reduction of them will help improve the PSM's performance significantly. We have presented the centralized PSM (C-PSM) to accomplish this goal. The AP obtains optimal beacon interval and listen intervals based on the traffic characteristics of the clients. The jointly optimized intervals could avoid frequent unnecessary wake-ups and reduce simultaneous wake-ups which collectively translates into energy saving and reduction in the buffering delay. Moreover, C-PSM obtains optimal congestion windows to reduce buffering delay for clients assigned with large listen interval. It could also increase energy efficiency by the first-wake-up schedule, which determines the first wake-up times for clients, such that simultaneous wake-ups and channel collisions could be reduced further. Extensive simulation results show that C-PSM could obviously save more energy, increase energy efficiency and reduces the AP buffering delay under different traffic patterns compared with the standard PSM. When the number of clients increases, C-PSM still outperforms the standard PSM a lot. Furthermore, we have found that the advantages of C-PSM over the standard PSM depends on the power consumption model. C-PSM can obviously increases clients' energy efficiency only when the client's wake-up energy consumption is small, and the ratio of idle power to sleeping power is large.

In future works, we will improve the control algorithm by studying how to configure its input parameters. According to our simulation studies, their settings have a great impact on the scheme's performance. We will also extend the scheme to support more traffic models and distributions for the packet size, burst size, and packet intervals. At the same time, we will extend our C-PSM into many real application environments. For example, we will study a real wired-wireless network where the senders are equipped with feedback channels, such as TCP

senders. We will evaluate whether C-PSM is also effective for this type of senders. In this case, the AP should consider the characteristics of wired path when it determines the set of optimal parameters.

Appendix A

A List of Notations

Table A.1: Notations used in chapter 3.

Symbols	**Meanings**
e_c	Column vector including c unit elements
$\mathbf{E}[\ ]$	Expectation operation
c	Number of clients in a WLAN
C_{max}	Optimal channel capacity in bps
R	Real channel capacity in bps
α	Ratio of R to C_{max}
W	Bandwidth in Hz
S	Signal power consumption
N	Noise power consumption
P	Transmission power consumption of one device
A	Attenuation which is the ratio of P to S

Symbols	Meanings
M_i	The i^{th} wireless client, $i \in \{1, \ldots, c\}$
q_i	The i^{th} queue, $i \in \{1, \ldots, c\}$
$\boldsymbol{SW}$	Walk times of queues with the mean s_i for q_i
SW_0	Walk time of queues in PG's downlink phase with the mean s_0
$\boldsymbol{Tx}$	Transmission periods of queues, Tx_i for q_i
$\boldsymbol{Rx}$	Receiving periods of queues, Rx_i for q_i
$\boldsymbol{\Theta}$	Busy periods of queues, Θ_i for q_i
$\boldsymbol{V}$	Vacation periods of queues, V_i for q_i
C	Cycle period
$\boldsymbol{E}$	Energy consumption of all queues per cycle, E_i for q_i
$\boldsymbol{F}$	Sizes of packets for queues, F_i for q_i
N_i	Noise power of the wireless channel for q_i
g_i	Power gain of the wireless channel between AP and q_i
$\boldsymbol{K}$	$K_i = g_i N_i$ for q_i
$\boldsymbol{H}$	$H_i = 2\mathbf{E}[F_i]/\alpha W$ for q_i
$\boldsymbol{\lambda_d}$	Vector of arrival rates for downlink, $\lambda_{d,i}$ for q_i
$\boldsymbol{\lambda_u}$	Vector of arrival rates for uplink, $\lambda_{u,i}$ for q_i
$\boldsymbol{\lambda}$	Vector of arrival rates, $\lambda_i = \lambda_{d,i} + \lambda_{u,i}$ for q_i
$\boldsymbol{\beta}$	Vector of downland percentage, β_i for q_i
$\mathbf{b}$	Vector of average service times for all packets, b_i for q_i
$\mathbf{b_u}$	Vector of average service times for downlink packets, $b_{u,i}$ for q_i
$\mathbf{b_d}$	Vector of average service times for uplink packets, $b_{d,i}$ for q_i in MG
$\overline{b_D}$	Average service time for downlink packet in PG
ρ	Vector of workload (i.e., utilization ratio) of queues , ρ_i for q_i

Symbols	Meanings
P_{max}	Maximal power consumption of client
P_{min}	Minimal power consumption of client
P_R	Receiving power consumption of client, constant
P_I	Idle power consumption of client, constant
P_V	Sleeping power consumption of client, constant
P_{AP}	Transmission power consumption of the AP
$\mathbf{P}$	Vector of transmission power in clients
$\mathbf{T_D}$	Vector of average delay of downlink packets in queues
$\overline{T_D}$	Average delay of all downlink packets
$\mathbf{T_U}$	Vector of average delay of uplink packets in queues
$\overline{T_U}$	Average delay of all uplink packets
$[\mathbf{b_{min}}, \mathbf{b_{max}}]$	Feasible region of average service times
$[\mathbf{b^P_{min}}, \mathbf{b^P_{max}}]$	Region of average service times constrained by power limitations
$[\mathbf{b^\rho_{min}}, \mathbf{b^\rho_{max}}]$	Region of average service times constrained by stability
n	Iteration time
σ	Step size
ε_1, ε_2	Accuracy tolerance parameters in the iterative algorithm
$^{(0)}$	Superscript for the initializations in the iterative algorithm
*	Superscript of optimal solutions/schedules
$\mathbf{P^*_A}$	Optimum of $\mathbf{P}$ obtained from the algorithm
$\mathbf{P^*_N}$	Optimum of $\mathbf{P}$ obtained from the exhaustive method
e_P	Differential rate between $\mathbf{P^*_A}$ and $\mathbf{P^*_N}$
d	Average delay

Symbols	Meanings
σ_T	Delay variance
r_e	Energy inflation ratio
r_d	Delay inflation ratio
ERR	Energy reduction ratio
DRR	Delay reduction ratio
z	Energy efficiency metric of all clients, $z = \frac{\mathbf{E}[C]\sum_{i=1}^{c}\lambda_i F_i}{\mathbf{E}[E]}$ (bpJ)
φ	Metric of delay-energy tradeoff
ξ	Abbreviation of energy consumption of $\mathbf{E}[E]$
l	Arrival rate of one queue in a symmetric system
q_T	Typical queue
$_t$	Subscript for results of the typical queue
$_o$	Subscript for results of other queues
$_s$	Subscript for results of the whole system
$\_MG$	Subscript for results of MG schedule
$\_PG$	Subscript for results of PG schedule

Table A.2: Notations used in chapter 4.

Symbols	Meanings
c	Number of clients in a infrastructure network
s_j	The jth wireless client, $j = 1, \ldots, c$
$\mathbf{T}$	Vector of inter-packet arrival time distributions for c clients, T_j is for s_j
Δ	Vector of average inter-packet arrival times for c clients, δ_j is for s_j
λ	Total arrival rate of frames (in packets per second)

Symbols	Meanings
ρ	Workload/utilization ratio of a network
b	Mean service time of one frame with PSM
b_{min}	Minimum service time for one data frame
β	Beacon interval
Γ	Vector of listen interval parameters, γ_j is for s_j
Θ	Vector of minimal contention windows, θ_j is for s_j
t_u	Length of time unit in the simulator
t_s	Length of time slot in IEEE 802.11 protocol
t_{sim}	Simulation period
m_p	Maximal number of packets which arrive at one client during t_{sim}
$\boldsymbol{A}$	Matrix of packet arrival times including c rows and $maxPacket$ columns
β_{min}	Minimal length of BI
n	Number of BI candidates
ϵ_β	Step size of BI candidate
ϵ_Θ	Step size of CW
$\boldsymbol{L}$	Vector of estimated LIs of clients, L_j is for s_j
α_j	scaling factor of LI, i.e., $L_j = \alpha_j \delta_j$ for s_j
Pr_0	Empty probability that no packet for one client arrives within one LI
ξ	Threshold of empty probability
β_i	The i^{th} BI candidate, $i = 1, \ldots, n$
$\Gamma_{i,v}$	Three LI candidates for β_i, $v = 1, 2, 3$
Γ_i^*	The best LI candidate for β_i

Symbols	Meanings
τ	Index of unit
ν	Index of BI epoch
$\mathcal{S}$	Set of clients in the WSP problem
m	Number of clients in $\mathcal{S}$
$k_i(\nu)$	The wake-up counter of client i at the ν^{th} BI epoch
$w_i(\nu+u)$	The wake-up indicator of client i at the $(\nu+u)^{th}$ BI epoch, $u=1,\ldots$
$N(\nu+u)$	Number of wake-up clients after the $(\nu+u)^{th}$ BI epoch, $u=1,\ldots$
β^*	Optimal BI
Γ^*	Optimal vector of LI parameters, and γ_j^* is for s_j
Θ^*	Optimal vector of CWs, and θ_j^* is for s_j
$\mathbf{r}^*$	Optimal vector of the first wake-up times for clients, and r_j^* is for s_j
P	The total power consumed by the clients in watt (W)
T	The total client throughput in bits per second (bps)
$R_{T/P}$	$\frac{T}{P}$, the total energy efficiency metric in bits per joule (bpJ)
d_j	The frame buffering delay of s_j's frames at the AP in millisecond
N_u	The total number of unnecessary wake-ups by all clients
N_w	The total number of wake-ups by all clients
$N_{u,j}$	The number of unnecessary wake-ups by s_j
$N_{w,j}$	The number of wake-ups by s_j
$R_{u/w}$	$\frac{N_u}{N_w}$, the ratio of unnecessary wake-ups by all clients
$R_{u/w,j}$	$\frac{N_{u,j}}{N_{w,j}}$, the ratio of unnecessary wake-ups by s_j

Symbols	Meanings
$N_{bB,k}$	The number of BIs when k clients are involved in contention, $k \geq 2$
N_B	The number of BIs
$R_{bB/B,k}$	$\frac{N_{bB,k}}{N_B}$, ratio of the simultaneous wake-ups by k clients
N_c	The total number of collided frames by AP and all clients
N_t	The total transmission attempts by AP and all clients
$R_{c/t}$	$\frac{N_c}{N_t}$, ratio of the collided frames
η_P	Performance index of power-saving
η_T	Performance index of throughput
$\eta_{R_{T/P}}$	Performance index of energy efficiency
η_D	Performance index of AP buffering delay
$R_{I/S}$	Ratio of idle power to sleeping power
e_w	Energy consumption of one client when it wakes up, in joule

Appendix B

Variables and Flow Charts in the Simulator

Section 4.2 has presented the simulator architecture and introduced the main functions of module `BI_AP_CL`, module `notBI_AP` and module `notLI_CL`. In this appendix, we use the flow charts (Figure B.1-B.5) to illustrate how these modules operate. The variables used in these modules are shown in Table B.1[1]. According to the output variables, we obtain the energy consumption and throughput of each client, the number of wake-ups, the number of PS-Poll/ACK/Data frames, the delay of each data frame at the AP and the number of collisions. Therefore, we evaluate the performance of PSM in an infrastructure wireless network.

