國 立 交 通 大 學

電信工程研究所

博 士 論 文

行動寬頻無線網路之適應性通訊與省電機制

Adaptive Communication and Power-saving Mechanisms
in Mobile Broadband Wireless Networks

研 究 生：徐仲賢

指導教授：方凱田　博士

中 華 民 國 九 十 八 年 十 一 月

行動寬頻無線網路之適應性通訊與省電機制

Adaptive Communication and Power-saving Mechanisms in Mobile Broadband Wireless Networks

研 究 生：徐仲賢　　　　　　　　Student: Chung-Hsien Hsu

指導教授：方凱田 博士　　　　　　Advisor: Dr. Kai-Ten Feng

國立交通大學

電信工程研究所

博士論文

A Dissertation
Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in
Communication Engineering
Hsinchu, Taiwan

2009 年 11 月

行動寬頻無線網路之適應性通訊與省電機制

研究生：徐仲賢　　　　　　　　　　　　指導教授：方凱田 博士

國立交通大學　電信工程研究所

摘　　要

近幾年來，隨著移動式網際網路與無線多媒體應用的需求增長，刺激著寬頻無線存取技術的研究發展。在各式各樣的寬頻無線存取技術發展之際，以 IEEE 802.16 標準為基礎的移動式全球互通微波存取（Mobile WiMAX）是目前已經開始提供服務的領先技術。此篇論文論及一些在以 IEEE 802.16 為基礎的移動式寬頻無線網路之問題。更具體地說，四個議題在此篇論文中被探討：應用於 IEEE 802.16 點對多點（Point-to-multipoint, PMP）網路使用者（Subscriber Station, SS）間的適應性直接通訊方法；針對每對欲採用直接通訊方式的移動式使用者（Mobile Station, MS），以預測移動與干擾為基礎的排程演算法；在 IEEE 802.16m 睡眠模式運作中的傾聽視窗（Listening Window）調整機制；以及在該模式中以統計方式控制睡眠視窗（Sleep Window）的方法。

在第一個議題中，一個有彈性且免競爭的通訊方式被提出以提供 IEEE 802.16 點對多點網路使用者間進行直接通訊。基地台（Base Station, BS）控制和安排特定的時段給兩個正在進行封包傳送的使用者。在所提出的方法中，封包傳送的運作方式將依據基地台和使用者間的頻道狀態（Channel Condition）主動地在直接與間接通訊間切換。這樣的機制將會增加頻寬的利用率進而提升使用者生產量（User Throughput）。

以上述所提及之直接通訊為基礎，對於每對預期採用直接通訊的移動式使用者之排程演算法在第二個議題中被探討。藉由分析與估測每對欲通訊的移動式使用者的干擾範圍（Interference Region）和可實行範圍（Feasible Region），提出的演算法將適當地把該對移動式使用者進行直接通訊的時期安排在下載子框架（Downlink Subframe）或上傳子框架（Uplink Subframe），亦或者將其通訊方式轉換成原本的間接通訊方式，即經由基地台傳送。此外，合併該排程演算法與提出的預測使用者移動之機制，更新移動式使用者位置資訊所帶來的控制花費（Control Overhead）將被有效率地降低。

另一方面，為了改進移動式使用者在睡眠模式中具有即時性傳輸資料量（Real-time Traffic）時的能量保存效能，一個有彈性的傾聽視窗調整機制將在第三個探討議題中被提出。不同於 IEEE 802.16e-2005 標準中所規範的以視窗為基礎之運作方式，在提出的機制中將以週期為運作之基礎。該機制根據儲存的封包數目、欲重新傳送的封包數目以及可容忍的延遲限

制（Delay Constraint），動態地調整每個睡眠周期（Sleep Cycle）中傾聽視窗與睡眠視窗的比例，以達到節省能量之目的。值得一提的是作者於 IEEE 802.16m 國際標準制訂會議中提出該機制，其概念順利被採納於目前的 IEEE 802.16m 標準草稿中。

至於改善移動性使用者在睡眠模式中具有非即時性傳輸資料量（Non-real-time Traffic）時的能量保存效能，一個以 IEEE 802.16m 睡眠模式運作的概念為基礎，利用統計方式控制睡眠視窗的方法則是在第四個議題中被提出探討。該機制建構一個不連續時間的馬可夫控制的卜瓦松程序（Discrete-time Markov-modulated Poisson process, dMMPP）以表達非即時性傳輸資料的狀態。此外，利用部分環境可知馬可夫決策程序（Partially Observable Markov Decision Process, POMDP）來推測目前傳輸資料量的狀態。藉由該猜測的狀態加上可容忍的延遲限制和佇列大小（Queue Size）的考量，兩個次佳的策略在此方法中被提出，以達到改善能量保存之目的。

Adaptive Communication and Power-saving Mechanisms

in Mobile Broadband Wireless Networks

Student: Chung-Hsien Hsu   Advisor: Dr. Kai-Ten Feng

Institute of Communication Engineering
National Chiao Tung University

ABSTRACT

The growing demand for mobile Internet and wireless multimedia applications has motivated the development of broadband wireless-access technologies in recent years. While various broadband wireless-access technologies are under development at this time, mobile WiMAX based on IEEE 802.16 standards is the leading technology already in service today. This dissertation deals with some problems regarding IEEE 802.16-based mobile broadband wireless networks. Specifically, four topics are investigated: an adaptively direct communication approach for subscriber stations (SSs) in an IEEE 802.16 point-to-multipoint (PMP) network, a predictive motion and interference-based scheduling algorithm for mobile stations (MSs) with direct communication, a flexible listening window adjustment approach, and a statistical sleep window control approach for IEEE 802.16m sleep mode operations.

A flexible and contention-free communication approach is proposed to support direct communication between SSs in an IEEE 802.16 PMP network. The base station (BS) is coordinating and arranging specific time periods for the two SSs that are actively involved in packet transmission. According to the channel conditions among the BS and SSs, the packet transmission operation is switched between the direct link and indirect links in the proposed approach, which improves the bandwidth utilization and consequently increases the user throughput.

Based on the procedures mentioned above, a scheduling algorithm for each pair of MSs that is expected to conduct direct communication is presented. Both the interference region and feasible region for the pair of MSs to perform direct communication are studied and calculated. Based on these two types of information, the algorithm properly arranges the MSs to conduct direct communication in either downlink subframe or uplink subframe, or to communicate with the conventional communication manner, i.e., via the BS. Furthermore, by incorporating the motion prediction mechanism with the scheduling algorithm, the control overheads regarding the updates of MSs positions can be efficiently reduced.

On the other hand, in order to enhance the efficiency of energy conservation, a flexible listening window adjustment approach for an MS in the sleep mode with real-time traffic is present. Instead of window-based operation specified in the IEEE Std. 802.16e-2005, a cycle-based mechanism is considered in the proposed approach. With the consideration of tolerable delay constraint, the approach dynamically adjusts the ratio of listening window to sleep window for each sleep cycle based on the number of both buffered and retransmitted packets. It is worthwhile to mention that the concept of flexible listening window adjustment approach has been proposed by the author and is adopted in the IEEE 802.16m standard draft.

For an MS in the sleep mode with non-real-time downlink traffic, a statistical sleep window control approach is provided based on the concepts of IEEE 802.16m sleep mode operation. The proposed approach constructs a discrete-time Markov-modulated Poisson process for representing the states of non-real-time traffic. Furthermore, a partially observable Markov decision process is exploited to conjecture the present traffic state. Based on the estimated traffic state and the considerations of tolerable delay and/or queue size, two suboptimal policies for maximizing energy conservation are proposed within the approach.

Acknowledgements

I would like to express my sincere gratitude to my advisor Professor Kai-Ten Feng for his guidance, motivation, and support throughout my graduate career at National Chiao Tung University. I have learned a lot from his scientific thinking methodology, rigorous work attitude, and his personality of integrity and generosity, which will continue to be good examples in my future research and life.

I would like to thank Professor Chih-Yung Chang, Professor Chung-Ju Chang, Professor Li-Chun Wang, Professor Ching-Yao Huang, and Professor Ming-Jer Tsai for their invaluable comments as members of my dissertation committee.

I would like to thank all my colleagues in Professor Feng's research group, Mobile Intelligent Network Technology Laboratory, for their patience and sincere friendship.

I would like to dedicate this thesis to all my family members. Without their continuous encouragement, I would not have completed my graduate study. Finally, I would like to thank my dearest wife, Candy, my lovely boy, Alvin, and my little girl, Candies, for their love, encouragement, and sacrifice.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The growing demand for mobile Internet and wireless multimedia applications has motivated the development of broadband wireless-access technologies in recent years. While various broadband wireless-access technologies are under development at this time, mobile WiMAX based on IEEE 802.16 standards is the leading technology already in service today. The mobile WiMAX is being deployed via 475 service providers in 140 countries worldwide, including the United States, Japan, Korea, Taiwan, Europe, Australia, Mideast, etc [1]. With the characteristics of providing high data rate and enabling various usage models, the mobile WiMAX is considered as a candidate for the next generation mobile broadband wireless networks (MBWNs) as well as the first generation of mobile Internet technology [2].

The IEEE 802.16 Working Group established by the IEEE Standards Board in 1999 has developed and published several versions of air interface standards for wireless metropolitan area networks (WMANs) [3]. The IEEE 802.16 standards define the structure of the physical (PHY) and medium access control (MAC) operations that occur between base station (BS) and subscriber stations (SSs). The IEEE Std. 802.16-2001 [4] is the initial solution to provide a fixed broadband wireless-access technology. The standard described an orthogonal frequency-division multiplexing (OFDM)-based point-to-multipoint (PMP) solution, wherein all the communication between the BS and SSs were controlled by the BS. The IEEE Std. 802.16a-2003 amendment [5] added the specifications of Mesh mode and lower-frequency operations

to the original standard. In the Mesh mode, communication can occur directly between SSs with potential packet forwarding via other SSs. Both of the standards were revised and consolidated as the IEEE Std. 802.16-2004 [6] published in 2004 for fixed broadband wireless networks (FBWNs). The issues of mobility management and energy conservation were addressed in the IEEE Std. 802.16e-2005 amendment [7], which was utilized as the basic document to standardize the current mobile WiMAX system.

The IEEE 802.16 Maintenance Task Group is chartered to maintain IEEE 802.16 standards through ongoing maintenance activities. The latest revision is the IEEE Std. 802.16-2009 [8] published on May 2009. This revision supersedes and makes obsolete IEEE Std. 802.16-2004 and all of its subsequent amendments and corrigenda. In this standard, one of the major modifications is the removal of the Mesh mode, which results in that the PMP mode becomes the only solution in IEEE 802.16 networks. However, the inefficiency within the PMP mode occurs while two SSs are intended to conduct communication. It is required for the data packets between the SSs to be forwarded by the BS even though the SSs are adjacent with each others. Due to the packet rerouting process, communication bandwidth is wasted which consequently increases control overhead and packet-rerouting delay.

On the other hand, since January 2007, the IEEE 802.16 Working Group has embarked on the development of a new amendment of the IEEE 802.16 standard in the Task Group m. The new standard is focusing on the specification of an advanced air interface and system architecture to meet the requirements of the International Telecommunication Union - Radiocommunication/International Mobile Telecommunications (ITU-R/IMT)-advanced for fourth-generation (4G) systems. In other words, advanced services with characteristics of higher data rate and higher mobility are expected to be supported by the next-generation mobile WiMAX based on the new IEEE 802.16m standard. Since mobility is considered a key feature in MBWNs, how to prolong the battery lifetime of mobile stations (MSs) has been recognized as one of the critical issues. There are existing power-saving mechanisms specified in the IEEE Std. 802.16e-2005; while improved operations for energy conservation are also considered desirable for the IEEE 802.16m system.

Therefore, in this dissertation, the following two issues for IEEE 802.16-based MBWNs are studied and addressed. How to enhance the bandwidth utilization as well as user throughput in the PMP networks? And how to improve the energy conservation for MSs?

1.1 Dissertation Overview

Major contributions of this dissertation are listed as follows:

- An adaptively direct communication approach for SSs in an IEEE 802.16 OFDM/TDD-based PMP network (Chapter 2)

- A predictive motion and interference-based scheduling algorithm for MSs with direct communication (Chapter 3)

- A flexible listening window adjustment approach for IEEE 802.16m sleep mode operation (Chapter 4)

- A statistical sleep window control approach for IEEE 802.16m sleep mode operation (Chapter 5)

A brief overview of each chapter are given in subsequent paragraphs.

Chapter 2 provides a flexible and contention-free approach to support direct communication between SSs in an IEEE 802.16 OFDM/TDD-based PMP network. The BS is coordinating and arranging specific time periods for the two SSs that are actively involved in packet transmission. According to the channel conditions among the BS and SSs, the packet transmission operation is switched between the direct link and indirect links in the proposed approach, which improves the bandwidth utilization and consequently increases the user throughput. A comprehensive architecture design associated with the extended frame structures for the proposed approach are described. Moreover, an analytical model is also conducted to justify the correctness and effectiveness of the proposed scheme.

In **Chapter** 3, a scheduling algorithm for each pair of MSs that is expected to conduct direct communication is presented. Both the interference region and feasible region for the pair

of MSs to perform direct communication are studied and calculated. Based on these two types of information, the algorithm properly arranges the MSs to conduct direct communication in either DL subframe or UL subframe, or to communicate with the conventional communication manner, i.e., via the BS. Furthermore, since MSs may move around, a motion prediction mechanism is proposed and is considered within the scheduling algorithm, which effectively reduces the control overheads regarding the updates of MSs positions.

A flexible listening window adjustment approach for an MS in the sleep mode with real-time traffic is proposed in **Chapter** 4. Instead of window-based operation specified in the IEEE Std. 802.16e-2005, a cycle-based mechanism is considered in the proposed approach. With the consideration of tolerable delay constraint, the approach dynamically adjusts the ratio of listening window to sleep window for each sleep cycle based on the number of both buffered and retransmitted packets. Therefore, the performance of both energy efficiency and packet loss rate are improved in the proposed scheme. It is worthwhile to mention that the concept of flexible listening window adjustment approach has been proposed by the authors and is adopted in the IEEE 802.16m standard draft.