0. The index of the current unit is initialized as $\tau = 1$. At the beginning of one unit, the main program firstly judges whether any device has finished receiving m_p data frames. If it is true, at least one element of **stopTimes**

[1] $\mathbf{0}(a, b)$ is a matrix with a rows b columns whose elements are all 0. $\mathbf{1}(a, b)$ is a matrix with a rows b columns whose elements are all 1. When $a = 1$ or $b = 1$, the matrix is reduced to a vector. We use the bold form to show a matrix or a vector.

will be larger than 0, and the simulation will stop. If the simulation goes on, the main program will judge whether it is time for the AP to send a beacon after updating τ according to **counterU**. If the current unit is a BI epoch (i.e., $\tau = \frac{\beta}{t_u}$), module `BI_AP_CL` will be called. If it is not a BI epoch, the AP will operate module `notBI_AP`, and the PSM-clients will operate module `notLI_CL` individually. After that, the main program simulates the next unit after increasing τ by one. The flow chart of the main program including module `BI_AP_CL` is shown as Figure B.1.

1. Module `BI_AP_CL`:

   The AP creates a beacon frame, sets **TIM** and initializes **existBeacon**, **justwake** and $simC$. $TIM(j) = 1$ if at least one data frame for s_j is buffered in the AP (i.e., $A(j, frames(j) + 1) \leq \tau$, where $frames(j)$ is the number of data frames which have been received by s_j); otherwise, $TIM(j) = 0$. If the channel is busy (i.e., $busy = 0$), the beacon will be sent successfully, and the AP will set $counterU(c + 1) = beacon\_units - 1$. Then the channel is recorded as busy (i.e., $busy = 1$), and the number of beacons (i.e., $beacons$) is increased by 1.

   The program also judges whether the current unit is a LI epoch for s_j, $\forall j$. If it is the LI epoch for s_j (i.e., $\tau = \frac{\beta\gamma_j}{t_u}$), $send\_dest(j)$ will be set as 1 to record that s_j is receiving the beacon frame. If $TIM(j) = 1$, then the number of awakened clients ($simC$) will increase one. If s_j is originally in the sleep mode (i.e., $powerMode(j) = 4$), the wake-up action will be recorded (i.e., $justwake(j) = 1$), and the number of wake-ups ($wakeups(j)$) will be increased by 1. Then, s_j updates **powerUnit** (i.e., $powerUnits(j, powerMode(j)) += counter - sum(powerUnits(j, :))$) and

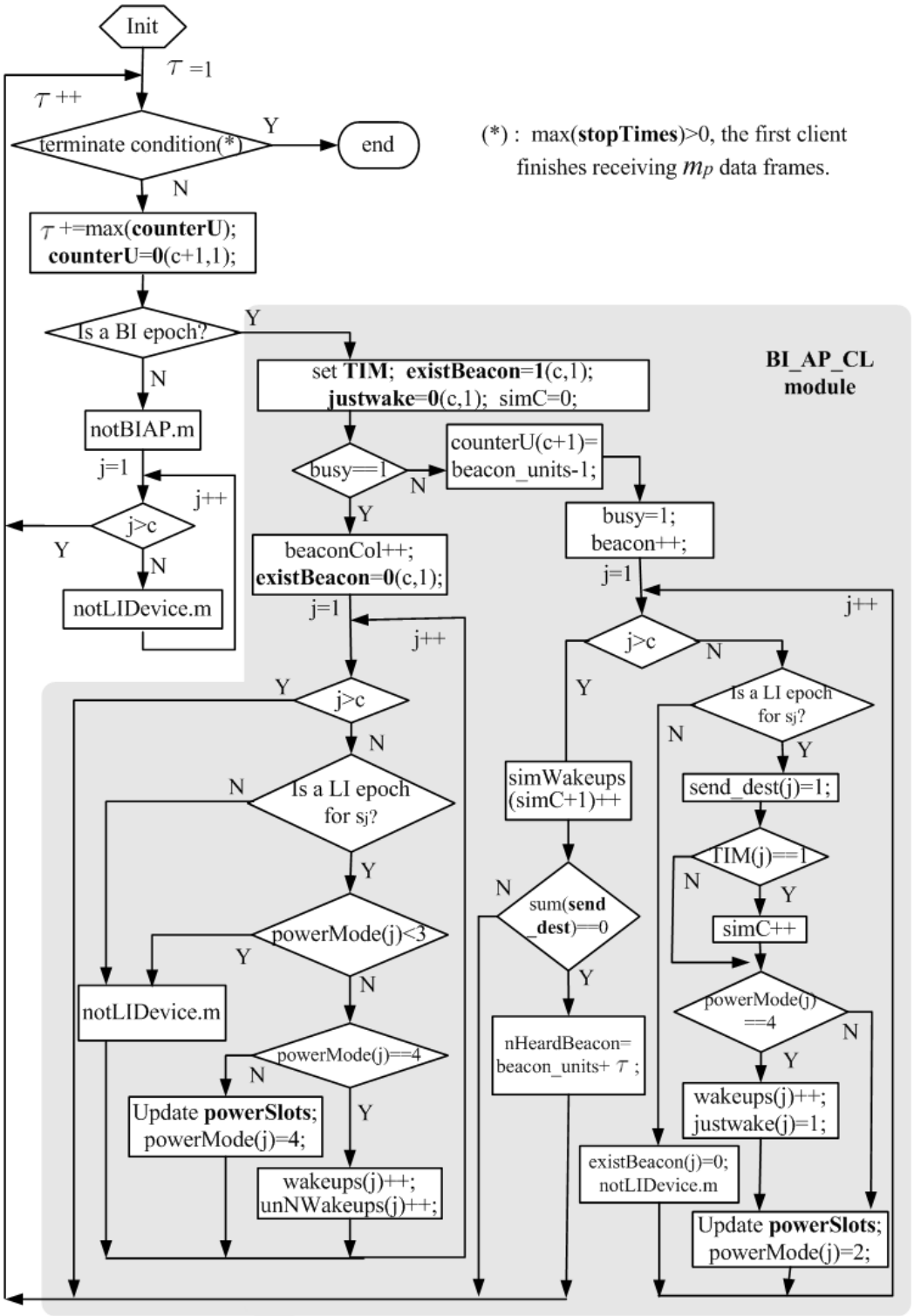


Figure B.1. The main program including module `BI_AP_CL`.

sets $powerMode(j) = 2$ to record that it goes to the receiving mode.[2] If it is not a LI epoch, s_j will unset $exstBeacon(j)$ and operate module `notLI_CL`. Finally, **simWakeups** records that $simC$ clients wake up to receive buffered frames in this BI where $simWakeups(simC + 1)$ is increased by 1. When the beacon frame is not received by any client (i.e., $sum(send\_dest) = 0$), the index of the unit when this beacon transmission will finish is recorded in $nheardBeacon$.

If the AP fails to send a beacon frame when $busy = 1$, the number of beacon collision ($beaconCol$) will be increased by 1, and all elements of **existBeacon** will be set as zero. If the current unit is not a LI epoch for s_j or s_j is receiving or sending any frame, s_j will operate module `notLI_CL`. If the current unit is a LI epoch for s_j, the idle s_j will update **powerUnits** and then go to the sleeping mode (i.e., $powerMode(j) = 4$). And the sleeping s_j will wake up without hearing any beacon. If then, the number of wake-ups ($wakeups(j)$) and the number of unnecessary wake-ups ($unNWakeups(j)$) will both increase one.

2. Module `notBI_AP`:

   When $nHeardBeacon > 0$, the AP is sending a beacon frame which is not heard by any client. If $\tau = nHeardBeacon$, this beacon transmission will finish, and the channel resource will be released by $busy = 0$. At the same time, $nHeardBeacon$ is reset as zero. If $nHeardBeacon = 0$ and $sum(send\_dest) > 0$, the module will not take any action, because the AP has been sending one beacon frame or one data frame. When $nHeardBeacon > 0$ and $sum(send\_dest) = 0$, the idle AP determines

[2] In a matrix $\mathbf{A}$ with a rows b columns, $A(:, b)$ ($A(a, :)$) shows a vector of the elements in the b^{th} (a^{th}) column (row). The function 'sum' obtains the sum of all elements in one vector.

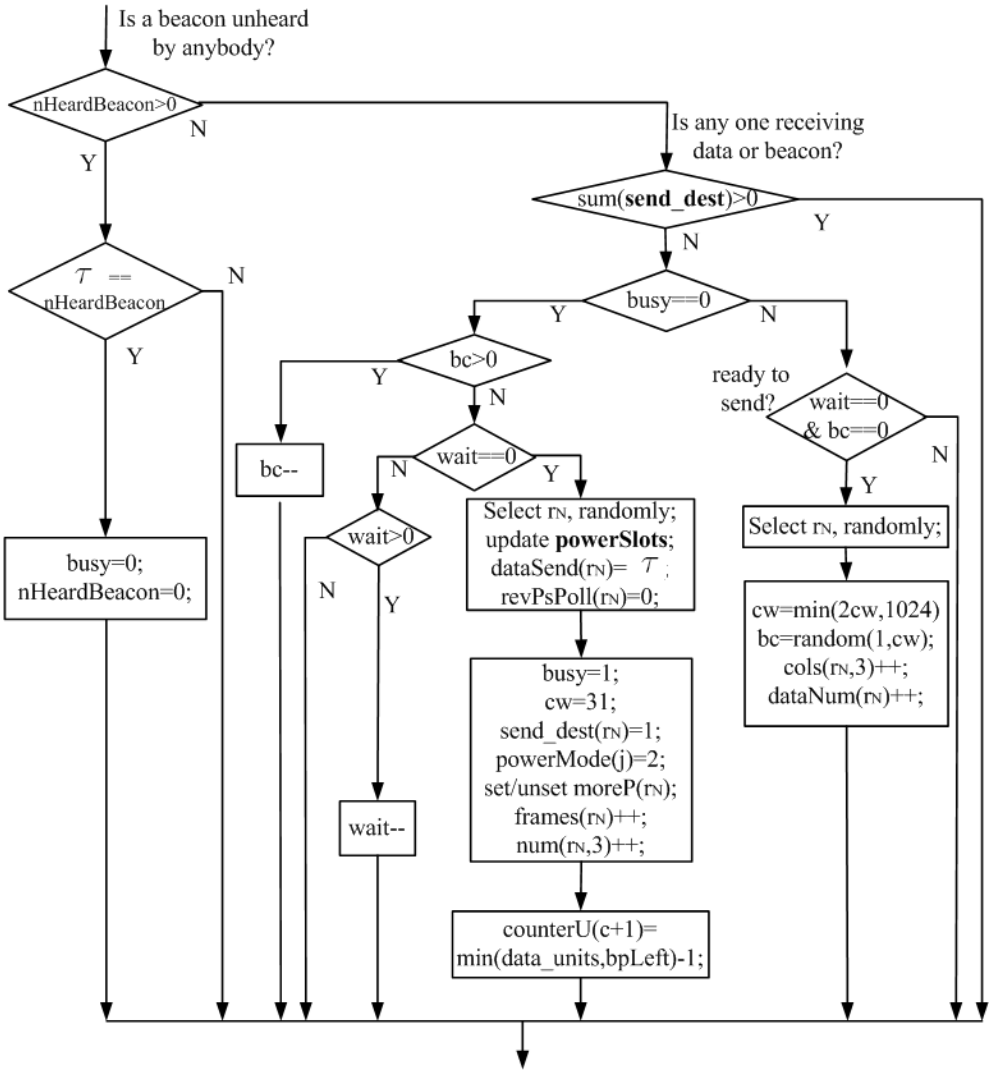


Figure B.2. The `notBI_AP` module.

whether to send one data frame according to DCF.

If the channel is idle (i.e., $busy = 0$), the backoff timer (bc) and the waiting timer of SIFS/DIFS ($wait$) will be checked. If $wait < 0$, the module will not take any action, since the AP has not any frame to send. If $wait > 0$ and $bc > 0$, bc will be decreased by 1. If $wait > 0$ and $bc = 0$, $wait$ will be decreased by 1. If $bc = 0$ and $wait = 0$, the AP will randomly select an awaken client whose index is denoted as r_N, according to the indicator vector of PS-Poll frames, **revPsPoll**. Then, the AP updates $revPsPoll(r_N) = 0$, $dataSend(r_N) = \tau$ and **powerUnits** to record that it begins to send one frame to s_{r_N}. Moreover, the busy channel is busy (i.e., $busy = 1$), s_{r_N} enters the receiving mode (i.e., $send\_dest(r_N) = 1$ and $powerMode(j) = 2$), and the AP's CW returns to the default value (i.e., 31). The AP also checks whether more frames for s_{r_N} are buffered. If $A(j, frames(j)+1) \leq \tau$, the More Data bit will be set by $moreP(r_N) = 1$. $frames(r_N)$ and the number of data frames sent to s_{r_N} (i.e., $num(r_N, 3)$) both increase one. And the AP sets $counterU(c+1) = data\_units - 1$.

If the channel is busy, the AP will do nothing except when $bc = 0$ and $wait = 0$. If the exception happens, the AP will fail to send a data frame to the selected awakened due to channel collision. The AP doubles its CW and resets bc according to the updated CW. Finally, $num(r_N, 3)$ and the number of data collision $cols(r_N, 3)$ both increase one.

3. module `notLI_CL`:

When s_j ($\forall j$) is not in a LI epoch, it operates module `notLI_CL` according to its current power mode. Four submodules are designed to deal with each client on different power modes. If s_j is in sleep mode, submodule

`notLI_CL_Sleep` will not take any action. Only when all clients are sleeping and the AP does not take any action (i.e., $counterU(1) = 0$), submodule `notLI_CL_Sleep` will set $counterU(j) = bpLeft - 1$ when $bpLeft > 1$, $j = 1, \ldots, c$. If s_j is active, the following three submodules: `notLI_CL_Tx`, `notLI_CL_Rx` and `notLI_CL_Idle` are performed accordingly.

3.1. Submodule `notLI_CL_Tx`:

If s_j is in the transmission mode (i.e., $powerMode(j) = 1$), submodule `notLI_CL_Tx` will take actions only when one frame transmission is finished. If $ack(j) = 1$, s_j will release the channel by $busy = 0$ and reset the indicator of ACK by $ack(j) = 0$ after finishing an transmission of ACK frame. If s_j has received m_p frames, it will update the stop time of the simulation by $stopTimes(j) = \tau$. Otherwise, s_j will check whether it has more buffered frames in the AP. If $moreP(j) = 1$, s_j will go to idle mode (i.e., $powerMode(j) = 3$), create a PS-Poll frame (i.e., $psPoll(j) = 1$) and prepare for sending it by setting $Waits(j)$. If s_j has not any buffered frame in the AP, it will go to sleep (i.e., $powerMode(j) = 4$) and unset $psPoll(j) = 0$, $Waits(j) = -1$. After that, s_j updates $powerUnits(j, 1)$.

If $psPoll(j) = 1$, s_j will release the channel by $busy = 0$ and reset the indicator of PS-Poll by $psPoll(j) = 0$ after an transmission of PS-Poll frame. s_j goes to idle mode (i.e., $powerMode(j) = 3$) after updating $Waits(j) = -1$ and $powerUnits(j, 1)$. It also records the PS-Poll arrival on the AP by $revPsPoll(j) = 1$ and $psPollArl(j) = counter$. When $wait = -1$, the AP will set it as SIFS (in units) to prepare for sending one data frame.

If s_j is sending a data frame when $ack(j) = 0$ and $psPoll(j) = 0$,

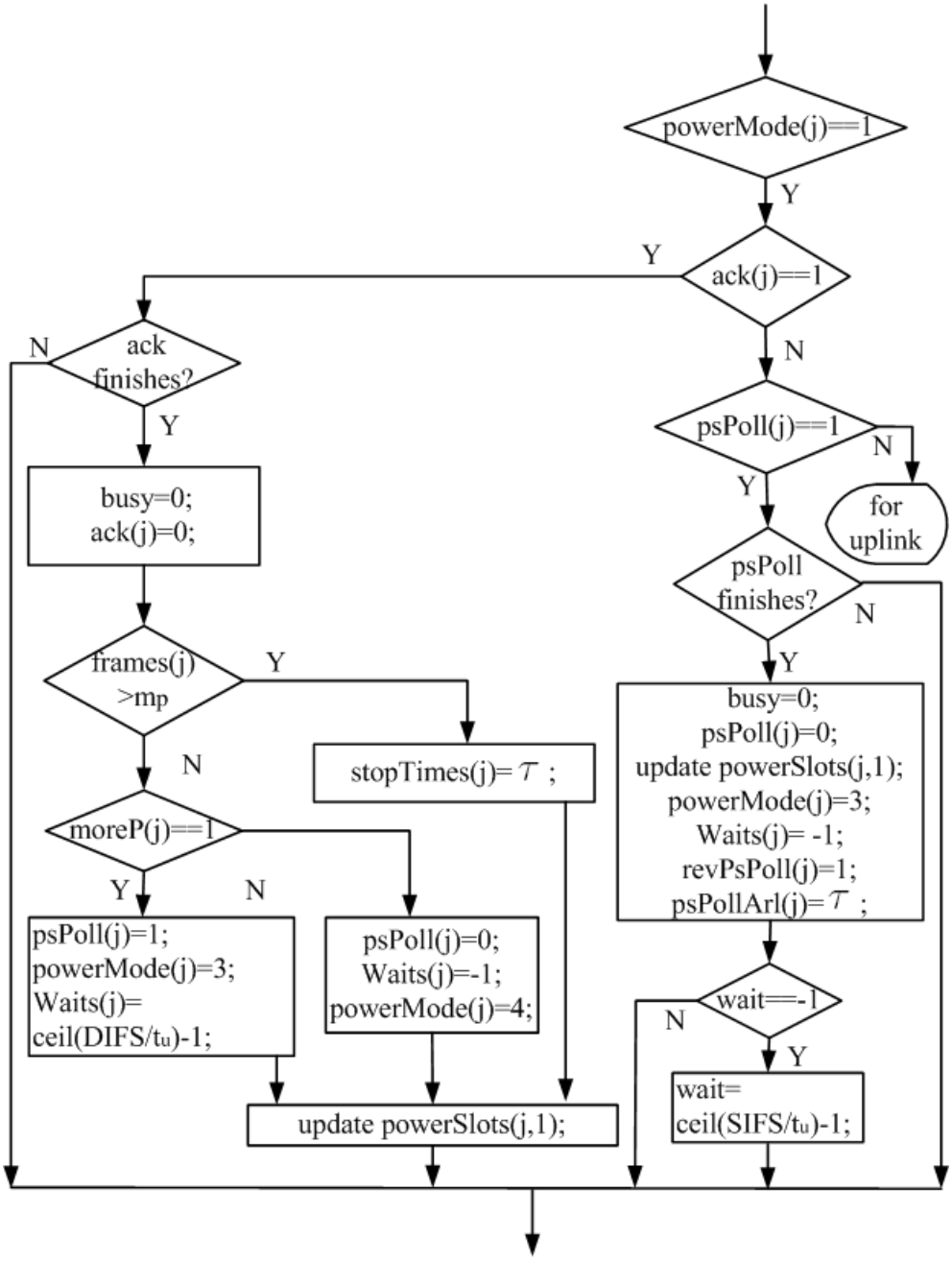


Figure B.3. The `notLI_CL_Tx` submodule.

the uplink module should be called. We will discuss this case in the future, since it is still under construction.