For an MS in the sleep mode with non-real-time downlink traffic, a statistical sleep window control approach, based on the concepts of IEEE 802.16m sleep mode operation, is provided in **Chapter** 5. The proposed approach constructs a discrete-time Markov-modulated Poisson process (dMMPP) for representing the states of non-real-time traffic. Furthermore, a partially observable Markov decision process (POMDP) is exploited to conjecture the present traffic state. Based on the estimated traffic state and the considerations of tolerable delay and/or queue size, two suboptimal policies for maximizing energy conservation are proposed within the approach.

Each of the following four chapters are began with an introduction and some preliminaries to describe the subject problem. The proposed approach, analytical model, and performance evaluation are presented in the subsequent sections. Finally, the concluding remarks are given in the end of each chapter.

Chapter 2

Adaptively Direct Communication Approach for Subscriber Stations in Broadband Wireless Networks

2.1 Introduction

The IEEE 802.16 standards for wireless metropolitan area networks (WMANs) are designed to satisfy various demands for high capacity, high data rate, and advanced multimedia services [9]. The medium access control (MAC) layer of IEEE 802.16 networks supports both point-to-multipoint (PMP) and mesh modes for packet transmission [6]. Based on application requirements, it is suggested in the standard that only one of the modes can be exploited by the network components within the considered time period, and the PMP mode is considered the well-adopted one. In the PMP mode, packet transmission is coordinated by the base station (BS), which is responsible for controlling the communication with multiple subscriber stations (SSs) in both downlink (DL) and uplink (UL) directions. All the traffic within an IEEE 802.16 PMP network can be categorized into two types, including inter-cell traffic and intra-cell traffic. For the inter-cell traffic, the source/destination pair of each traffic flow are located in different cells. On the other hand, the intra-cell traffic is defined if they are

situated within the same cell. The inefficiency within the PMP mode occurs while two SSs are intended to conduct packet transmission, i.e., the intra-cell traffic between the SSs. It is required for data packets between the SSs to be forwarded by the BS even though the SSs are adjacent with each others. Due to the packet rerouting process, the communication bandwidth is wasted which consequently increases the packet-rerouting delay.

In order to alleviate the drawbacks resulted from the indirect transmission, a directly communicable mechanism between SSs should be considered in IEEE 802.16 networks. Several direct communication approaches have been proposed for different types of networks. The direct-link setup (DLS) protocol is standardized in the IEEE 802.11z draft standard to support direct communication between two SSs in wireless local networks [10]. However, the DSL protocol is designed as a contention-based mechanism, which does not guarantee the access of direct link setup and data exchanges between two SSs. The dynamic slot assignment (DSA) scheme for Bluetooth networks is proposed in [11, 12], which is primarily implemented based on the characteristics of the Bluetooth standard [13]. Since the frame structures and the medium access mechanisms are different among these wireless communication technologies, both the DLS protocol and DSA scheme cannot be directly applied to IEEE 802.16 networks. For the IEEE 802.16 PMP networks, the virtual direct link access (VDLA) mechanism is proposed in [14], which partially overlaps the DL and UL subframes within a single frame. The source SS and destination SS are scheduled in the overlapped time intervals in order to accomplish the direct transmission. However, since the channel conditions among the BS and the two SSs can be different, it will not always attainable to assume that both the source and destination SSs process the common burst profile in the VDLA scheme.

In this work, an adaptive point-to-point communication (APC) approach is proposed to support direct communication for SSs. The APC approach is designed as a flexible and contention-free scheme especially for the time division duplexing-based IEEE 802.16 PMP networks. The complete data structures and procedures for implementing direct communication are proposed. According to the relative locations and channel conditions among the BS and SSs, the packet transmission operation is switched between the direct link and in-

direct links in the APC approach, which results in enhanced network throughput. While the direct link approach is selected, the required bandwidth, communication overhead, and packet latency can be greatly reduced. The effectiveness of the APC approach is evaluated and validated via both the numerical analysis and extensive simulation studies. It can be shown that the proposed APC approach outperforms the conventional IEEE 802.16 transmission mechanism in terms of user throughput, communication overhead, and packet latency. In summary, the contributions of this work are listed as follows: (a) the proposal of a comprehensive architectural design associated with the extended frame structures for conducting direct communication; (b) an adaptive communication approach that can dynamically select the most efficient packet transmission scheme between the direct communication and indirect communication; (c) an approach that is fully compatible and can be directly integrated with the existing protocols defined in the IEEE 802.16 standard; and (d) numerical analysis and extensive simulations to justify the effectiveness of the proposed scheme.

The remainder of this chapter is organized as follows. Section 2.2 briefly reviews the MAC frame structure and packet transmission mechanism in IEEE 802.16 PMP networks. The proposed APC approach, consisting of management structures, admission control scheme, and direct communication procedures, is described in Section 2.3; while the numerical analysis is carried in Section 2.4. Both the performance evaluation and validation of the APC approach are conducted in Section 2.5. Section 2.6 draws the concluding remarks.

2.2 Preliminaries

2.2.1 IEEE 802.16 OFDM/TDD Frame Structure

The PMP mode is considered the well-adopted network configuration in IEEE 802.16 networks wherein the BS is responsible for controlling all the communication among SSs. Two duplexing techniques are supported for the SSs to share common channels, i.e., time division duplexing (TDD) and frequency division duplexing. The MAC protocol is structured to support multiple physical (PHY) layer specifications in the IEEE 802.16 standard. In this

7
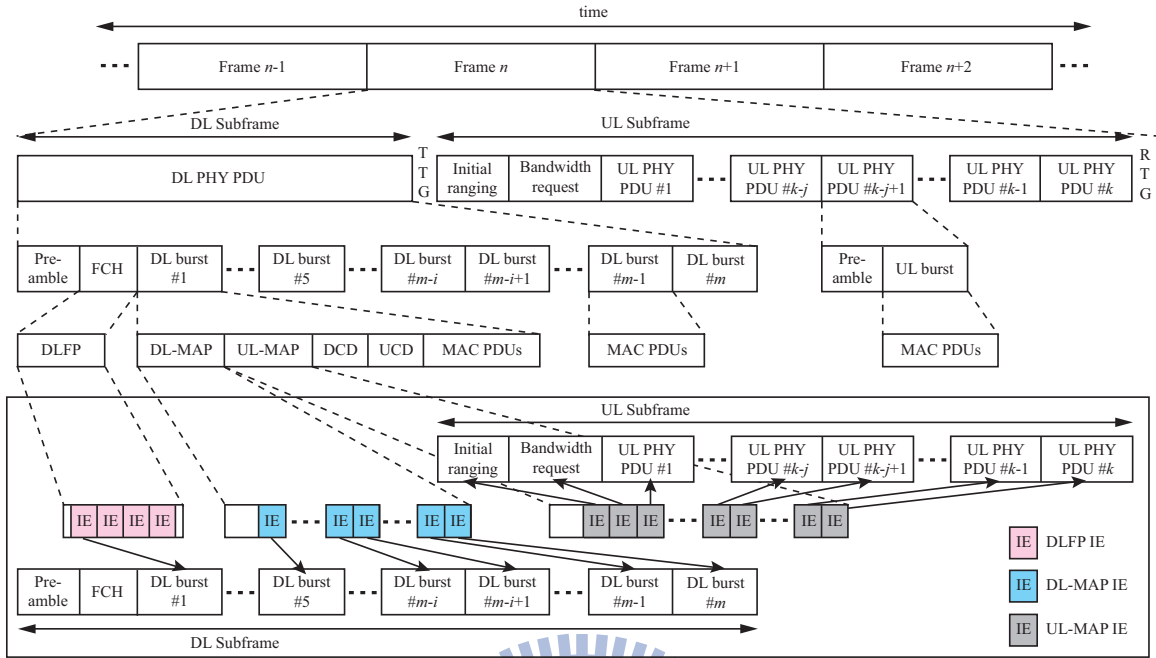
Figure 2.1: Schematic diagram of IEEE 802.16 PMP OFDM frame structure with TDD mode.

work, the WirelessMAN-OFDM PHY, utilizing the orthogonal frequency division multiplexing (OFDM), with TDD mode is exploited for describing the design of the proposed APC approach.

Fig. 2.1 illustrates the schematic diagram of the IEEE 802.16 PMP OFDM frame structure with TDD mode. It can be observed that each frame consists of a DL subframe and a UL subframe. The DL subframe contains only one DL PHY protocol data unit (PDU), which starts with a long preamble for PHY synchronization. The preamble is followed by a frame control header (FCH) burst and several DL bursts. A DL frame prefix (DLFP), which is contained in the FCH, specifies the burst profile and length for the first DL burst (at most four) via the information element (IE). It is noted that each DL burst may contain an optional preamble and more than one MAC PDUs that are destined for the same or different SSs. The first MAC PDU followed by the FCH is the DL-MAP message, which employs DL-MAP IEs to describe the remaining DL bursts. The DL-MAP message can be excluded in the case that the DL subframe consists of less than five bursts; nevertheless, it must still be sent out periodically to maintain synchronization. A UL-MAP message immediately following

the DL-MAP message denotes the usage of UL bursts via UL-MAP IEs. An interval usage code, corresponding to a burst profile, describes a set of transmission parameters, e.g., the modulation and coding type, and the forward error correction type. The DL interval usage code (DIUC) and UL interval usage code (UIUC) are specified in the DL channel descriptor (DCD) and UL channel descriptor (UCD) messages respectively. The BS broadcasts both the DCD and UCD messages periodically to define the characteristics of the DL and UL physical channels respectively.

On the other hand, as can be seen from Fig. 2.1, the UL subframe starts with the contention intervals that are specified for both initial ranging and bandwidth request. It is noted that more than one UL PHY PDU can be transmitted after the contention intervals. Each UL PHY PDU consists of a short preamble and a UL burst, where the UL burst transports the MAC PDUs for each specific SS. Moreover, the transmit/receive transition gap (TTG) and the receive/transmit transition gap (RTG) are inserted in between the DL and the UL subframes and at the end of each frame respectively. These two gaps provide the required time for the BS to switch from the transmit to receive mode and vice versa.

2.2.2 Packet Transmission Mechanism

A connection in IEEE 802.16 PMP networks is defined as a unidirectional mapping between the BS and an MS, which is identified by a 16-bit connection identifier (CID). Two kinds of connections, including management connections and transport connections, are defined in the IEEE 802.16 standard. The management connections are utilized for delivering MAC management messages; while the transport connections are employed to transmit user data. During the initial ranging of a SS, a pair of UL/DL basic connections are established, which belong to a type of the management connections. It is noted that a single Basic CID is assigned to a pair of UL/DL basic connections, which is served as the identification number for the corresponding SS. Thus the SS uses the individual transport CID to request bandwidth for each transport connection while the BS arranges the accumulated transmission opportunity by addressing the Basic CID of the SS.
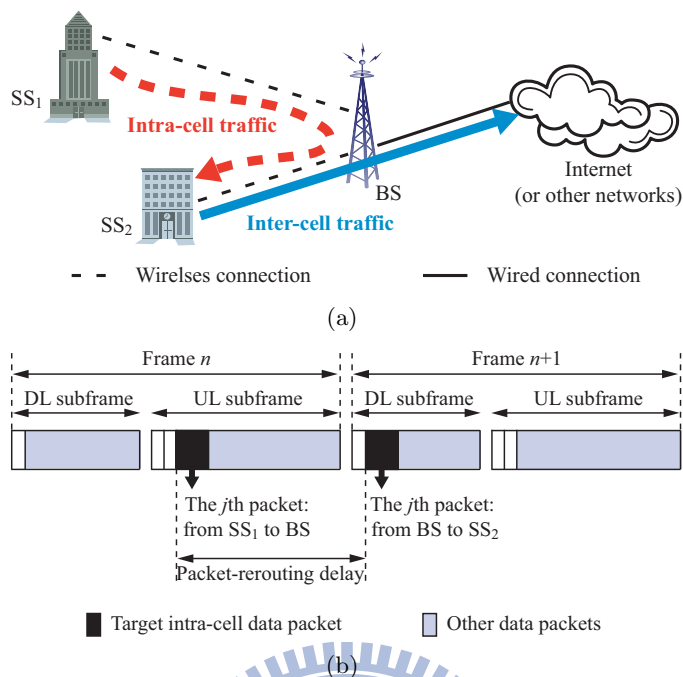
Figure 2.2: Example of packet transmission in IEEE 802.16 PMP networks: (a) network topology and (b) conventional transmission scheme in time sequence.

Fig. 2.2 depicts the conventional packet transmission mechanism of IEEE 802.16 PMP networks. An exemplified network topology that consists of one BS and two neighboring SSs is shown in Fig. 2.2(a). Two types of traffic exist in the network: inter-cell traffic and intra-cell traffic. For the inter-cell traffic, the source and the destination for each traffic flow are located in different cells, e.g., the traffic flow of SS_2 for accessing the Internet. On the other hand, the intra-cell traffic is defined while the source and destination are situated within the same cell network, such as the traffic flow between SS_1 and SS_2 in Fig. 2.2(a). Considering the scenario that SS_1 intends to communicate with its neighboring station SS_2, two transport connections are required to be established via the service flow management mechanism for the intra-cell traffic, i.e., the UL transport connection from SS_1 to the BS and the DL transport connection from the BS to SS_2. Fig. 2.2(b) illustrates the conventional transmission mechanism of IEEE 802.16 PMP networks in time sequence. In the most ideal case, the jth intra-cell packet, transmitted from SS_1 to the BS in the nth frame, will be forwarded to SS_2 in the $(n+1)$th frame by the BS. The rerouting process apparently requires twice of communication bandwidth

10

for achieving the intra-cell packet transmission, which consequently increases communication overhead resulted from the duplication of data packets. Moreover, the delay time for packet-rerouting can be more than one half of a frame duration while the packet transmission from the BS to SS_2 is postponed to a latter DL subframe.

2.2.3 Problem Statement

Based on the aforementioned drawbacks of the conventional transmission mechanism in IEEE 802.16 PMP networks, the object problem of this work is described as follows:

Problem 1 (Direct Communication Problem). *Given a pair of SSs that are actively involved in packet transmission, how to conduct efficient communication for the pair in order to increase the user throughput as well as to reduce the communication bandwidth and control overheads?*

2.3 Adaptive Point-to-point Communication (APC) Approach

In this section, the proposed APC approach is presented for IEEE 802.16 PMP networks. The concept of the proposed scheme is to design a point-to-point directly communicable mechanism for SSs. Based on the channel conditions among the BS and SSs, the packet transmission operation is switched automatically between the conventional indirect-link scheme and the proposed direct-link mechanism. Therefore, the following three conditions are designed to be satisfied in the proposed APC approach:

- *Flexible Condition.* This condition states that the APC approach is dynamically implemented. In other words, the direct communication between two SSs can be initiated, reactivated, and terminated by either the BS or the SS at any frame.

- *Contention-free Condition.* The inquiry for direct communication between two SSs may be requested by different pairs of SSs simultaneously. This condition indicates that the proposed APC approach should hold the contention-free processes, which include request, proceeding, and termination among those directly communicable pairs.

11
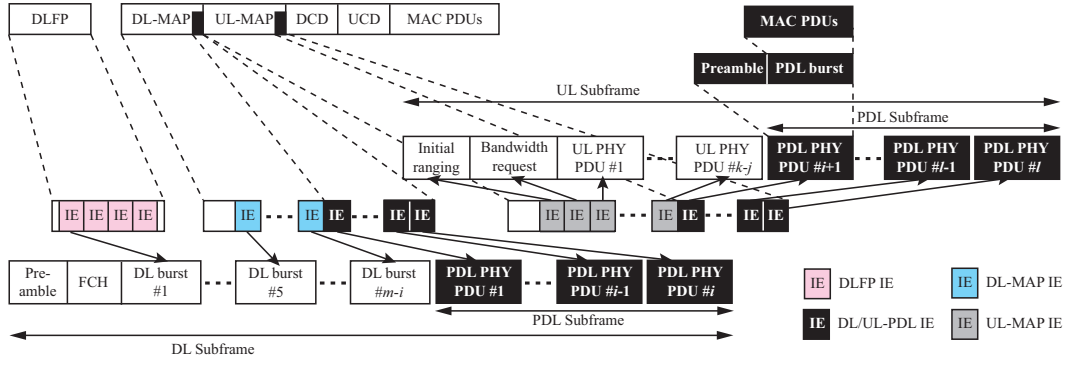
Figure 2.3: Schematic diagram of IEEE 802.16 PMP OFDM frame structure with APC approach.

- *Backward-compatible Condition.* This condition indicates that the proposed APC approach is compatible and can be directly integrated with the existing protocols that are defined in the IEEE 802.16 standard.

It can be expected that the APC approach reduces both the required communication bandwidth and the control overhead for intra-cell traffic and also eliminates the packet-rerouting delay time. Furthermore, the network throughput is enhanced due to the automatic selection of efficient transmission manner between two schemes.

2.3.1 Architecture and Management Structure

Comparing with the original IEEE 802.16 PMP frame structure, the following two modifications are exploited by the proposed APC approach: (*i*) one or more point-to-point direct link (PDL) PHY PDUs are appended after the original DL PHY PDU in DL subframe; and (*ii*) a specific number of UL PHY PDUs are replaced by the PDL PHY PDUs in UL subframe. The schematic diagram of the IEEE 802.16 PMP frame structure with APC approach is illustrated in Fig. 2.3, where the adjustments are depicted by the block rectangles. It can be observed that the proposed PDL subframe is designed to be a subset of a DL subframe and/or a UL subframe. A PDL subframe consists of one or more PDL PHY PDUs, which start with a short preamble followed by a PDL burst. Each PDL burst is designed to transport the MAC PDUs for each specific SS. Moreover, the DL-PDL IE and UL-PDL IE are designed to be included

in the DL-MAP and UL-MAP messages respectively, which are utilized to specify the burst profiles and lengths for the corresponding PDL bursts. In order to fully compatible with the existing IEEE 802.16 standard, three categories of management structures,including the DL-PDL IE and UL-PDL IE, the PDL subheader, and the PDL burst profile change request (PBPC-REQ) and response (PBPC-REP) messages, are proposed in the APC approach. The functionalities and formats of these structures are detailed as follows.

DL-PDL IE and UL-PDL IE

As shown in Fig. 2.3, the DL-MAP message defines the access to the DL channel; while the UL-MAP message characterizes and schedules the UL subframe. In other words, both the DL-MAP and UL-MAP messages adopt the IEs along with DIUC and UIUC to describe the burst profiles and lengths of the corresponding DL and UL bursts respectively. The proposed DL-PDL IE and UL-PDL IE are designed to depict burst profiles and lengths of their corresponding PDLs in the DL and UL subframes respectively. The format of the DL-PDL IE is defined as in Table 2.1, which is considered as a new type of the extended DIUC dependent IE within the OFDM DL-MAP IE. The extended DIUC field identifies the IE type; while the size of IE is indicated in the length field. The duration field specifies the length of the corresponding PDL burst in the OFDM symbols. On the other hand, the proposed UL-PDL IE (as defined in Table 2.2) is designed to be a new type of the UL extended IE that is contained in the OFDM UL-MAP IE. It is noticed that both the proposed DL-PDL IE and UL-PDL IE are designed to conform to the formats of the DL-MAP dummy IE and UL-MAP dummy IE, respectively, which are defined in the IEEE 802.16 standard for supporting newly developed IEs.

PDL Subheader

In the original IEEE 802.16 standard, six types of subheaders have been specified immediately followed by the generic MAC header. The proposed PDL subheader is designed to be a new type of per-PDU subheader that can be transmitted in both the DL and UL directions. It

13

Table 2.1: OFDM DL-PDL IE format

Syntax	Size (bit)
DL_PDL_IE() {	-
Extended DIUC	4
Length	4
DIUC	4
Duration	12
}	-

Table 2.2: OFDM UL-PDL IE format

Syntax	Size (bit)
UL_PDL_IE() {	-
Extended UIUC	4
Length	4
UIUC	4
Padding nibble, if needed	4
}	-

is utilized to implement the request, response, announcement, and termination of the APC approach. The presence of the PDL subheader is indicated by a reserved bit in the generic MAC header. Table 2.3 illustrates the format of the PDL subheader; while its encoded type field is shown in Table 2.4. The location type (LT) field (as shown in Table 2.5) within the PDL subheader is adopted to either request or present the location information (LI) IE which includes the latitude, longitude, and altitude information [15]. The usages of the location information will further be explained in Subsection 2.3.2.

PBPC-REQ and PBPC-REP messages

In the IEEE 802.16 standard, the adaptive modulation and coding (AMC) mechanism is exploited as the link adaption technique to improve network performance on time-varying channels. The BS selects an adequate modulation and coding scheme (MCS) for a SS based on the reported signal-to-interference and noise ratio (SINR) value. The selected MCS is specified in the burst profile and is represented by the DIUC and UIUC value for the DL and UL directions respectively. Moreover, the BS permits the changes in the DIUC value that are suggested by the SS via the burst profile change request message. Similarly, both the

14

Table 2.3: PDL subheader format

Syntax	Size (bit)
PDL Subheader() {	-
Type	3
Location Type (LT)	2
if (Type == 000) {	-
SSID	48
LI_IE()	64
}	-
if (Type == 001)	-
CID	16
if (Type == 010) {	-
if (LT == 01)	-
LI_IE()	64
}	-
if (Type == 011) {	-
SINR	8
if (LT == 10)	-
LI_IE()	64
}	-
if (Type == 1xx) {	-
CID	16
CID	16
for ($i = 1; i <= n; i++$)	-
CID	16
}	-
Reserved	3
}	-

Table 2.4: Type field encodings for PDL subheader

Type	PDL subheader type
000	PDL request from source SS to BS.
001	PDL request from BS to destination SS.
010	PDL response from source SS to BS.
011	PDL response from destination SS to BS.
100	PDL termination request from SS.
101	PDL announcement for termination.
110	PDL announcement for confirming the request.
111	PDL request from BS to considered SSs.

Table 2.5: LT field encodings for requesting/presenting location information

LT	DL/UL
00	Not request/present.
01	Request/present source SS's location information.
10	Request/present destination SS's location information.
11	Request both SSs' location information./-

Table 2.6: PBPC-REQ/REP message format

Syntax	Size (bit)
PBPC-REQ/REP_Message_Format(){	-
Management Message Type	8
Reserved	4
CID	16
PIUC	4
Configuration Change Count	8
}	-

proposed PBPC-REQ and PBPC-REP messages are designed to change the MCS for PDLs, where the format of both messages are shown in Table 2.6. The PBPC-REQ message is utilized to request the adjustment of the PDL interval usage code (PIUC) value for the PDL burst. The BS will respond with the proposed PBPC-REP message for either confirming or denying the alteration in the PIUC value. It is noticed that the PIUC value represents either the DIUC in DL direction or the UIUC in UL direction. The operation of the PBPC-REQ and PBPC-REP messages will be shown in the following subsection.

2.3.2 Admission Control Scheme

In the APC approach, some criteria should be exploited to determine the execution of direct communication between two SSs. A two-tiered admission control scheme for a BS and two attached SSs is presented in this subsection. In wireless communication system, the data transmission range for each station is proportional to its corresponding transmission power. In order to avoid additional power consumption resulted from the APC approach, the distance factor is considered as the first-tiered constraint, energy-oriented (EO) constraint, which is

Table 2.7: OFDM modulation and coding schemes

MCS index	Modulation	Coding rate	Coded block size (byte)	Receiver SNR (dB)
0	BPSK	1/2	24	3.0
1	QPSK	1/2	48	6.0
2	QPSK	3/4	48	8.5
3	16-QAM	1/2	96	11.5
4	16-QAM	3/4	96	15.0
5	64-QAM	2/3	144	19.0
6	64-QAM	3/4	144	21.0

defined as

$$\mathcal{C}_1: \quad D(SS_s, SS_d) \leq D(SS_s, BS),$$

where $D(x, y)$ denotes the relative distance between x and y; while the source SS and destination SS of an intra-cell traffic is represented as SS_s and SS_d respectively. In other words, the transmission power utilized by SS_s for achieving direct communication is adjusted to be equal to or less than that as specified in the conventional IEEE 802.16 mechanism.

On the other hand, for the purpose of enhancing the efficiency for data transmission, channel conditions among the BS, SS_s, and SS_d should be taken into account. Different MCSs associated with various number of data bits are adopted for data transmission under different channel conditions. Based on the channel states and the corresponding MCS, the second-tiered constraint, throughput-oriented (TO) constraint, is defined as

$$\mathcal{C}_2: \quad T_{PDC}(SS_s, SS_d) \geq T_{Conv}(SS_s, SS_d),$$

where $T(SS_s, SS_y)$ represents the raw user throughput defined as "*number of bits per second that is received by the destination SS_d while the source is SS_s*". In other words, the raw user throughput resulted from the PDC approach (T_{PDC}) should be at least equal to or higher than that from the conventional IEEE 802.16 mechanism (T_{Conv}). The values of both T_{PDC} and T_{Conv} are derived as the description in the following paragraph.

Table 2.7 shows the supported MCSs that are specified in the IEEE 802.16 standard. In

the considered OFDM system, the raw data rate R_d of the MCS with index ξ is represented as

$$R_d[\xi] = \frac{B_u[\xi]}{T_s}, \tag{2.3}$$

where T_s is the OFDM symbol duration. The notation $B_u[\xi]$ indicates the number of uncoded bits per OFDM symbol of the MCS with index ξ, which is obtained as

$$B_u[\xi] = N_d \cdot \log_2 M \cdot R_c[\xi], \tag{2.4}$$

where N_d denotes the number of data subcarriers and $R_c[\xi]$ is the coding rate of the MCS with index ξ. The value of the parameter M depends on the adopted MCS, i.e., $M = 2$ for BPSK, $M = 4$ for QPSK, $M = 16$ for 16-QAM, and $M = 64$ for 64-QAM. Moreover, the OFDM symbol duration T_s can be acquired as

$$T_s = T_b + T_g = T_b + G \cdot T_b = \frac{1+G}{\triangle f}, \tag{2.5}$$

where T_b and T_g represent the useful symbol time and the cyclic prefix (CP) time respectively. The notation G denotes the ratio of T_g to T_b. The subcarrier spacing $\triangle f$ is obtained as

$$\triangle f = \frac{F_s}{N_s} = \frac{8000}{N_s} \cdot \left\lfloor \frac{n \cdot BW}{8000} \right\rfloor, \tag{2.6}$$

where N_s indicates the number of total subcarriers. The notation F_s represents the sampling frequency with its value specified by the IEEE 802.16 standard as in (2.6), where n is the sampling factor and BW is the channel bandwidth. By substituting (2.6) into (2.5), the OFDM symbol time can be approximated as

$$T_s = \frac{N_s}{F_s} \cdot (1 + G) \approx \frac{N_s}{n \cdot BW} \cdot (1 + G). \tag{2.7}$$

With (2.4) and (2.7), the raw data rate R_d of the MCS with index ξ in (2.3) becomes

$$R_d[\xi] \approx \frac{N_d \cdot \log_2 M \cdot R_c[\xi] \cdot n \cdot BW}{N_s \cdot (1 + G)}.$$

(2.8)

Based on (2.8), the raw user throughput by adopting the PDC approach is acquired as

$$T_{PDC}(SS_s, SS_d) = R_d[\xi_{(s,d)}],$$

(2.9)

where $\xi_{(s,d)}$ represents the index of the MCS that will be assigned to the direct link between SS_s and SS_d. On the other hand, the raw user throughput in the conventional IEEE 802.16 mechanism is constrained by the two-hop transmission, i.e., from SS_s to BS and from BS to SS_d. Thus the T_{Conv} can be obtained as

$$T_{Conv}(SS_s, SS_d) = \frac{1}{2} R_d[\phi_{(s,d)}],$$

(2.10)

where

$$\phi_{(s,d)} = \min\left[\xi_{(s,BS)}, \xi_{(BS,d)}\right].$$

(2.11)

The notation $\xi_{(s,BS)}$ denotes the index of the MSC utilized in the link between SS_s and the BS; while that is assigned to the link between the BS and SS_d is represented as $\xi_{(BS,d)}$.

2.3.3 Direct Communication Procedures

Based on the aforementioned management structures and admission control scheme, the direct communication procedures of the APC approach are explained in this subsection.