3.2. Submodule `notLI_CL_Rx` [3]:

If s_j is in the receiving mode (i.e., $powerMode(j)$=2), submodule `notLI_CL_Rx` will take actions only when one frame reception is finished. If $existBeacon(j) = 1$, s_j will reset $send\_dest(j) = 0$, $existBeacon(j) = 0$ and $wait = -1$ after receiving a beacon frame. If all awakened clients finish receiving the beacon frame, the channel will be released by $busy = 0$. Next, s_j will check whether it has buffered frames in the AP. If $TIM(j) = 1$, s_j will go to idle mode (i.e., $powerMode(j) = 3$), create a PS-Poll (i.e., $psPoll(j) = 1$) and prepare for sending it by setting $Waits(j) = DIFS$ and $BC(j) = CW(j)$. If there is not any buffered frame for s_j (i.e., $TIM(j) = 0$), s_j will go to sleep (i.e., $powerMode(j) = 4$) and set $psPoll(j) = 0$. In this case, s_j wakes up unnecessarily if $justwake(j) = 1$, and then the number of unnecessary wake-ups ($unNWakeups(j)$) increases one. Finally, s_j updates $powerUnits(j, 2)$.

If s_j finishes receiving a data frame, it will release the channel by $busy = 0$, reset $send\_dest(j) = 0$ and $wait = -1$. It also goes to the idle mode (i.e., $powerMode(j) = 3$), creates an ACK frame (i.e., $ack(j) = 1$) and prepares for sending it by setting $Waits(j) = SIFS$. Similarly, s_j updates $powerUnits(j, 2)$.

3.3. Submodule `notLI_CL_Idle`:

If s_j is in the idle mode (i.e., $powerMode(j)$=3), submodule `notLI_CL_Idle` will let it transmit a frame when the conditions are satisfied. If the

[3] This submodule will later implement the ACK reception in clients for uplink traffic.

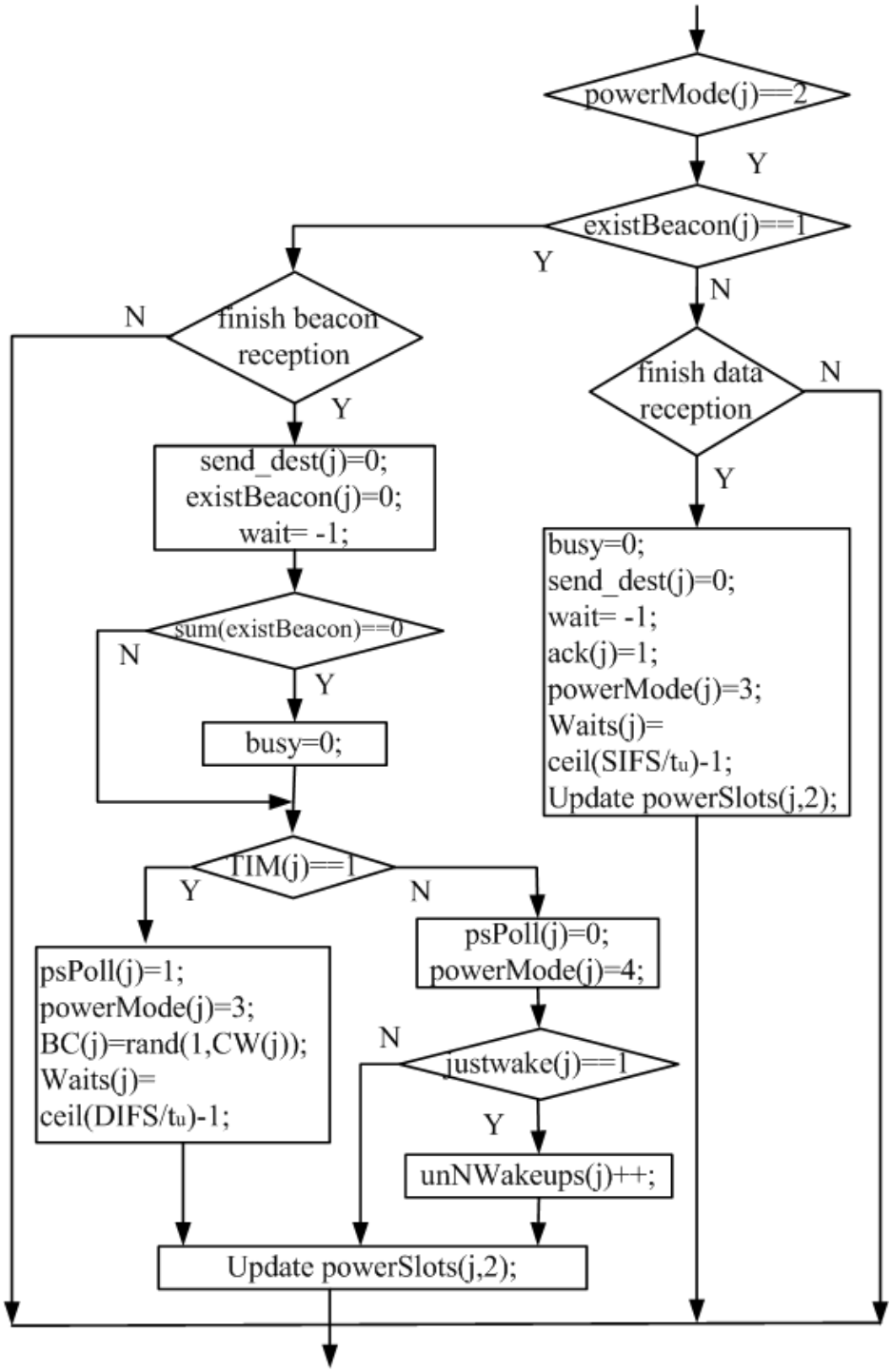


Figure B.4. The `notLI_CL_Rx` submodule.

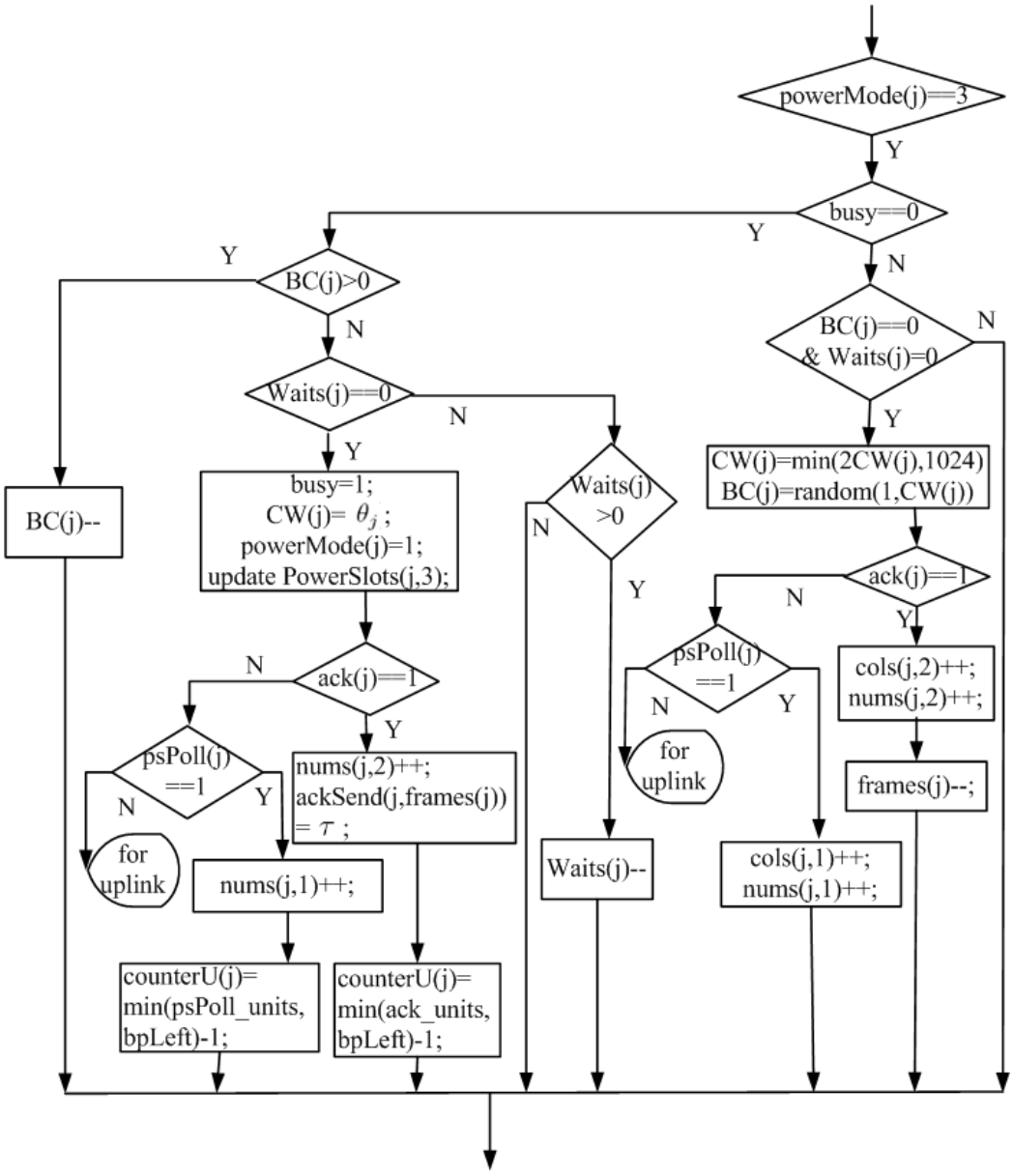


Figure B.5. The `notLI_CL_Idle` submodule.