Considering a basic IEEE 802.16 PMP network that consists of a BS and two SSs, an intra-cell traffic flow is existed between the source SS_s and destination SS_d. Two transport connections are established for packet transmission, i.e., the UL transport connection (with $CID = K_1$) from SS_s to the BS and the DL transport connection (with $CID = K_2$) from the BS to SS_d. It is noted that K_1 and K_2 are denoted as specific CID values.

The initialization of direct communication is achieved by conducting the procedure of link
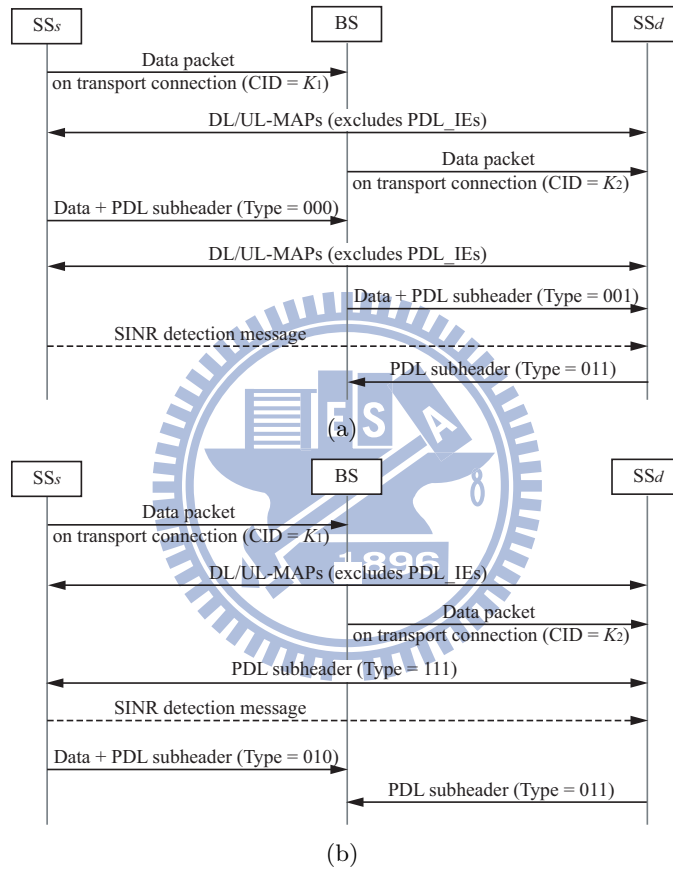
Figure 2.4: Schematic diagram of link request and information collection procedure for direct communication: (a) SS-initiated procedure and (b) BS-initiated procedure.

request and information collection. The pair of SS_s and SS_d that anticipates to establish a direct link are required to provide their location information and channel conditions to the BS. It is assumed that the SSs can acquire their location information by either using the GPS or performing network-based location estimation techniques [16, 17]. The collected information is utilized in the admission control scheme mentioned above. Fig. 2.4 illustrates an exemplified message flows for the initialization procedure of direct communication.

In the case that SS_s intends to conduct direct communication with SS_d, as shown in Fig. 2.4(a), it attaches a PDL subheader to a data packet that will be delivered to the BS. The PDL subheader is utilized to request a direct link establishment with SS_d, where the Type field in the subheader is denoted as 000 and the 48-bit MAC address of SS_d is filled in the SSID field (as shown in Tables 2.3 and 2.4). The LT field is assigned as 01 and the location information of SS_s will be filled into the corresponding LI IE. As the BS receives the request PDL subheader from SS_s, the BS will attach a PDL subheader (with Type field = 001) to the data packet and conduct the transmission to SS_d. The transport CID K_1 will be carried in the subheader for indicating that SS_s is the source of the direct communication, and it will send an SINR detection message via that connection. If the BS does not possess the location information of SS_d, it will set the LT field as 10 for requesting the location information of SS_d.

Furthermore, the BS will arrange a DL burst for SS_s with the assignment in the corresponding DL-MAP message. The CID of the UL transport connection (i.e., CID = K_1) will be recorded in the CID field of the DL-MAP IE; while the DIUC field is set with the value that corresponds to the BPSK 1/2 MCS. Specifically, the DL burst for the connection with CID K_1 is prepared for SS_s to transmit an SINR detection message to SS_d by using the BPSK 1/2 MCS. It is noticed that the UL transport CID will be assigned in the DL subframe, which will not happen by adopting the conventional mechanism. Therefore, the SS_s is aware that the DL burst for the UL transport CID is utilized to transmit the SINR detection message. After receiving the PDL subheader and SINR detection message from the BS and SS_s respectively, SS_d will transmit a response PDL subheader with Type 011. The

21

average SINR value calculated from the SINR detection message will also be recorded in the response PDL subheader. It is noted that the location information of SS_d is carried in the response PDL subheader if it is required by the BS.

On the other hand, the BS-initiated direct communication procedure is shown in Fig. 2.4(b). Contrary to the SS-initiated procedure, the BS actively announces the link request along with the PDL subheader (with Type field = 111) to the specific SSs, i.e, SS_s and SS_d. The Basic CIDs for both SS_s and SS_d are specified within the first two CID fields of the PDL subheader as shown in Table 2.3; while the corresponding UL transport CID K_1 is written in the third CID field. As the SS_s receives the requesting PDL subheader with LT = 01 or 11 from the BS, the SS_s will attach a PDL subheader (with Type field = 010) to a data packet that will be delivered to the BS. Correspondingly, the SS_s will utilize the replying PDL subheader to provide its location information that is requested by the BS. The remaining procedures of the BS-initiated APC approach are similar to that of the SS-initiated case, such as the SINR detection procedure between SS_s and SS_d, and the response PDL subheader from SS_d.

The BS executes the admission control procedure after it received the response PDL subheader transmitted from SS_d. Based on the collected information, the aforementioned two-tiered control scheme is exploited by the BS to either confirm or deny the direct communication request between SS_s and SS_d. It is noted that the constraints $\mathcal{C}_1$ and $\mathcal{C}_2$ can be exploited either jointly or separately. The performance of the separately implemented mechanisms, i.e., the APC-TO and the APC-EO approaches, will be evaluated via the simulation results in Subsection 2.5.2. The confirming results will be broadcasted along with the PDL subheader (with Type field = 110) by the BS. The Basic CIDs for both SS_s and SS_d are written within the the first two CID fields of the PDL subheader as shown in Table 2.3; while the corresponding confirmed connections are recorded in the remaining CID fields. In the case that all the connections belonging to the indirect transmission from SS_s to SS_d are confirmed by the BS, the individual CIDs will be replaced by the Basic CID of SS_s. In other words, the Basic CID of SS_s will be filled into the third CID field in order to reduce the excessive control
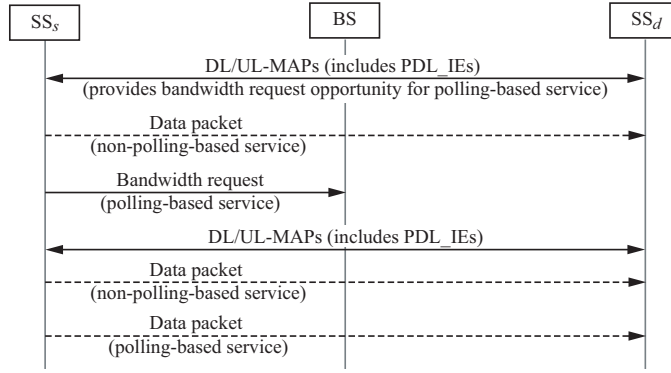
Figure 2.5: Schematic diagram of bandwidth request procedure in APC approach.

overhead caused by the individually confirmed CIDs. Consequently, the BS will arrange the corresponding PDL bursts for conducting direct communication in subsequent frames.

After receiving the confirmation announcement, the considered SSs will activate the procedure of direct communication. According to the received MAPs associated with the PDL IEs, SS_s will conduct packet transmission directly to SS_d within the PDL bursts. Moreover, SS_d will continuously observe and evaluate the cannel condition for the direct link with the adaptation to an appropriate MCS. The calculated SINR is compared with the receiver SNR range of the current MCS (as listed in Table 2.7) by SS_d. If the existing MCS is observed to be improper for the current channel condition, SS_d will initiate a PBPC-REQ message to the BS for suggesting an appropriate MCS via the PIUC value. Consequently, the BS will respond with a PBPC-REP message with a recommended PIUC value. If the PIUC values from both the PDPB-REQ and the PBPC-REP messages are perceived to be the same, the request for the change of MCS will be accepted. If the condition is not satisfied, the PIUC value will remain unaltered.

It is worthwhile to mention that the bandwidth requests are conducted by an SS based on individual transport connection. On the other hand, the bandwidth grant from the BS is executed according to the accumulated requests from the SS. In other words, the bandwidth grant is addressed to the Basic CID of the corresponding SS, not to the individual transport CIDs. As a result, the CID specified for the PDL burst becomes the Basic CID of SS_s. Furthermore, in order to fully compatible with the existing specification, the procedures for the

23

bandwidth request and allocation as specified in the IEEE 802.16 standard are implemented within the proposed APC approach. Fig. 2.5 illustrates the bandwidth request procedure while the APC approach is adopted. It can be observed that the BS preserves PDL bursts for non-polling based service periodically. Furthermore, the BS will continue to provide unicast bandwidth request opportunity for polling-based service based on the original transport CIDs of SS_s. The unicast bandwidth grant of the polling-based service will consequently be assigned to PDL bursts based on the Basic CID of SS_s.

The procedure for the link termination occurs as one of the following two conditions is satisfied: (a) the channel condition of the direct link is becoming worse than that from the indirect channels (i.e., via the BS); (b) the direct communication is determined to be ceased. It is noted that the link termination can be initiate by either the BS or the SSs. In the SS-initiated termination procedure, the SS will transmit a termination PDL subheader (with Type field = 100). As the message is received by the BS, it will broadcast an announcement along with a PDL subheader (with Type field = 101) to both SS_s and SS_d regarding the termination of the direct communication link. On the other hand, for the BS-initiated termination procedure, the termination information is actively announced by the BS. As a result, the BS and the associated SSs will return to adopt the original packet transmission mechanism as defined in the IEEE 802.16 standard.

Fig. 2.6 depicts the flowchart of switching process for packet transmission in the proposed APC approach. The entire process is constructed by the aforementioned procedures. It can be observed that the communication operation switched between direct and indirection manners is dominated by the constraint $\mathcal{C}_2$. In other words, the APC approach always selects the most efficient transmission manner for intra-cell traffic based on the channel conditions among the BS and SSs. It is noted that the channel conditions can be obtained via calculating the SINR value periodically. Consequently, it can be expected that the network throughput is enhanced while the proposed APC approach is exploited.
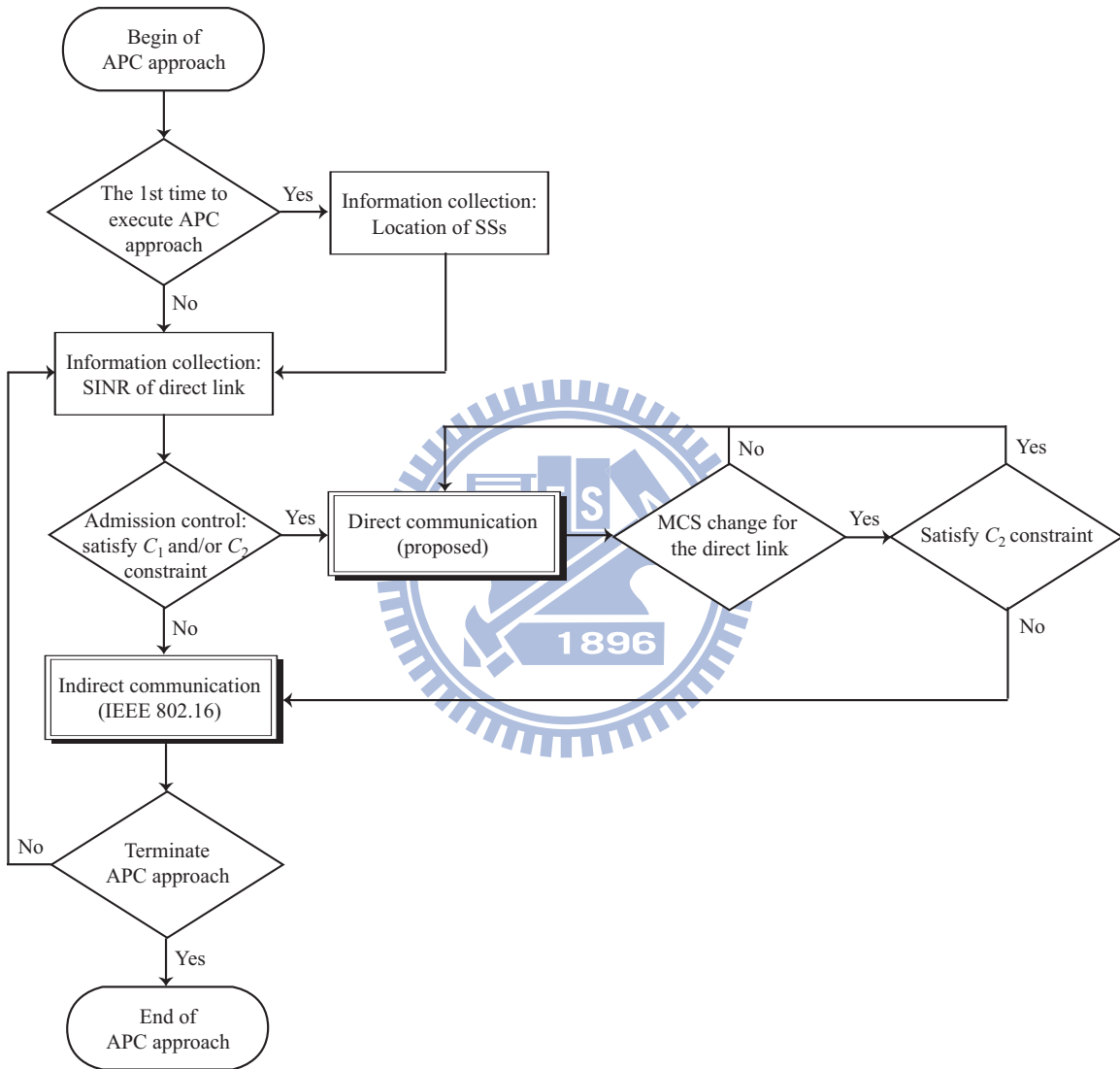
Figure 2.6: Flow diagram of APC approach.

2.4　Numerical Analysis

In this section, an analytical study of the IEEE 802.16 PMP OFDM frame structure with TDD mode is conducted. This study aims at analyzing the saturation user throughput as well as the corresponding overhead that can be achieved in IEEE 802.16 PMP networks. The user throughput is defined as the data bits per second which are received by the destination station in the considered network. Both the conventional packet transmission mechanism and the proposed APC approach are analyzed while taking into account both the MAC and PHY overheads. Since the saturation user throughput of the IEEE 802.16 network is the primary concern in this study, only one BS and two SSs are considered as shown in Fig. 2.2(a). It is assumed that both inter-cell traffic and intra-cell traffic flows exist in the considered network. Each inter-cell traffic flow is attached to either a UL or a DL connection; while the intra-cell traffic flow is resulted from the combination of both a UL and a DL connections. Therefore the number of inter-cell connections for a specific SS_k can be represented as

$$n_k = n_k^d + n_k^u,$$ (2.12)

where n_k^d and n_k^u are the numbers of inter-cell connections for SS_k in the DL and UL directions respectively. On the other hand, the number of intra-cell connections for SS_k can be represented as

$$m_k = m_k^d + m_k^u + \overline{m}_k^d + \overline{m}_k^u,$$ (2.13)

where m_k^d and m_k^u denote the numbers of one-hop intra-cell connections in the DL and UL directions for SS_k respectively. The term one-hop indicates that BS is assigned as either the source or the destination station for the intra-cell connection. On the other hand, the numbers of two-hop intra-cell connections in the DL and UL directions for SS_k are represented as $\overline{m}_k^d$ and $\overline{m}_k^u$. In other words, both the source and destination stations are designated to SSs; while SS_k belongs to one of these two SSs.

2.4.1 Conventional Mechanism

In this subsection, both the saturation user throughput and the corresponding overhead for the conventional mechanism of the IEEE 802.16 network will be derived. Considering that S_f represents one frame duration in the unit of OFDM symbols, the total number of available OFDM symbols S_{av} within a frame can be obtained as

$$S_{av} = S_f - S_{ttg} - S_{rtg},\qquad(2.14)$$

where S_{ttg} and S_{rtg} denote the durations of TTG and RTG in OFDM symbols respectively. Without loss of generality, it is assumed that a frame duration is evenly partitioned by the DL and UL subframes for analytical convenience. Different durations of DL and UL subframes can also be analyzed in similar manner. However, the ratio of DL subframe to UL subframe should be a constant value for all the BSs in an IEEE 802.16 network. If the ratio is different for each BS, significant inter-cell interference will appear in the entire network. Therefore, the durations of both the DL subframe (S_f^d) and the UL subframe (S_f^u) in OFDM symbols can be represented as

$$S_f^d = \left\lfloor \frac{S_{av}}{2} \right\rfloor,\qquad(2.15)$$

$$S_f^u = S_{av} - S_f^d.\qquad(2.16)$$

In order to obtain the user throughput, both the MAC and PHY overheads should be removed from the available raw bandwidth. According to the assumptions as mentioned above, the MAC and PHY overheads for a DL subframe in the conventional mechanism can be computed as

$$S_{oh}^d = S_{lp} + S_{fch} + \left\lceil \frac{2B_{gmh} + 2B_{crc} + B_{map}^d + B_{ie}^d + B_{map}^u + 5B_{ie}^u}{C_{map}} \right\rceil,\qquad(2.17)$$

where S_{lp} and S_{fch} are the durations of long preamble and FCH in OFDM symbols, which are regarded as the PHY overhead. The notations B_{gmh} and B_{crc} represent the size of generic

MAC header and CRC in bytes; while B_{map}^d, B_{map}^u, B_{ie}^d, and B_{ie}^u are the size of DL-MAP, UL-MAP, DL-MAP IE, and UL-MAP IE in bytes respectively. It is noted that all these parameters are designated as the MAC overhead. Moreover, the bytes per OFDM symbol for transmitting these MAP messages is denoted as C_{map}, whose value depends on the selection of the MCS. Since only two SSs are considered in the analytical model, three DL bursts are described in the FCH, i.e., for broadcasting MAP messages and transmitting data packets to SS_1 and SS_2 individually. The DL-MAP message contains only an end of map IE; while the UL-MAP message associated with UL-MAP IEs describe five bursts, including initial ranging, bandwidth request, two data grants for SS_1 an SS_2 individually, and end of map IE. Furthermore, both the DL-MAP and UL-MAP messages are individually attached with a generic MAC header and a CRC as denoted in (2.17). On the other hand, the MAC and PHY overhead for a UL subframe consists of two contention intervals (i.e., the initial ranging and the bandwidth request) and the short preambles. Since there are two SSs in considered, two short preambles will be transmitted in total by both SS_1 and SS_2. Therefore, the overhead for a UL subframe can be obtained as

$$S_{oh}^u = S_{ci} + 2S_{sp}, \tag{2.18}$$

where S_{ci} and S_{sp} are the durations of contention intervals and a single short preamble in OFDM symbols respectively.

Recall that the main objective of this study is to determine the saturation user throughput and the corresponding overhead of IEEE 802.16 PMP networks. The focus is to compute the maximum number of MAC PDUs that can be transmitted and received by a station within a frame respectively. It is assumed that there is always a packet to be transmitted in each the connection for any of the SS. Moreover, based on connection-oriented feature of the IEEE 802.16 MAC protocol, the resource scheduling within the BS is implemented on a per connection basis. Therefore, each connection is designed to receive a fair share of service. In other words, the available bandwidth for an SS in a frame depends on the number of connections it possesses. Based on (2.12), (2.13), (2.15) and (2.17), the maximum number of

MAC PDUs that can be transmitted by the BS to SS_k during a DL subframe is computed as

$$\varphi_k^d(B_{pkt}, C_k) = \left\lfloor \frac{(S_f^d - S_{oh}^d) \cdot \frac{n_k^d + m_k^d + \overline{m}_k^d}{\sum_{i=1}^2 (n_i^d + m_i^d + \overline{m}_i^d)} \cdot C_k}{B_{gmh} + B_{pkt} + B_{crc}} \right\rfloor, \tag{2.19}$$

where B_{pkt} represents the mean size of MAC service data unit (SDU) in bytes; while C_k is the bytes per OFDM symbol for SS_k to transmit packets. Similarly, the maximum number of MAC PDUs that is transmitted by SS_k to the BS during a UL subframe can be obtained as

$$\varphi_k^u(B_{pkt}, C_k) = \left\lfloor \frac{(S_f^u - S_{oh}^u) \cdot \frac{n_k^u + m_k^u + \overline{m}_k^u}{\sum_{i=1}^2 (n_i^u + m_i^u + \overline{m}_i^u)} \cdot C_k}{B_{gmh} + B_{pkt} + B_{crc}} \right\rfloor. \tag{2.20}$$

According to the definition of the user throughput, only the packets that are received by the destination stations should be considered. The maximum number of effective MAC PDUs for SS_k in a DL subframe will still be the same as (2.19), i.e.,

$$\varepsilon_k^d(B_{pkt}, C_k) = \varphi_k^d(B_{pkt}, C_k). \tag{2.21}$$

However, the maximum number of effective MAC PDUs for SS_k in a UL subframe becomes

$$\varepsilon_k^u(B_{pkt}, C_k) = \left\lfloor \frac{(S_f^u - S_{oh}^u) \cdot \frac{n_k^u + m_k^u}{\sum_{i=1}^2 (n_i^u + m_i^u + \overline{m}_i^u)} \cdot C_k}{B_{gmh} + B_{pkt} + B_{crc}} \right\rfloor. \tag{2.22}$$

It is noted that the parameter $\overline{m}_k^u$ as appeared in (2.20) is not counted since it denotes the number of two-hop intra-cell connections in the UL direction, which should not be considered in the computation of effective throughput in (2.22). Based on (2.21) and (2.22), the user throughput of SS_k in the DL and UL subframes, respectively, can be derived as

$$T_k^d(B_{pkt}, C_k) = \frac{\varepsilon_k^d(B_{pkt}, C_k) \cdot 8 B_{pkt}}{L_f}, \tag{2.23}$$

$$T_k^u(B_{pkt}, C_k) = \frac{\varepsilon_k^u(B_{pkt}, C_k) \cdot 8 B_{pkt}}{L_f}, \tag{2.24}$$

where L_f denotes the duration of a frame in seconds. Combining (2.23) and (2.24), the saturation user throughput that is achieved in the IEEE 802.16 PMP network can therefore be computed as the sum of the user throughput of each SS, i.e.,

$$T_{conv}^{max}(B_{pkt}, C_k) = \sum_{k=1}^{2} \left[T_k^d(B_{pkt}, C_k) + T_k^u(B_{pkt}, C_k) \right].$$ (2.25)

Furthermore, the corresponding overhead in terms of time per frame can be derived from (2.17) and (2.18) as

$$\begin{aligned} O_{conv}^{max}(B_{pkt}, C_k) &= (S_{ttg} + S_{rtg} + S_{oh}^d + S_{oh}^u) \cdot L_s \\ &+ \sum_{k=1}^{2} \frac{[\varphi_k^u(B_{pkt}, C_k) - \varepsilon_k^u(B_{pkt}, C_k)] \cdot (B_{gmh} + B_{pkt} + B_{crc})}{C_k} \cdot L_s \\ &+ \sum_{k=1}^{2} \frac{[\varepsilon_k^d(B_{pkt}, C_k) + \varepsilon_k^u(B_{pkt}, C_k)] \cdot (B_{gmh} + B_{crc})}{C_k} \cdot L_s, \end{aligned}$$ (2.26)

where L_s is the duration of an OFDM symbol in seconds. It is noted that the second term in (2.26) represents the overhead caused by the duplication of the data packets; while the MAC overhead of the received data packets is specified in the last term.

2.4.2 APC Approach

In the proposed APC approach, the BS will arrange specific time interval for the SSs to conduct direct communication. Since the saturation user throughput of IEEE 802.16 PMP network is the main concern of this study, it is assumed that all the direct connections are scheduled at the end of the UL subframe regardless of any specific scheduling algorithm. Therefore, the MAC and PHY overhead for a DL subframe by adopting the APC approach can be computed as

$$\mathbb{S}_{oh}^d = S_{lp} + S_{fch} + \left\lceil \frac{2B_{gmh} + 2B_{crc} + B_{map}^d + B_{ie}^d + B_{map}^u + 7B_{ie}^u + 2B_{ie}^{updl}}{C_{map}} \right\rceil,$$ (2.27)

where B_{ie}^{updl} represents the size of DL-PDL IE in bytes. It is noted that since the UL-PDL IE is an extended IE within the UL-MAP IE, additional two UL-MAP IEs and two UL-PDL IEs are required to arrange the direct transmission between the SSs. Furthermore, the overhead for a UL subframe with the APC approach becomes

$$\mathbb{S}_{oh}^u = S_{ci} + 4S_{sp}. \tag{2.28}$$

It is noted that an additional short preamble is required for the PHY synchronization of SS_1 and SS_2 respectively. Recall that each connection receives a fair share of service is assumed in this study. With (2.14), (2.27), and (2.28), the duration of PDL subframe in OFDM symbols can be obtained as

$$\mathbb{S}_{pdl} = (S_{av} - \mathbb{S}_{oh}^d - \mathbb{S}_{oh}^u) \cdot \frac{\sum_{i=1}^2 \overline{m}_i^u}{\sum_{i=1}^2 (n_i + m_i - \overline{m}_i^d)}. \tag{2.29}$$

Since only two SSs are considered in this study, the number of two-hop intra-cell connections in the UL direction for SS_1 equals to that in the DL direction for SS_2 and vice versa, i.e., $\overline{m}_1^u = \overline{m}_2^d$ and $\overline{m}_2^u = \overline{m}_1^d$. Therefore, the numbers of direct connections and total connections in the considered network are $\sum_{i=1}^2 \overline{m}_i^u$ and $\sum_{i=1}^2 (n_i + m_i - \overline{m}_i^d)$ respectively. The number of remaining OFDM symbols in a frame by exploiting the APC approach becomes

$$\mathbb{S}_{av} = S_f - S_{ttg} - S_{rtg} - \mathbb{S}_{pdl}. \tag{2.30}$$

Similar to (2.16), the duration of DL subframe ($\mathbb{S}_f^d$) and UL subframe ($\mathbb{S}_f^u$) in OFDM symbols by adopting the APC approach can be represented as

$$\mathbb{S}_f^d = \left\lfloor \frac{\mathbb{S}_{av}}{2} \right\rfloor, \tag{2.31}$$

$$\mathbb{S}_f^u = \mathbb{S}_{av} - \mathbb{S}_f^u. \tag{2.32}$$

31

Consequently, the maximum number of MAC PDUs that can be transmitted by the BS to SS_k during a DL subframe can be computed as

$$\vartheta_k^d(B_{pkt}, C_k) = \left\lfloor \frac{(\mathbb{S}_f^d - \mathbb{S}_{oh}^d) \cdot \frac{n_k^d + m_k^d}{\sum_{i=1}^{2}(n_i^d + m_i^d)} \cdot C_k}{B_{gmh} + B_{pkt} + B_{crc}} \right\rfloor. \tag{2.33}$$

It is noted that the number of two-hop intra-cell connections in the DL direction $\overline{m}_k^d$ as appeared in (2.19) is not considered in (2.33) due to the adoption of the proposed APC approach. On the other hand, the maximum number of MAC PDUs that are delivered by SS_k during a UL subframe is acquired as

$$\vartheta_k^u(B_{pkt}, C_k) = \left\lfloor \frac{(\mathbb{S}_f^u - \mathbb{S}_{oh}^u) \cdot \frac{n_k^u + m_k^u}{\sum_{i=1}^{2}(n_i^u + m_i^u)} \cdot C_k}{B_{gmh} + B_{pkt} + B_{crc}} \right\rfloor + \left\lfloor \frac{\mathbb{S}_{pdl} \cdot \frac{\overline{m}_k^u}{\overline{m}_1^u + \overline{m}_2^u} \cdot C_{pdl}}{B_{gmh} + B_{pkt} + B_{crc}} \right\rfloor, \tag{2.34}$$

where C_{pdl} represents the bytes per OFDM symbol for direct transmission. It is noted that the second term in (2.34) specifies the maximum number of MAC PDUs that are transmitted by SS_k within the direct connections. Furthermore, the maximum number of effective MAC PDUs for SS_k in a DL subframe is equal to (2.33); while that in a UL subframe will be the same as (2.34). Based on (2.33) and (2.34), the user throughput of SS_k in the DL and UL subframes, respectively, can be obtained as

$$\mathbb{T}_k^d(B_{pkt}, C_k) = \frac{\vartheta_k^d(B_{pkt}, C_k) \cdot 8 B_{pkt}}{L_f}, \tag{2.35}$$

$$\mathbb{T}_k^u(B_{pkt}, C_k) = \frac{\vartheta_k^u(B_{pkt}, C_k) \cdot 8 B_{pkt}}{L_f}. \tag{2.36}$$