channel is idle (i.e., $busy = 0$), the backoff timer ($BC(j)$) and the waiting timer ($Waits(j)$) of s_j will be checked. If $Waits(j) < 0$, the submodule will not take any action, since s_j has not any frame to send. If $BC(j) > 0$ and $Waits(j) > 0$, $BC(j)$ will decrease one. If $BC(j) = 0$ and $Waits(j) > 0$, $Waits(j)$ will decrease one. If $BC(j) = 0$ and $Waits(j) = 0$, s_j will send a frame and occupy the channel (i.e., $busy = 1$). After that, it resets the CW by the minimal value (i.e., $CW(j) = \theta_j$), updates $powerMode(j) = 1$ and $powerUnits(r_N, 3)$. If the frame is an ACK frame (i.e., $ack(j) = 1$), s_j will record the transmission time by $ackSend(j, frames(j)) = \tau$, increase its number of ACKs ($nums(j, 2)$) by 1 and set $counterU(j) = ack\_units - 1$. If the frame is a PS-Poll frame (i.e., $psPoll(j) = 1$), s_j will increase its number of PS-Polls ($nums(j, 1)$) by 1 and set $counterU(j) = psPoll\_units - 1$.

If the channel is busy, s_j is frozen except when $BC(j) = 0$ and $Waits(j) = 0$. If this exception happens, the client will fail to send the frame due to a channel collision. Then s_j doubles $CW(j)$ and resets $BC(j)$ according to the updated $CW(j)$. If the frame is an ACK frame (i.e., $ack(j) = 1$), s_j will increase its number of ACKs ($nums(j, 2)$) and its number of ACK collisions ($cols(j, 2)$) by 1 individually. Moreover, the $frames(j)^{th}$ frame will be retransmit after decreasing $frames(j)$ by 1. Similarly, if the frame is a PS-Poll frame (i.e., $psPoll(j) = 1$), s_j will increase its number of PS-Polls ($nums(j, 1)$) and its number of PS-Poll collisions ($cols(j, 1)$) by 1 individually.

When the uplink traffic exists, the idle s_j possibly prepares to send a data frame when $psPoll(j) = 0$ and $ack(j) = 0$. This case will be

implemented in the future.

Table B.1: A list of variables used in the simulator.

Variable	**Default**	**Definition**
Inputs variables		
c	2	Number of PSM-enabled clients, $j = 1, \ldots, c$
t_{sim}	30	Maximal arrival time of data frames
β	0.1s	Length of beacon interval
Γ	$\mathbf{1}(c, 1)$	Vector of listen intervals of c clients, γ_j for s_j
Θ	31*$\mathbf{1}(c, 1)$	Vector of minimal CWs of c clients, θ_j for s_j
$\mathbf{A}$	$\mathbf{0}(c, m_p)$	Matrix of arrival times of data frames for clients, $A(j, :)$ is for s_j, and m_p is the maximal number of data frames for one client during t_{sim}
data_units	57	Number of units for sending a data frame
beacon_units	31	Number of units for sending a beacon frame
psPoll_units	25	Number of units for sending a PS-Poll frame
ack_units	25	Number of units for sending an Ack frame
Outputs variables		
powerUnits	$\mathbf{0}(c, 4)$	Matrix of units when each client is in Tx/Rx/ Idle/Sleep modes, $powerUnits(j, :)$ is for s_j
dataSend	$\mathbf{0}(c, m_p)$	Matrix of time when AP sends data frames to each client, $dataSend(j, :)$ is for s_j
ackSend	$\mathbf{0}(c, m_p)$	Matrix of time when each client returns ACKs, $ackSend(j, :)$ is for s_j
frames	$\mathbf{0}(c, 1)$	Vector of frames received by c clients
beacons	0	Number of successful beacon frames

Variable	Default	Definition
beaconLIs	$\mathbf{0}(c, 1)$	Vector of beacons heard by c clients
wakeups	$\mathbf{0}(c, 1)$	Vector of wake-ups on c clients
unNWakeups	$\mathbf{0}(c, 1)$	Vector of unnecessary wake-ups on c clients
simWakeups	$\mathbf{0}(c+1, 1)$	Vector of the number of BIs when i clients receive data, $i = 0, 1, \ldots, c$
cols	$\mathbf{0}(c, 3)$	Matrix of PS-Poll/ACK/data collisions on each client, $cols(j, :)$ is for s_j
nums	$\mathbf{0}(c, 3)$	Matrix of the numbers of PS-Poll/ACK /data frames on each client, $nums(j, :)$ is for s_j
stopTimes	$\mathbf{0}(c, 1)$	Vector of time when AP receives the last ACK frame from clients
State variables		
τ	1	Index of the current unit
counterU	$\mathbf{0}(c+1, 1)$	Vector of updated units, $counterU(j)$ increased by s_j, and $counterU(c+1)$ is increased by AP
$bpLeft$	0	Number of units to the coming BI epoch
cw	31	Congestion windows of the AP
bc	0	Back-off timer of the AP
$wait$	-1	Waiting timer of the AP, $wait \leq DIFS$
$busy$	0	Indicator whether the channel is busy
$nHeardBeacon$	0	Unit when the transmission of a beacon frame which is not received by any client finishes
revPsPoll	$\mathbf{0}(c, 1)$	Vector of PS-Poll indicators of c clients,

Variable	Default	Definition
		$revPsPoll(j) = 1$ when a PS-Poll is from s_j
psPollArl	$\mathbf{0}(c, 1)$	Vector of arrival times of PS-Poll frames; AP receives a PS-Poll frame from s_j at $psPollArl(j)$
send_dest	$\mathbf{0}(c, 1)$	Vector of indicators whether c clients are receiving frame, $send\_dest(j) = 1$ when s_j is in Rx mode
moreP	$\mathbf{0}(1, c)$	Vector of More Data Bits of c clients, $moreP(j) = 1$ if frames are buffered for s_j
powerMode	$4 * \mathbf{1}(c, 1)$	Vector of power modes of c clients, $powerMode(j, i)$ $= 1/2/3/4$ when s_j is in Tx/Rx/Idle/Sleep mode
pMCounter	$\mathbf{0}(c, 1)$	Vector of the beginning unit of current power mode on c clients
ACK	$\mathbf{0}(c, 1)$	Vector of ACK indicators of c clients, $ack(j) = 1$ when an ACK from s_j exists
psPoll	$\mathbf{0}(c, 1)$	Vector of PS-Poll indicators of c clients, $psPoll(j) = 1$, when a PS-Poll from s_j exists
TIM	$\mathbf{0}(c, 1)$	Vector of TIM in a beacon, $TIM(j) = 1$ when AP buffers frames for s_j
$simC$	0	Number of awaken clients during a BI
CW	Γ	Congestion windows of c clients
BC	$\mathbf{0}(1, c)$	Vector of back-off timers of c clients
Waits	$-1 * \mathbf{1}(1, c)$	Vector of waiting timers of c clients
existBeacon	$\mathbf{0}(c, 1)$	Vector of beacon indicators of c clients, when s_j is receiving a beacon $existBeacon(j) = 1$
justwake	$\mathbf{0}(c, 1)$	Vector records wake-up actions of c clients

Bibliography

[1] The network simulator - ns-2. http://www.isi.edu/nsnam/ns/.

[2] Onoe rate control. http://madwifi.org/browser/trunk/athrate/onoe.

[3] Prism chipset data sheet an96141.1. http://www.intersil.com, 1997.

[4] Power consumption and energy efficiency comparisons of WLAN products. Technical report, Atheros Communications, Inc., 2003.

[5] A. Acampora and S. Krishnamurthy. A new adaptive MAC layer protocol for wireless ATM networks in harsh fading and interference environments. In *Proc. ICUPC*, 1997.

[6] S. Agarwal, S. Krishnamurthy, R. Katz, and S. Dao. Distributed power control in ad hoc wireless networks. In *Proc. PIMRC*, 2001.

[7] L. Alonso and R. Agusti. Automatic rate adaptation and energy-saving mechanisms based on cross-layer information for packet-switched data network. *IEEE Radio Communications*, pages 15–20, March 2004.

[8] E. Altman, A. Kumar, D. Kumar, and R. Venkatesh. Cooperative and non-cooperative control in IEEE 802.11 WLANs. Technical report, INRIA, March 2005.

[9] N. Amitay. Distributed switching and control with fast resource assignment/handoff for personal communication systems. *IEEE JSAC, SAC-11*, 1993.

[10] M. Anand, E. Nightingale, and J. Flinn. Self-tuning wireless network power management. In *Proc. ACM MOBICOM*, 2003.

[11] G. Anastasi, M. Conti, E. Gregori, and A. Passarella. Experimental analysis of an application-independent energy management scheme for Wi-Fi hotspots. In *Proc. ISCC 2004*, 2004.

[12] G. Anastasi, M. Conti, E. Gregori, and A. Passarella. A performance study of power-saving polices for Wi-Fi hotspots. *The International Journal of Computer and Telecommunications Networking*, 45(3):295–318, 2004.

[13] G. Anastasi, M. Conti, E. Gregori, and A. Passarella. 802.11 power-saving mode for mobile computing in Wi-Fi hotspots: Limitations, enhancements and open issues. *ACM/Springer Wireless Networks*, 2007.

[14] E. Ayanoglu, S. Paul, T. Laporta, and K. Sabnani. AIRMAIL: a link-layer protocol for wireless networks. *Wireless Networks*, 1(1):47–60, 1995.

[15] C. Bae and W. Stark. Energy and bandwidth efficiency in wireless networks. In *Proc. International Conference on Communications, Circuits and Systems*, 2006.

[16] V. Baiamonte and C. F. Chiasserini. Saving energy during channel contention in 802.11 WLANs. *Mobile Networks and Applications*, 11(2):287–296, 2006.

[17] A. Bakre and B.R. Badrinath. I-TCP: indirect TCP for mobile hosts. In

Proc. the 15th International Conference on Distributed Computing Systems, pages 136–143, 1995.

[18] H. Balakrishnan, S. Seshan, and R. Katz. Improving reliable transport and handoff performance in cellular wireless networks. *Wireless Networks*, 1(4), 1995.

[19] S. Bansal, R. Shorey, R. Gupta, and A. Misra. Energy efficiency and capacity for TCP traffic in multi-hop wireless networks. *Wireless Networks*, 12(1):5–12, Feb 2006.

[20] A. Belghith, A. Belghith, and M. Molnar. Enhancing psm efficiencies in infrastructure 802.11 networks. *International Journal of Computing and Information Sciences (IJCIS)*, 5(1):13–23, April 2007.

[21] R. Berry and R. Gallager. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, 48(5):1135–1149, 2002.

[22] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2003.

[23] I. Bettesh and S. Shamai. Optimal power and rate control for fading channels. In *Proc. IEEE Vehicular Technology Conference*, 2001.

[24] V. Bhargavan, A. Demers, S. Shenker, and L. Zhang. MACAW: Media access protocol for wireless LAN's. In *Proc. ACM SIGCOMM*, 1994.

[25] G. Bianchi and F. Borgonovo et al. C-PRMA: A centralized packet reservation multiple access for local wireless communications. *IEEE Transactions on Vehicular Technology*, 46(2), 1997.

[26] J. Bicket. Bit-rate selection in wireless networks. Master's thesis, MIT, 2005.

[27] E. Biyikoglu, B. Prabhakar, and A. Gamal. Energy-efficient packet transmission over a wireless link. *IEEE/ACM Transactions on Networking*, 10(4), 2002.

[28] G. Boggia, P. Camarda, L. Grieco, and S. Mascolo. Energy efficient feedback-based scheduler for delay guarantees in ieee 802.11e networks. *Computer Communications*, 29(13-14):2680–2692, Aug 2006.

[29] R. Bolla, F. Davoli, and C. Nobile. A RRA-ISA multiple access protocol with and without simple priority schemes for real-time and data traffic in wireless cellular systems. *Mobile Networks and Applications*, 2(1), 1997.

[30] I. Broustis, J. Eriksson, S. Krishnamurthy, and M. Faloutsos. Implications of power control in wireless networks: A quantitative study. In *Proc. Passive and Active Measurements Conference (PAM)*, 2007.

[31] K. Brown and S. Singh. M-TCP: TCP for mobile cellular networks. *Computer Communication Review*, 27(5), 1997.

[32] C-Guys, Inc. *SD-Link 11b Specifications*, 2004.

[33] MATLAB Central. 802.11b PHY matlab code. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3213\&objectType=FILE.

[34] MATLAB Central. IEEE 802.11a WLAN model. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3540&objectType=FILE.

[35] S. Chakraborty, Y. Dong, D. Yau, and J. Lui. On the effectiveness of movement prediction to reduce energy consumption in wireless communication. *IEEE Transactions on Mobile Computing*, 5(2), 2006.

[36] A. Chandra, V. Gummalla, and J. O. Limb. Wireless medium access control protocols. *IEEE Communications Surveys*, 2, 2000.

[37] A. Chandrakasan and R. Brodersen. Minimizing power consumption in digital CMOS circuits, 1995.

[38] A. Chandrakasan, S. Sheng, and R. Brodersen. Low-power CMOS digital design. *IEEE Journal of Solid-State Circuits*, 27(4):473–484, 1992.

[39] C. Chang, K. Chen, M. You, and J. Chang. Guaranteed quality-of-service wireless access to ATM networks. *IEEE Journal of Selected Areas in Communications*, 15(1):106–118, 1997.

[40] J. Chang and L. Tassiulas. Maximum lifetime routing in wireless sensor networks. *IEEE/ACM Transactions on Networking (TON)*, 12(4):609–619, 2004.

[41] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris. SPAN: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. In *Proc. the 7th ACM International Conference on Mobile Computing and Networking*, July 2001.

[42] D. Chen, J. Li, and J. Ma. Multiple access protocol for WLAN based on adaptive token passing with fairness guarantee. In *Proc. the 20th International Conference on Advanced Information Networking and Applications*, 2006.

[43] J. Chen, K. Sivalingam, and P. Agrawal. Performance comparison of battery power consumption in wireless multiple access protocols. *ACM/Baltzer Wireless Networks*, 5(6):445–460, 1999.

[44] J. Chen, K. Sivalingam, P. Agrawal, and R. Acharya. Scheduling multimedia services in a low-power MAC for wireless and mobile ATM networks. *IEEE Transactions on Multimedia*, 1(2), 1999.

[45] M. Chiang. Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control. *IEEE Journal on Selected Areas in Communications*, 23(1), 2005.

[46] M. Chiang and J. Bell. Balancing supply and demand of bandwidth in wireless cellular networks: Utility maximization over powers and rates. In *Proc. IEEE INFOCOM*, 2004.

[47] Z. Chou, C. Hsu, and S. Hsu. UPCF: a new point coordination function with qos and power management for multimedia over wireless lans. *IEEE/ACM Transactions on Networking (TON) archive*, 14(4), 2006.

[48] G. Choudhury and H. Takagi. Comments on "Exact Results for Nonsymmetric Token Ring Systems". *IEEE Trans. Communications*, 38(8):1125–1127, August 1990.

[49] Cisco Systems, Inc. *Data Sheet of CISCO AIRONET 802.11A/B/G WIRELESS CARDBUS ADAPTER*, 2004.

[50] Cisco Systems, Inc. *Cisco Aironet Wireless LAN Client Adapters Installation and Configuration Guide (Technical Specifications)*, Feb 2005.

[51] R. Cooper, S. Niu, and M. Srinivasan. Setups in polling models: does it

make sense to set up if no work is waiting? *Journal of Applied Probability*, 36:585–592, 1999.

[52] D. De Couto, D. Aguayo, B. Chambers, and R. Morris. Performance of multihop wireless networks: Shortest path is not enough. In *Proceedings of the First Workshop on Hot Topics in Networks (HotNets-I)*. Proc. ACM SIGCOMM, 2002.

[53] R. Dittmann and F. Hubner. Discrete-time analysis of a cyclic service system with gated limited service. Technical report, University of Wurzburg, 1993.

[54] Y. Dong and D. Yau. Adaptive sleep scheduling for energy-efficient movement-predicted wireless communication. In *Proc. IEEE International Conference on Network Protocols (ICNP)*, 2005.

[55] S. Doshi, S. Bhandare, and T. Brown. An on-demand minimum energy routing protocol for a wireless ad hoc network. *Mobile Computing and Communications Review*, 6(3), 2002.

[56] F. Douglis, P. Krishnan, and B. Marsh. Thwarting the power-hungry disk. In *Proc. Winter USENIX Conference*, 1994.

[57] J. Ebert, S. Aier, G. Kofahl, A. Becker, B. Burns, and A. Wolisz. Measurement and simulation of the energy consumption of an WLAN interface. Technical report, Technical University Berlin, Telecommunication Networks Group, June 2002.

[58] L. Feeney and M. Nilsson. Investigating the energy consumption of a wireless network interface in an ad hoc networking environment. In *Proc. IEEE INFOCOM*, 2001.

[59] M. Ferguson and Y. Aminetzah. Exact results for nonsymmetric token ring systems. *IEEE Trans. Communications*, 33(3):223–231, March 1985.

[60] L. Foulds. *Optimization Techniques*. Springer-Verlag New York Inc., 1981.

[61] C. Fricker and R. Jaibi. Monotonicity and stability of periodic polling models. *Queueing Systems*, 15:211–38, 1994.

[62] A. Fukuda, K. Mukumoto, and W. Gang. Slotted idle signal multiple access scheme for two-way centralized wireless communication networks. *IEEE Transactions on Vehicular Technology*, 43(2), 1994.

[63] C. Fullmer and J. Carcia. Floor acquisition multiple access (FAMA) for packet-radio networks. In *Proc. ACM SIGCOMM*, 1995.

[64] A. Gamal, C. Nair, B. Prabhakar, E. Biyikoglu, and S. Zahedi. Energy-efficient scheduling of packet transmissions over wireless networks. In *Proc. IEEE INFOCOM*, 2002.

[65] M. Gast. *802.11 Wireless Networks: The Definitive Guide*. O'Reilly Media, Inc., 2 edition, 2005.

[66] J. Geier. 802.11 medium access methods. http://www.wi-fiplanet.com/tutorials/article.php/1548381, November 2002.

[67] S. Ghez, S. Verdu, and S. Schwartz. Stability properties of slotted Aloha with multipacket reception capability. *IEEE Transaction on automatic control*, 33(7), July 1988.

[68] J. Gomez, A. Campbell, M. Naghshineh, and C. Bisdikian. Conserving transmission power in wireless ad hoc networks. In *Proc. IEEE Conference on Network Protocols (ICNP)*, November 2001.