As a result, the saturation user throughput by adopting the APC approach can be acquired as

$$\mathbb{T}_{apc}^{max}(B_{pkt}, C_k) = \sum_{k=1}^{2} \left[\mathbb{T}_k^d(B_{pkt}, C_k) + \mathbb{T}_k^u(B_{pkt}, C_k) \right]. \tag{2.37}$$

Similarly, the corresponding overheads in terms of time per frame is obtained as

$$\mathbb{O}_{apc}^{max}(B_{pkt}, C_k) = (S_{ttg} + S_{rtg} + \mathbb{S}_{oh}^d + \mathbb{S}_{oh}^u) \cdot L_s$$
$$+ \sum_{k=1}^{2} \frac{[\vartheta_k^d(B_{pkt}, C_k) + \vartheta_k^u(B_{pkt}, C_k)] \cdot (B_{gmh} + B_{crc})}{C_k} \cdot L_s. \qquad (2.38)$$

2.5 Performance Evaluation

In this section, the analytical results obtained from the previous section will be validated. Furthermore, extensive simulations are performed to evaluate the performance of the proposed APC approach in comparison with the conventional packet transmission mechanism in IEEE 802.16 PMP networks. It is noted that simulations are carried out by an event-driven IEEE 802.16 MAC simulator written in MATLAB. All the required procedures and functions for uplink/downlink data transmission and uplink bandwidth request/grant have been implemented in the simulator.

2.5.1 Validation of Analytical Results

Simulations are performed in order to validate the analytical models as derived in the previous section. One BS along with two SSs as shown in Fig. 2.2(a) are considered in the network, where 20 traffic flows are assumed. Since the saturation user throughput is the main concern in the derivation of the analytical models, the packet buffer for each connection is designed to be nonempty during the entire simulation time. Furthermore, same MCS is utilized within both the DL and UL transmission for each SS.

Fig. 2.7 illustrates both the simulation and analytical results of saturation user throughput over various percentage of intra-cell traffic flows for the proposed APC approach and the conventional IEEE 802.16 packet transmission scheme. It can be observed that the same highest saturation user throughput are reached in both schemes while all the flows belong to the inter-cell traffic. The reason is attributed to the definition of user throughput, wherein the BS is treated as either a source or a destination station for an inter-cell traffic flow. By adopting the conventional mechanism, it can be expected that the saturation user throughput
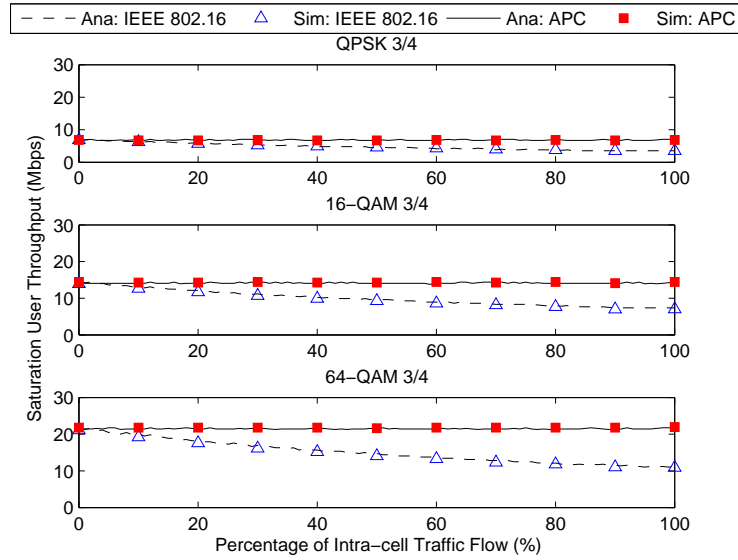
Figure 2.7: Validation of analytical results: saturation user throughput versus percentage of intra-cell traffic flows.

decreases as the percentage of intra-cell traffic flows is augmented because the packets are required to be forwarded via the BS. On the other hand, by exploiting the proposed APC approach, the packets are directly transmitted from the source to the destination, which results in the same level of saturation user throughput.

The performance of the corresponding overhead for the compared schemes are shown in Fig. 2.8. Owing to the duplication of forwarding packets via the the BS, it can be expected that the overhead from the conventional scheme increases as the percentage of intra-cell traffic flows is augmented. Nevertheless, the overhead resulted from the proposed APC approach remains at the same level under all the scenarios since the packets are directly transmitted from the source to the destination.

2.5.2 Performance Comparison

In order to evaluate the effectiveness of the proposed APC approach, a more realistic simulation setup is performed in this subsection. A 19 cell-based wrap around topology [18] is considered as the simulation layout. Each cell consists of a centered BS and 20 uniformly distributed SSs. The mandatory path loss model defined as $PL(dB) = 130.19 + 37.6 \log_{10}(R/1000)$ for
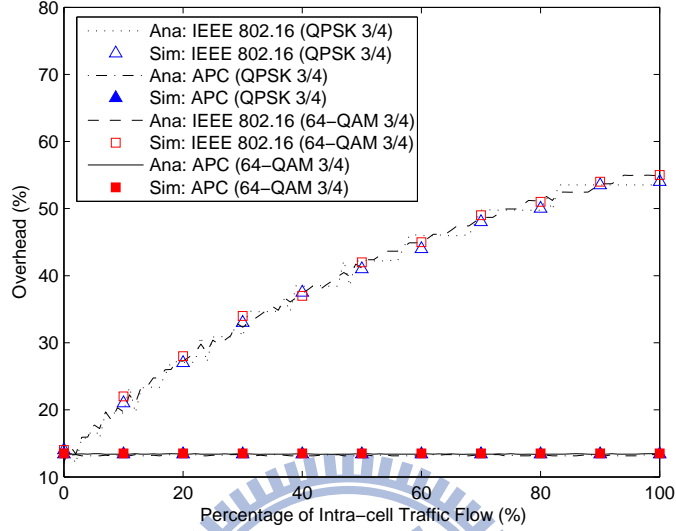
Figure 2.8: Validation of analytical results: overhead versus percentage of intra-cell traffic flows.

Table 2.8: Simulation parameters for APC approach

Parameter	Value
OFDM symbol duration	34 μs
Maps modulation	BPSK
Data modulation	QPSK, 16-QAM, 64-QAM
Frame duration	10 ms
SSTTG/SSRTG	1 OFDM symbol
Initial ranging interval	5 OFDM symbols
Bandwidth request interval	5 OFDM symbols
Simulation time	1 minute

IEEE 802.16 system [18] is adopted to reflect the channel conditions, where R is the distance between transmitter and receiver in unit of meter. Based on this path loss model, the maximum value of R for each MCS is listed in the last column of Table ??. It is noted that all of these values are calculated without the consideration of interference caused by other transmissions. In the simulation, both the intra-cell and inter-cell traffic flows are considered; while the source and destination of each flow are randomly selected. The packet arrival process of each flow is assumed to follow a Poisson process with rate λ in unit of packet per frame. The size of each packet is selected to follow the exponential distribution with the mean value of 200 bytes. Since the scheduling algorithm is not specified in the IEEE 802.16 standard, the deficit round robin (DRR) [19] and weighted round robin (WRR) [20] algorithms are selected as the BS's DL and UL schedulers respectively. The DRR algorithm is also utilized by the SS to share the UL grants that are provided by the BS for its corresponding UL connections. Table 2.8 lists the parameters that are adopted within the simulation. Figs. 2.9 to 2.11 show the performance comparisons among the proposed APC approach with the TO constraint (denoted as APC-TO) and EO constraint (denoted as APC-EO) respectively, and the conventional scheme (denoted as IEEE 802.16). Furthermore, the comparisons of direct communication conducted in either the DL subframe or the UL subframe are also shown in these figures.

Fig. 2.9 illustrates the performance of user throughput, overhead, and latency under various traffic loads λ for each compared schemes. In this comparison, 200 intra-cell traffic flows with the same λ are considered. It is intuitive that the user throughput, overhead, and latency increase as the traffic load is augmented for all the schemes. Due to the effect of direct communication conducted among SSs, the performance of proposed APC approaches outperform that of the conventional scheme under various traffic loads. Noted that the latency for the conventional mechanism consists of queuing and transmission latency; while additional latency for direct link setup is considered in the APC approaches. It is observed that comparably lower latency can be obtained via the proposed APC approaches than the conventional scheme since the direct link setup is conducted along with the original data transmission.
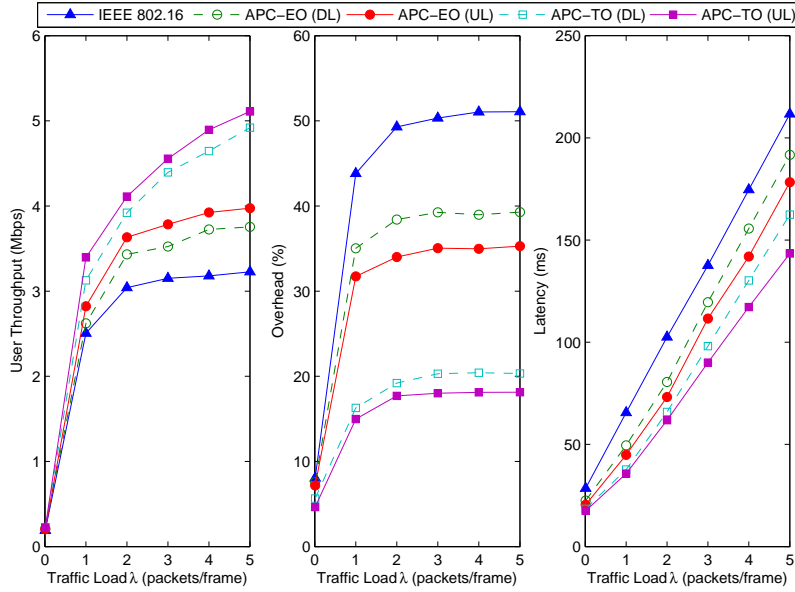
Figure 2.9: Performance comparison of user throughput, overhead, and latency versus traffic load (λ).

Comparing between the APC-TO and APC-EO mechanisms, relative better performance is observed in the APC-TO scheme. This can be attributed to the reason that the EO constraint is considered more stringent compared to the TO constraint in terms of the criterion for direct transmission. Therefore, less pairs of SSs are allowed to conduct the direct communication while the APC-EO approach is exploited, which results in relative lower performance in comparison with the APC-TO scheme. On the other hand, for each APC approach, it is observed that conducting direct transmissions in UL subframes has better performance than that in DL subframes. The reason can be contributed to the potential inter-cell interferences to neighbor-cell SSs that are introduced by conducting direct communication in DL subframes, which can consequently decrease the entire system performance. If the direct transmissions are arranged in UL subframes, the conventional communication pairs will not be interfered since all the SSs are served as transmitters during those UL subframes.

Performance comparisons with an increasing number of intra-cell traffic ranging from 0 to 400 is shown in Fig. 2.10. The packet arrival process of each flow is assumed to follow a Poisson process with rate $\lambda = 2$ in unit of packet per frame. As can be expected that the
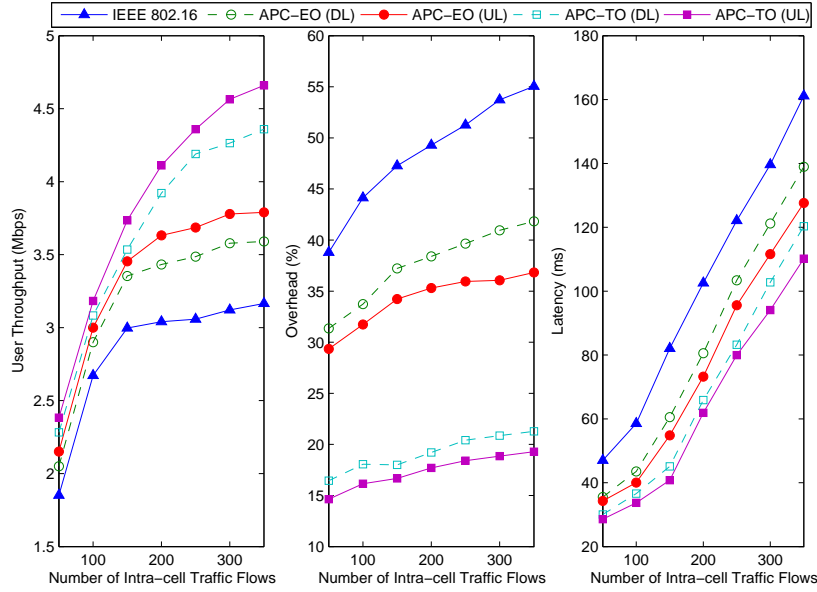
Figure 2.10: Performance comparison of user throughput, overhead, and latency versus number of intra-cell traffic flows (with $\lambda = 2$).

user throughput, overhead, and latency increase as the number of intra-cell traffic flows is enlarged. The performance of proposed APC approaches outperform that of the conventional scheme. In the conventional scheme, the intra-cell traffic is required to be forwarded by the BS. Consequently, more than twice of the communication bandwidth is necessitate for the packet transmission. By adopting the proposed APC approaches, the intra-cell traffic can be directly transmitted from the source to the destination, which resulted in conserved bandwidth. Consequently, relative higher user throughput, lower overhead, and lower latency are acquired in the APC approaches. Similarly, the APC-TO approach outperforms the APC-EO scheme; meanwhile, arrangement of direct communication in UL subframes outperforms that in DL subframes for each APC approach.