[69] K. Govil, E. Chan, and H. Wasserman. Comparing algorithms for dynamic speed-setting of a low-power CPU. In *Proc. ACM MOBICOM*, pages 13–25, 1995.

[70] A. Grilo and M. Nunes. Link-adaptation and transmit power control for unicast and multicast in IEEE 802.11a/h/e WLANs. In *Proc. IEEE Conference on Local Computer Networks (LCN)*, 2003.

[71] Z. Haas and J. Deng. Dual busy tone multiple access (DBTMA)a multiple access control scheme for ad hoc networks. *IEEE Transactions on Communications*, 50(6), July 2002.

[72] R. Haines and A. Aghvami. Indoor radio environment considerations in selecting a media access control protocol for wideband radio data communications. In *Proc. IEEE ICC*, 1994.

[73] J. Havinga and J. Smit. Energy-efficient wireless networking for multimedia applications. *Wireless Communications and Mobile Computing*, 1, 2001.

[74] P. Havinga. Energy efficiency of error correction on wireless systems. In *Proc. IEEE Wireless Communications and Networking Conference*, 1999.

[75] P. Havinga and G. Smit. E2MaC: an energy efficient MAC protocol for multimedia traffic. Technical report, University of Twente, 1998.

[76] P. Havinga and G. Smit. Qos scheduling for energy-efficient wireless communication. In *Proc. ITCC*, 2001.

[77] P. Havinga, G. Smit, and M. Bos. Energy efficient adaptive wireless network design. In *Proc. IEEE ISCC*, 2000.

[78] Y. He, R. Yuan, X. Ma, J. Li, and C. Wang. Scheduled PSM for minimizing energy in wireless lans. In *Proc. IEEE ICNP*, 2007.

[79] C. Hein. CSIM-parallel process and diagrams simulator. Technical report, Lockheed-Martin ATL, 2004.

[80] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proc. the 33rd Hawaii International Conference on System Sciences*, 2000.

[81] G. Holland and N. Vaidya. Analysis of TCP performance over mobile ad hoc networks. *Wireless Networks*, 8:275–288, 2002.

[82] G. Holland, N. Vaidya, and P. Bahl. A rate-adaptive mac protocol for multi-hop wireless networks. In *Proc. ACM MOBICOM*, 2001.

[83] T. Holliday, A. Goldsmith, and P. Glynn. Optimal power control and source-channel coding for delay constrained traffic over wireless channels. In *Proc. IEEE ICC*.

[84] F. Hu and N.K. Sharma. Enhancing wireless internet performance. *IEEE Communications Surveys and Tutorials*.

[85] X. Huang and I.Rubin. Bit-per-joule performance of power saving ad hoc networks with a mobile backbone under distance aware routing. In *Proc. IEEE GLOBECOM 2006*, 2006.

[86] IEEE-SA Standards Board. *IEEE Std 802.11b: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*, September 1999.

[87] IEEE-SA Standards Board. *IEEE Std 802.11i: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications–MAC Security Enhancements*, June 2004.

[88] IEEE-SA Standards Board. *IEEE Std 802.11e: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications–MAC Quality of Service Enhancements*, September 2005.

[89] Siemens Energy & Automation Inc. Node management: Dcf, pcf and ipcf. http://www.sea.siemens.com/scalance/tech.html.

[90] S. Jayashree, B. Manoj, and C. Murthy. Next step in MAC evolution: Battery awareness? In *Proc. IEEE Globecom*, 2004.

[91] Y. Jeong, J. Park, J. Ma, and D. Kim. An enhanced power save mode for ieee 802.11 station in ad hoc networks. In *Proc. IFIP International Federation for Information*, page 414C420, 2004.

[92] M. Johnsson. HiperLAN/2 - the broadband radio transmission technology operating in the 5 ghz frequency band. http://easy.intranet.gr/paper_20.pdf.

[93] C. Jones, K. Sivalingam, P. Agrawal, and J. Chen. A survey of energy efficient network protocols for wireless networks. *Wireless Networks*, 7(4), July 2001.

[94] E. Jung and N. H. Vaidya. An energy efficient MAC protocol for wireless LANs. In *Proc. IEEE INFOCOM*, 2002.

[95] E. Jung and N. H. Vaidya. A power control MAC protocol for Ad Hoc networks. In *Proc. ACM MOBICOM*, 2002.

[96] A. Kamerman and L. Monteban. WaveLAN II: A high-performance wireless lan for the unlicensed band. *Bell Labs Technical Journal*, 1997.

[97] P. Karn. MACA- a new channel access method for packet radio. In *Proc. the ARRL/CRRL Amateur Radio 9th Computer Networking Conference*, 1990.

[98] M. Karol, Z. Liu, and K. Eng. An efficient demand-assignment multiple access protocol for wireless packet (ATM) networks. *Wireless Networks*, 1(3), Sept 1995.

[99] J. Kim, S. Kim, S. Choi, and D. Qiao. CARA: Collision-aware rate adaptation for IEEE 802.11 WLANs. In *Proc. IEEE INFOCOM*, 2006.

[100] Y. Kim, H. Jung, H. Lee, and K. Cho. Mac implementation for IEEE 802.11 wireless LAN. In *Proc. ATM (ICATM 2001) and High Speed Intelligent Internet Symposium*, 2001.

[101] Y. Kim, J. Yu, and S. Choi. SP-TPC: A self-protective energy efficient communication strategy for IEEE 802.11 WLANs. In *Proc. IEEE VTC*, 2004.

[102] L. Kleinrock. *Queueing Systems, Vol.2: Computer Applications*, volume 2. Wiley, 1976.

[103] L. Kleinrock and F. Tobagi. Packet switching in radio channels: Part icarrier sense multiple access modes and their throughput-delay characteristics. *IEEE Transactions on Communications*, Dec 1975.

[104] G. Koole and P. Nain. An explicit solution for the value function of a priority queue. *Queueing Systems*, 47:251–285, 2004.

[105] R. Krashinsky and H. Balakrishnan. Minimizing energy for wireless web access with bounded slowdown. *Wireless Networks*, 11:135–148, 2005.

[106] S. Kumar, V. S. Raghavan, and J. Deng. Medium access control protocols for ad hoc wireless networks: A survey. *Ad-Hoc Networks Journal*, 4(3):326–358, May 2006.

[107] S. Kurkowski, T. Camp, and M. Colagrosso. MANET simulation studies: the incredibles. *Mobile Computing and Communications Review*, 9(4), 2005.

[108] S. Kwon, J. Lee, and D. Cho. Power management scheme considering priority in wireless lan. In *Proc. IEEE Vehicular Technology Conference*, 2005.

[109] UCLA Parallel Computing Laboratory. Global mobile information systems simulation library. http://pcl.cs.ucla.edu/projects/glomosim/.

[110] M. Lacage, M. Manshaei, and T. Turletti. Rate adaptation: A practical approach. In *Proc. ACM MSWiM*, 2004.

[111] J. Lee, C. Rosenberg, and E. Chong. Energy efficient schedulers in wireless networks: design and optimization. *Mobile Networks and Applications*, 11(3), June 2006.

[112] J. Lee, C. Rosenberg, and K. Chong. An opportunistic power-saving mode and scheduler design for wireless local area networks. In *Proc. WCNC 2006*, 2006.

[113] H. Lei and A. Nilsson. Queuing analysis of power management in the IEEE 802.11 based wireless lans. *IEEE Transactions on Wireless Communications*, 6(4), 2007.

[114] P. Lettieri, C. Schurgers, and M. Srivastava. Adaptive link layer strategies for energy efficient wireless networking. *Wireless Networks*, 5(5), 1999.

[115] S. Lim, C. Yu, and C. Das. RCAST: A randomized communication scheme for improving energy efficiency in manets. In *Proc. the 25th IEEE ICDCS*, pages 123–132, June 2005.

[116] H. Lin, S. Huang, and R. Jan. A power-saving scheduling for infrastructure-mode 802.11 wireless LANs. *Computer Communications*, 29:3483–3492, 2006.

[117] J. Liu and S. Singh. ATCP: TCP for mobile ad hoc networks. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 19(7), 2001.

[118] Lucent. *IEEE 802.11 WaveLAN PC Card: User's Guide*, p.a1 edition.

[119] C. Margi. *Energy Consumption Trade-offs in Power Constrained Networks*. PhD thesis, University of California Santa Cruz, June 2006.

[120] B. Marsh, F. Douglis, and P. Krishnan. Flash memory file caching for mobile computers. In *Proc. the 27th Hawaii Conference on Systems Science*, 1994.

[121] S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, and R. Wang. TCP Westwood: Bandwidth estimation for enhanced transport over wireless links. In *Proc. ACM MOBICOM*, pages 16–21, July 2001.

[122] F. Meshkati, H. Poor, and S. Schwartz. Energy-efficient resource allocation in wireless networks: An overview of game-theoretic approaches. *IEEE SIGNAL PROCESSING MAGAZINE*, May 2007.

[123] R. Min and A. Chandrakasan. A framework for energy-scalable communication in high-density wireless networks. In *Proc. International symposium on Low power electronics and design (ISLPED)*, pages 36–41, 2002.

[124] R. Min and A. Chandrakasan. Top five myths about the energy consumption of wireless communication. *Mobile Computing and Communications Review*, 1(2), 2002.

[125] A. Misra and S. Banerjee. MRPC: Maximizing network lifetime for reliable routing in wireless environments. In *Proc. IEEE Wireless Communications and Networking Conference*, 2002.

[126] J. Monks, V. Bharghavan, and W. W. Hwu. A power controlled multiple access protocol for wireless packet networks. In *Proc. IEEE INFOCOM*, pages 219–228, 2001.

[127] S. Nath, Z. Anderson, and S. Seshan. Choosing beacon periods to improve response times for wireless HTTP clients. In *Proc. the ACM International Workshop on Mobility Management and Wireless Access (MobiWac)*, 2004.