Fig. 2.11 depicts the influence from the percentage of inter-cell traffic flows to the performance of these three schemes. In this comparison, 200 traffic flows with the same packet arrival rate $\lambda = 2$ are considered. It can be seen that the user throughput decreases as the percentage of intra-cell traffic flows is increased for all the schemes, wherein the conventional scheme has the worst throughput performance. This can be attributed to the reason that the
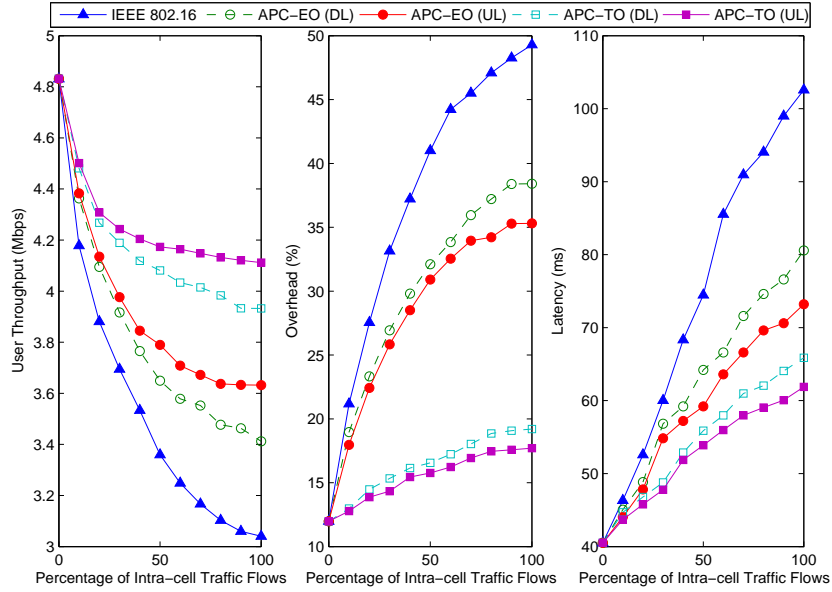
Figure 2.11: Performance comparison of user throughput, overhead, and latency versus percentage of intra-cell traffic flows (with $\lambda = 2$).

intra-cell traffic is required to be forwarded by the BS in the conventional scheme. However, there can be opportunities for the proposed APC approaches to conduct direct communication among SSs. On the other hand, due to the same reason, the overhead and the latency for all the schemes increase as the percentage of intra-cell traffic flows is enlarged. Similarly, the APC-TO approach outperforms the APC-EO scheme, and relative better performance is resulted from direct communication that are conducted in UL subframe comparing with that in DL subframes for each APC scheme. The merits of proposed APC approach can therefore be observed.

2.6 Concluding Remarks

In this work, a flexible and contention-free adaptive point-to-point communication (APC) approach is proposed. A comprehensive architectural design associated with the extended frame structure for the proposed APC approach is developed to be backward-compatible with the IEEE 802.16 standard. The proposed scheme arranges specific time periods from the BS

to achieve point-to-point direct communication among the SSs while they are adjacent with each others. According to the relative locations and channel conditions among the BS and SSs, the packet transmission operation is switched between the direct communication and indirect communication. Based on both the analytical and simulation results, it is observed that the proposed APC approach can outperform the conventional IEEE 802.16 transmission mechanism under different network scenarios.
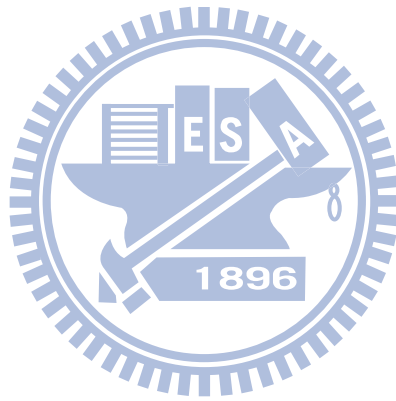
Chapter 3

Predictive Motion and Interference-based Scheduling Algorithms for Direct Communication of Mobile Stations
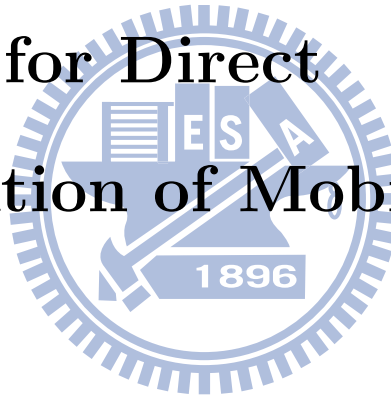
3.1   Introduction

In previous chapter, an adaptive communication approach for a pair of subscriber stations (SSs) that are actively involved in packet transmission is introduced. According to the relative locations and channel conditions among the base station (BS) and SSs, the communication operation for the pair is switched automatically between the direct manner and indirect manner, which results in the reduction of the communication bandwidth, control overhead, and packet latency. However, the issues of inter-cell interference and node mobility were not considered in the approach.

The inter-cell interference results from the inappropriate arrangement of direct communication. For example, in a downlink (DL) subframe of IEEE 802.16 networks, all the SSs are served as receivers with the BS served as the transmitter. In the case that a pair of SSs

41

proceeds with direct communication in the DL subframe, the neighbor-cell SSs that are close to the considered pair will suffer from additional inter-cell interference resulted from the direct communication. Consequently, the network performance can be severely degraded. On the other hand, the location information of SSs are considered as one of the necessary components to conduct direct communication. Since the SSs may move around in the networks, how to efficiently acquire the positions of SSs becomes an important issue.

In this work, a scheduling algorithm is proposed for each pair of mobile stations (MSs) that are expected to conduct direct communication. The interference region and feasible region for the pair of MSs to conduct direct transmission are studied and calculated. Based on these two types of information, the predictive interference-based scheduling (PIS) algorithm is presented to properly arrange the feasible opportunities (i.e., in either DL subframes or UL subframes) for conducting direct communication. Furthermore, with the consideration of mobility of MSs, a motion prediction mechanism for the MS is proposed. Combining the predication mechanism with the PIS algorithm, a predictive motion and interference-based scheduling (PMIS) algorithm is given to reduce the control overhead regarding the updates of MSs positions. The efficiency of the proposed PIS and PMIS algorithms are evaluated and compared via simulations. In terms of user throughput, simulation results show that the proposed PIS and PMIS algorithm outperforms original adaptive point-to-point (APC) approach. Furthermore, the control overhead resulted from the PIS algorithm is significantly reduced by the PMIS approach.

The remainder of this chapter is organized as follows. Section 3.2 briefly describes the considered signal propagation model and mobility model, associated with the formulation of the target problem. The interference region and feasible region for a direct communication pair are explained in Section 3.3. Section 3.4 describes the proposed PIS algorithm; while the PMIS algorithm is given in Section 3.5 by combining a motion predication mechanism with the PIS algorithm. Section 3.7 draws the conclusions.

3.2 Preliminaries

3.2.1 Signal Propagation Model

In this work, the effects of both path loss and shadowing are considered to characterize signal propagation. The path loss is caused by dissipation of the power radiated by the transmitter as well as the effects of propagation channel; while the attenuation of signal power resulted from obstacles between the transmitter and receiver is reflected by the shadowing. Since it is difficult to obtain a signal model that characterizes path loss accurately across a range of different environments, the considered path loss model with a simplified formulation is defined as [21]

$$P_L = K + 10\tau \log_{10}(\frac{d}{d_0}) + \psi \tag{3.1}$$

with

$$K = 20 \log_{10}(\frac{4\pi d_0}{\lambda}), \tag{3.2}$$

where K is an intercept which is the free-space path loss at reference distance d_0, and λ is the wavelength in meters. The variable τ is the path loss exponent; while d is the distance between the transmitter and receiver. The shadowing component is denoted as ψ, which is a zero-mean Gauss-distributed random variable with variance σ_ψ^2.

3.2.2 Mobility Model

The Gauss-Markov mobility (GMM) model [22] is considered in this work to represent the motion of each MS. Comparing with the random walk mobility model [23] that results in sharp turns and sudden stops, a more realistic motion trajectory can be obtained with the GMM model. Let α_k and V_k denote the moving direction (with respect to the positive x-axis) and velocity of an MS at a discrete time instant t_k respectively. Based on the GMM mode,
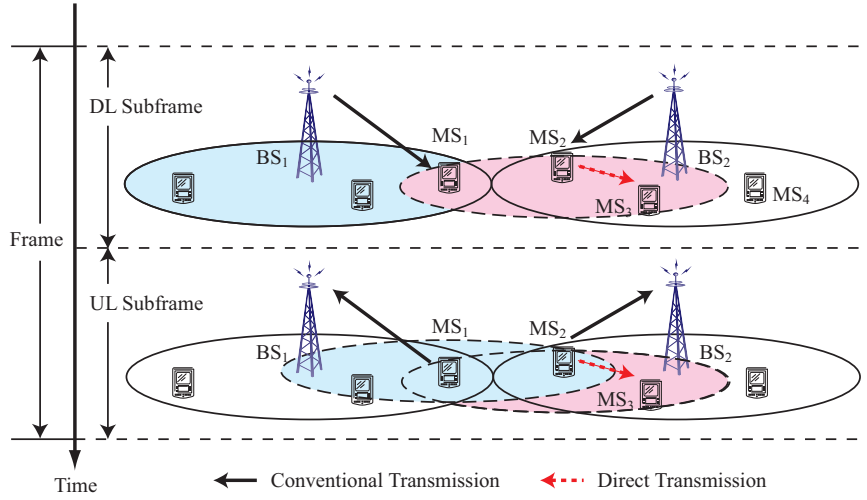
Figure 3.1: Schematic diagram of transmission scheduling for IEEE 802.16 TDD-based PMP networks.

both the moving direction and velocity can be formulated as [24]

$$\alpha_k = \gamma_{k-1}\alpha_{k-1} + (1 - \gamma_{k-1})\bar{\alpha} + \sqrt{(1 - \gamma_{k-1}^2)}\Upsilon_{\alpha_{k-1}} \tag{3.3}$$

and

$$V_k = \varrho_{k-1}V_{k-1} + (1 - \varrho_{k-1})\bar{V} + \sqrt{(1 - \varrho_{k-1}^2)}\Upsilon_{V_{k-1}}, \tag{3.4}$$

where $\bar{\alpha}$ and $\bar{V}$ represent the asymptotic mean of the moving direction and velocity as $t_k \to \infty$. The notations $\Upsilon_{\alpha_{k-1}}$ and $\Upsilon_{V_{k-1}}$ are zero-mean Gaussian-distributed random variables; while γ_{k-1} and ϱ_{k-1} are time-varying parameters that represent different levels of randomness as $0 \leq \gamma_{k-1}, \varrho_{k-1} \leq 1$. Two extreme cases correspond to the linear motion (as $\gamma_{k-1} = \varrho_{k-1} = 1$) and Brownian motion (as $\gamma_{k-1} = \varrho_{k-1} = 0$).

3.2.3 Problem Statement

In an IEEE 802.16 point-to-multipoint (PMP) network, the BS is responsible for controlling the communication with multiple MSs in both DL and UL directions. Adopting the time division duplexing (TDD) technique, time domain is divided into successive frame durations.

Each frame consist of a DL subframe and a UL subframe. In DL subframe, the BS is served as the only transmitter within each cell; while all the attached MSs are receivers. On the other hand, in UL subframes, packet transmission is conduced by the MSs to the BS. An exemplified network consisting of two neighboring cells are shown in Fig. 3.1. In the DL subframe, BS_1 and BS_2 are the transmitters in the network. With the consideration of appropriate power allocation for each BS, MS_2 will successfully receive the packets transmitted from BS_2 without severely interference caused by BS_1. On the other hand, in UL the subframe, a pre-defined duration is scheduled for an MS communicate with its serving BS. In the case that MS_1 and MS_2 are scheduled within the same duration, MS_2 will not suffer from the interference caused by MS_1 since both of them are transmitters. However, the interference-free situation will not be guaranteed while the direct communication is available between MSs. For example, as shown in Fig. 3.1, if MS_2 directly communicates with MS_3 in the UL subframe, the interference will be obtained by MS_1 since it is located within MS_2's transmission range. Nevertheless, no additional interference will be introduced if the direct communication is conducted in the DL subframe.