[128] V. Naware and L. Tong. Stability of queues in slotted ALOHA with multiple antennas. In *Proc. The Allerton Conference*, 2002.

[129] S. Nedevschi, L. Popa, and G. Iannaccone. Reducing networking energy consumption via sleeping and rate-adaptation. In *Proc. the 5th USENIX Symposium on Networked Systems Design and Implementation*, 2008.

[130] P. Nicopolitidis, G. Papadimitriou, and A. Pomportsis. Learning automata-based polling protocols for wireless LANs. *IEEE Transactions on Communications*, 51(3), 2003.

[131] P. Nuggehalli, V. Srinivasan, and R. Rao. Delay constrained energy efficient transmission strategies for wireless devices. In *Proc. IEEE INFOCOM*, 2002.

[132] P. Nuggehalli, V. Srinivasan, and R. Rao. Energy efficient transmission scheduling for delay constrained wireless networks. *IEEE Transactions on Wireless Communications*, 5(3), 2006.

[133] Inc. OPNET Technologies. Opnet modeler documentation and support. www.opnet.com.

[134] S. Pack and Y. Choi. An adaptive power saving mechanism in ieee 802.11 wireless IP networks. *JOURNAL OF COMMUNICATIONS AND NETWORKS*, 7(2), 2005.

[135] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos. Content-aware resource allocation and packet scheduling for video transmission over wireless networks. *IEEE Journal on Selected Areas in Communications*, 25(4):106–118, 2007.

[136] P. Pant and T. Castelli. Simulation of a wireless network using the 802.11 mac protocol. http://nislab.bu.edu/sc546/sc546Fall2002/wireless/FinalReport.pdf, 2002.

[137] Y. Park and G. Hwang. Performance modelling and analysis of the sleep-mode in IEEE802.16e WMAN. In *Proc. IEEE Vehicular Technology Conference*, 2007.

[138] W. Pattara-atikom and P. Krishnamurthy. Distributed mechanisms for quality of service in wireless LANs. *IEEE Wireless Communications Magazine*, June 2003.

[139] A. Pinheiro and J. DeMarca. Fair deterministic packet access protocol: F-RAMA. *Electronic Letters*, 32(25), Dec 1996.

[140] J. Postel. Transmission Control Protocol. *RFC 793*, September 1981.

[141] The VINT Project. *The ns Manual*. UC Berkeley, LBL, USC/ISI, and Xerox PARC, July 2007.

[142] Proxim Wireless Corporation. *Data sheet of ORiNOCO 11a/b/g Combo-Card*, 2006.

[143] Proxim Wireless Corporation. *Data Sheet of ORiNOCO 11a/b/g PCI Card*, 2006.

[144] D. Qiao, S. Choi, A. Jain, and K. Shin. MiSer: An optimal low-energy transmission strategy for IEEE 802.11a/h. In *Proc. ACM MOBICOM*, 2003.

[145] D. Qiao, S. Choi, and K. Shin. Interference analysis and transmit power control in IEEE 802.11a/h wireless lans. *IEEE/ACM Transactions on Networking*, 15(5):1007–1020, 2007.

[146] D. Qiao, S. Choi, A. Soomro, and K. Shin. Energy-efficient PCF operation of IEEE 802.11a wireless LAN. In *Proc. IEEE INFOCOM*, 2002.

[147] D. Qiao and K. Shin. Smart power-saving mode for IEEE 802.11 wireless lans. In *Proc. IEEE INFOCOM*, 2005.

[148] B. Radunovic and J. Boudec. Rate performance objectives of multi-hop wireless networks. In *Proc. IEEE INFOCOM*.

[149] V. Raghunathan, S. Ganeriwal, and M. Srivastava. Energy efficient wireless packet scheduling and fair queuing. *ACM Transactions in Embedded Computing Systems (TECS)*, 3(1):3–23, February 2004.

[150] N. Riga and A. Medina et al. Transport services for energy constrained environments. In *Proc. ACM SIGCOMM'05, Work-in-progress Session.*, 2005.

[151] V. Rodoplu and H. Meng. Bit-per-joule capacity of energy-limited wireless networks. *IEEE Transactions on Wireless Communications*, 6(3):857–865, 2007.

[152] A. Sadek, K. Liu, and A. Ephremides. Cooperative multiple access for wireless networks: Protocols design and stability analysis. In *Proc. the 40th Annual Conference on Information Sciences and Systems*, 2006.

[153] Y. Sankarasubramaniam, I. Akyildiz, and S. Mclaughlin. Energy efficiency based packet size optimization in wireless sensor networks. In *Proc. the 1st IEEE Intl. Workshop on Sensor Network Protocols and Applications*, 2003.

[154] C. Schurgers, O. Aberthorne, and M. Srivastava. Modulation scaling for energy aware communication systems. In *Proc. International Symposium on Low Power Electronics and Design*, pages 96 – 99, 2001.

[155] C. Schurgers, V. Raghunathan, and M. Srivastava. Modulation scaling for real-time energy aware packet scheduling. In *Proc. Global Communications Conference*, pages 3653–3657, 2001.

[156] E. Shih, P. Bahl, and M. Sinclair. Wake on wireless: An event driven energy saving strategy for battery operated devices. In *Proc. MOBICOM*, 2002.

[157] T. Simunic, L. Benini, P. Glynn, and G.D. Micheli. Dynamic power management for portable systems. In *Proc. ACM MOBICOM*, 2000.

[158] S. Singh and C. Raghavendra. PAMAS power aware multi-access protocol with signalling for ad hoc networks. *ACM Computer Communication Review*, 28(3):5–26, 1998.

[159] OMNeT++ Community Site. Omnet++ discrete event simulation system. http://www.omnetpp.org/.

[160] Scalable Network Technologies (SNT). Qualnet software and documentation. http://www.scalable-networks.com/.

[161] P. Soni and A. Chockalingam. Performance analysis of UDP with energy efficient link layer on Markov fading channels. *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, 1(4), 2002.

[162] A. Srinivas and E. Modiano. Minimum energy disjoint path routing in wireless ad-hoc networks. In *Proc. ACM MOBICOM*, 2003.

[163] M. Stemm and R. Katz. Measuring and reducing energy consumption of network interfaces in hand-held devices. *IEICE Transactions on Communications*, E80-B(8):1125–31, 1997.

[164] K. Sundaresan, V. Anantharaman, H. Hsieh, and R. Sivakumar. ATP: A reliable transport protocol for ad-hoc networks. In *Proc. ACM MobiHoc*, 2003.

[165] W. Szpankowski. Stability conditions for some distributed systems: Buffered random access systems. *Advances in Applied Probability*, 26, 1994.

[166] H. Takagi. Queueing analysis of polling models: an update. *Stochastic Analysis of Computer and Communication Systems*, pages 267–318, 1990.

[167] E. Tan, L. Guo, S. Chen, and X. Zhang. PSM-throttling: Minimizing energy consumption for bulk data communications in WLANs. In *Proc. IEEE ICNP*, 2007.

[168] A. Tarello, J. Sun, M. Zafer, and E. Modiano. Minimum energy transmission scheduling subject to deadline constraints. In *Proc. IEEE WIOPT*, 2005.

[169] Inc. The MathWorks. Matlab software and documentation. http://www.mathworks.com/.

[170] Y. Tseng, C. Hsu, and T. Hsieh. Power-saving protocols for IEEE 802.11-based multi-hop ad hoc networks. In *Proc. IEEE INFOCOM*, 2002.

[171] H. Tyan. *Design, realization and evaluation of a component-based compositional software architecture for network simulation.* PhD thesis, Graduate School of The Ohio State University, 2002.

[172] M. Vuuren and E. Winands. Iterative approximation of k-limited polling systems. *Queueing Systems: Theory and Applications*, 55(3):161–178, 2007.

[173] P. Wan, G. Calinescu, X. Li, and O. Frieder. Minimum energy broadcast routing in static ad hoc wireless networks. *ACM Wireless Networking (WINET)*, 8(6):607–617, 2002 2002.

[174] M. Wiczanowski, Y. Chen, S. Stanczak, and H. Boche. Optimal energy control in energy-constrained wireless networks with random arrivals under stability constraints. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, 2005.

[175] J. Wieselthier, G. Nguyen, and A. Ephremides. On the construction of energy-efficient broadcast and multicast trees in wireless networks. In *Proc. IEEE INFOCOM*, 2000.

[176] W. Wong and D. Goodman. Integrated data and speech transmission using packet reservation multiple access. In *Proc. IEEE ICC*, 1993.

[177] C. Wu and V. Li. Receiver-initiated busy-tone multiple access in packet radio networks. In *Proc. ACM SIGCOMM*, 1987.

[178] G. Wu and K. Taira et al. An R-ISMA integrated voice/data wireless information system with different packet generation rates. In *Proc. IEEE ICC*, 1996.

[179] G. Wu, K. Mukumoto, and A. Fukuda. An integrated voice and data transmission system with idle signal multiple access-dynamic analysis. *IEEE Transactions on Communications*, 76(11), Nov 1993.

[180] H. Yan, R. Krishnan, S. Watterson, and D. Lowenthal. Client-centered energy savings for concurrent HTTP connections. In *Proc. ACM NOSSDAV*, 2004.

[181] F. Zhang and S. Chanson. Throughput and value maximization in wireless packet scheduling under energy and time constraints. In *Proc. IEEE RTSS*, 2003.

[182] F. Zhang and S. Chanson. Blocking-aware processor voltage scheduling for real-time tasks. *ACM Transactions in Embedded Computing Systems (TECS)*, 3(2):307–335, 2004.

[183] Z. Zhang and A. Acampora. Performance of a modified polling strategy for

broadband wireless LANs in a harsh fading environment. In *Proc. IEEE GLOBECOM*, 1991.

[184] M. Zorzi and R. Rao. Error control and energy consumption in communications for nomadic computing. *IEEE Transactions on Computers*, 46(3):279–289, 1997.