Based on these observations, it is motivated that a proper scheduling algorithm for direct communication pairs should be provided to enhance the performance of IEEE 802.16 TDD-based PMP networks. The direct communication scheduling (DCS) problem is formally described as follows:

Problem 2 (Direct Communication Scheduling Problem). *Given a pair of MSs that are actively involved in packet transmission, how to arrange a proper time duration for the pair to conduct direct communication without intervening the communication of the other MSs that are located within its transmission range?*

3.3   Feasible Region Analysis

In this section, the feasible region for a direct communication pair is derived from the analyzed interference region. The definition of both interference region and feasible region are described as follows:

Definition 1 (Interference Region). *Given a communication pair $\mathbf{L} = \{N_T, N_R\}$, where N_T and N_R denote the transmitter and receiver respectively, the interference region for the pair $\mathbf{L}$ is defined as a region $\mathbf{R_{IR}^L}$ within which N_R will be interfered by an unrelated transmitter $\tilde{N}_T$.*

Definition 2 (Feasible Region). *Given a communication pair $\mathbf{L} = \{N_T, N_R\}$, where N_T and N_R denote the transmitter and receiver respectively, the feasible region for the pair $\mathbf{L}$ is defined as a region $\mathbf{R_{FR}^L}$ within which the communication between N_T and N_R can be conducted successfully.*

Considering an on-going communication pair $\mathbf{L} = \{N_T, N_R\}$ and the distance between N_T and N_R is $D^{\mathbf{L}}$, the received signal power transmitted from N_T at N_R is denoted as $P_r^{\mathbf{L}}$, which can be derived from (3.1) and (3.2) as

$$P_r^{\mathbf{L}} = P_t^{\mathbf{L}} \left(\frac{\lambda}{4\pi d_0} \right)^2 \left(\frac{d_0}{D^{\mathbf{L}}} \right)^{\tau} 10^{\frac{-\psi^{\mathbf{L}}}{10}}, \tag{3.5}$$

where $P_t^{\mathbf{L}}$ is the transmitted power of the signal at N_T. Let $P_t^{\tilde{N}_T}$ represents the power of interference signal transmitted by $\tilde{N}_T$ and the distance between $\tilde{N}_T$ and N_R is D^I. Similarly, with (3.1) and (3.2), the received power of interference signal at N_R can be obtained as

$$P_r^I = P_t^{\tilde{N}_T} \left(\frac{\lambda}{4\pi d_0} \right)^2 \left(\frac{d_0}{D^I} \right)^{\tau} 10^{\frac{-\psi^I}{10}}. \tag{3.6}$$

The signal-to-interference ratio at N_R is given as $\Phi_{N_R} = P_r^{\mathbf{L}}/P_r^I$, which should be grater than a threshold $\mathfrak{T}_\Phi$. Based on (3.5) and (3.6), Φ_{N_R} can be acquired as

$$\Phi_{N_R} = \frac{P_r^{\mathbf{L}}}{P_r^I} = \frac{P_t^{\mathbf{L}}}{P_t^{\tilde{N}_T}} \left(\frac{D^I}{D^{\mathbf{L}}} \right)^{\tau} 10^{\frac{\psi^I - \psi^{\mathbf{L}}}{10}} \geq \mathfrak{T}_\Phi. \tag{3.7}$$

The relationship between D^I and $D^{\mathbf{L}}$ is acquired from (3.7) as

$$D^I \geq D^{\mathbf{L}} \sqrt[\tau]{\frac{\mathfrak{T}_\Phi}{\Theta}}, \tag{3.8}$$
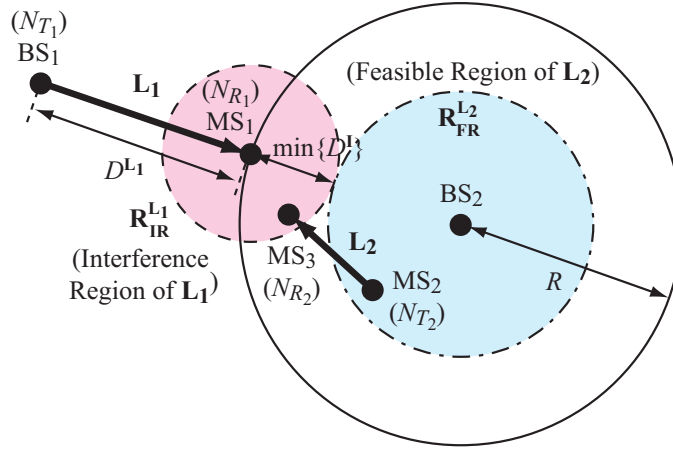
Figure 3.2: Schematic diagram of feasible region for direct communication pair $\mathbf{L_2} = \{N_{T_2}, N_{R_2}\}$ (i.e, MS_2 communicates with MS_3) in DL subframe.

where

$$\Theta = \frac{P_t^{\mathbf{L}}}{P_t^{\tilde{N}_T}} 10^{\frac{\psi^I - \psi^{\mathbf{L}}}{10}}. \tag{3.9}$$

Based on (3.8), the interference region for the pair $\mathbf{L} = \{N_T, N_R\}$ can be derived as a circular region centered at N_R with radius of $\min\{D^I\}$, i.e.,

$$\mathbf{R_{IR}^L} = \Gamma\left(N_R, D^{\mathbf{L}} \sqrt[\tau]{\frac{\mathfrak{T}_{\Phi}}{\Theta}}\right), \tag{3.10}$$

where $\Gamma(\cdot)$ is a function to calculate a circular area. In other words, if the transmitter $\tilde{N}_T$ exists within $\mathbf{R_{IR}^L}$, N_R will not successfully receive the signal transmitted from N_T.

3.3.1 Feasible Region in DL Subframe

In order to study the feasible region for a direct communication pair in a DL subframe, a simplified scenario is considered as shown in Fig. 3.2. It is assumed that $\mathbf{L_2} = \{N_{T_2} = MS_2, N_{R_2} = MS_3\}$ denotes a pair of MSs that are expected to conduct the direct communication served by BS_2, and a conventional communication pair $\mathbf{L_1} = \{N_{T_1} = BS_1, N_{R_1} = MS_1\}$ is served by the neighboring BS_1. It is noted that direct communication pairs should not intervene the conventional ones. Therefore, the distance between N_{T_1} and N_{R_1} (i.e., $D^{\mathbf{L_1}}$)

47
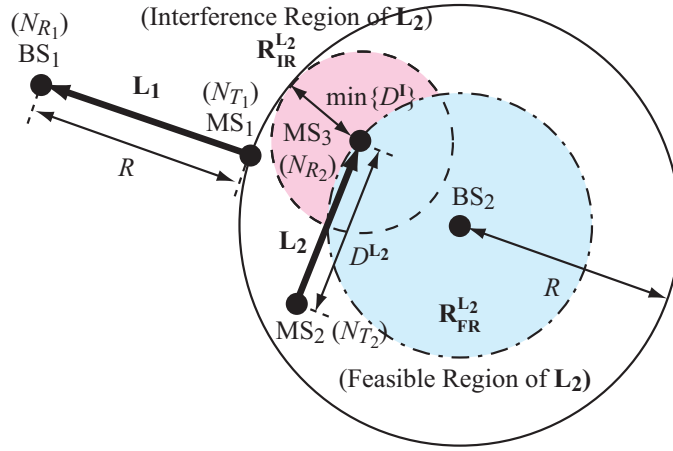
Figure 3.3: Schematic diagram of feasible region for direct communication pair $\mathbf{L}_2 = \{N_{T_2}, N_{R_2}\}$ (i.e, MS_2 communicate with MS_3) in UL subframe.

is assumed to be equal to the radius of the BS's transmission range R. In other words, the strictest case that MS_1 exists on the cell edge of BS_2 is considered. Intuitively, MS_2 should not exist within the interference region of $\mathbf{L}_1$, which can be obtained from (3.10) as

$$\mathbf{R}_{\mathbf{IR}}^{\mathbf{L_1}} = \Gamma\left(MS_2, R\sqrt[\tau]{\frac{\mathfrak{I}_\Phi}{\Theta}}\right). \tag{3.11}$$

As for MS_3, the receiver of $\mathbf{L}_2$, it can be located anywhere within the transmission range of MS_2. Therefore, for the scenario shown in Fig. 3.2, the feasible region of $\mathbf{L}_2$ is the service range of BS_2 except for the area overlapped by $\mathbf{R}_{\mathbf{IR}}^{\mathbf{L_1}}$.

Based on the aforementioned observations, the feasible region for a direct communication pair in a DL subframe can be inferred as follows:

Corollary 1. *Given a direct communication pair* $\mathbf{L} = \{N_T, N_R\}$ *served by a BS, the feasible region for* $\mathbf{L}$ *in a DL subframe (i.e.,* $\mathbf{R}_{\mathbf{FR,DL}}^{\mathbf{L}}$*) is a circular region centered at the BS with radius of* $R(1 - \sqrt[\tau]{\frac{\mathfrak{I}_\Phi}{\Theta}})$*, where* R *is the radius of BS's service range. If* N_T *is located within* $\mathbf{R}_{\mathbf{FR,DL}}^{\mathbf{L}}$*, the direct communication between* N_T *and* N_R *can be conducted successfully.*

3.3.2 Feasible Region in UL Subframe

It is similar to the case in the DL subframe, Fig. 3.3 illustrates a simplified scenario that consists of two communication pairs attached to two neighboring cells, respectively, in a UL subframe. As can be seen that $\mathbf{L}_1 = \{N_{T_1}, N_{R_1}\}$ is a conventional communication pair served by BS_1, where MS_1 is the transmitter and BS_1 is the corresponding receiver, i.e., $N_{T_1} = MS_1$ and $N_{R_1} = BS_1$. On the other hand, BS_2 serves a pair of MSs that are expected to conduct direct communication, i.e., $\mathbf{L}_2 = \{N_{T_1} = MS_2, N_{R_1} = MS_3\}$. Since MS_1 is a transmitter in the UL subframe, $\mathbf{L}_1$ will never be disturb by $\mathbf{L}_2$, but $\mathbf{L}_1$ may interfere with $\mathbf{L}_2$. In this case, the direct communication of $\mathbf{L}_2$ can be performed successfully if MS_1 located outside of the interference region of MS_3. Based on this observation, the feasible region for a direct communication pair in a UL subframe can be inferred as follows:

Corollary 2. *Given a direct communication pair $\mathbf{L} = \{N_T, N_R\}$ served by a BS, the feasible region for $\mathbf{L}$ in a UL subframe (i.e., $\mathbf{R^L_{FR,UL}}$) is a circular region centered at the BS with radius of $(R - D^{\mathbf{L}} \sqrt[\tau]{\frac{\tau_\Phi}{\Theta}})$, where R is the radius of BS's service range and $D^{\mathbf{L}}$ is the distance between N_T and N_R. If N_R is located within $\mathbf{R^L_{FR,UL}}$, the direct communication between N_T and N_R can be conducted successfully.*

It is worthwhile to mention that the feasible region for a direct communication pair is a fixed region in a DL subframe; while in the UL subframe, it is a variable region that is changed according to the distance between the transmitter and receiver of the pair.

3.4 Predictive Interference-based (PIS) Algorithm

In this section, the PIS algorithm is proposed to arrange a proper communication period for a pair of MSs that are expected to conduct direct communication. Based on the aforementioned feasible regions, communication of the pair will be arranged in either the DL subframe or the UL subframe, or be transferred to the conventional communication manner (i.e, via the BS). In the PIS algorithm, it is assumed that the accurate position information of the targeted MSs is known by the BS at each scheduling epoch. Let $\mathbf{L} = \{N_T, N_R\}$ denote the targeted
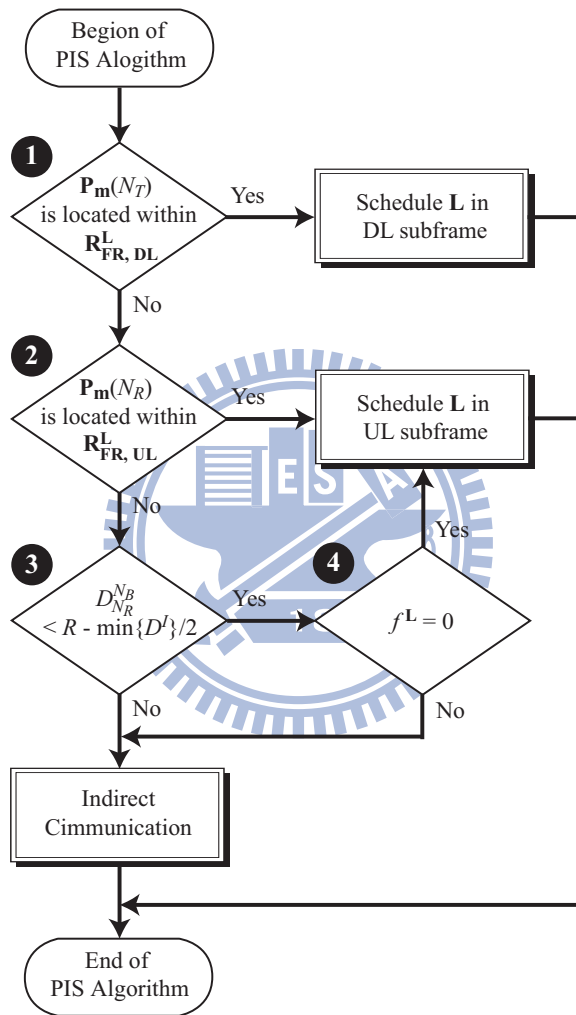
Figure 3.4: Flow diagram of PIS Algorithm.

direct communication pair, where N_T and N_R are MSs. Fig. 3.5 illustrates the flowchart of the PIS algorithm, which consists of four important steps. The detailed description of these steps are given in the following paragraphs.

Step 1: The availability of direct communication for **L** during the DL subframe is determined in this step. Let $\mathbf{P}_m^{N_T}$ represents the position of N_T at a scheduling epoch t_m. Based on Corollary 1, the DL feasible region of **L** at t_m is calculated as

$$\mathbf{R}_{\mathbf{FR,DL}}^{\mathbf{L}}(t_m) = \Gamma\left(\mathbf{P}^{N_B}, R(1 - \sqrt[\gamma]{\frac{\mathfrak{T}_\Phi}{\Theta}})\right), \tag{3.12}$$

where $\Gamma(x, y)$ is a function to calculate a circular area centered at x with radius of y; while $\mathbf{P}^{N_B}$ and R denotes the position of the BS (N_B) and its service range, respectively. If $\mathbf{P}_m^{N_T}$ is covered by $\mathbf{R}_{\mathbf{FR,DL}}^{\mathbf{L}}(t_m)$, **L** will be scheduled in the DL subframe and the scheduling will be completed; otherwise, it will go to Step 2.

Step 2: This step is utilized to determine if the direct communication for **L** can be conducted in the UL subframe. Based on Corollary 2, the UL feasible region of **L** at t_m is obtained as

$$\mathbf{R}_{\mathbf{FR,UL}}^{\mathbf{L}}(t_m) = \Gamma\left(\mathbf{P}^{N_B}, R - D^{\mathbf{L}}\sqrt[\gamma]{\frac{\mathfrak{T}_\Phi}{\Theta}}\right), \tag{3.13}$$

where $D^{\mathbf{L}}$ is the distance between N_T and N_R. Let $\mathbf{P}_m^{N_R}$ denotes the position of N_R at t_m. If $\mathbf{P}_m^{N_T}$ is covered by $\mathbf{R}_{\mathbf{FR,UL}}^{\mathbf{L}}(t_m)$, **L** will be scheduled in the UL subframe and the scheduling will be completed; otherwise, it will go to Step 3.

The feasible regions are estimated under the assumption that there are MSs with conventional communication on the boundary edge of the scheduling BS's service range. Moreover, the conventional communication pairs in UL subframes will not be affected by the direct communication pairs. Based on these two reasons, a heuristic schedule is considered and is executed in Steps 3 and 4.

Step 3: The heuristic scheme assumes that the MSs with conventional communication may exist $\min\{D^I\}/2$ away from the boundary edge of the BS's service range, where $\min\{D^I\}$

51

is the radius of interference region for **L**. Therefore, the radius of $\mathbf{R^L_{FR,UL}}$ becomes $R - \min\{D^I\}/2$. In Step 3, the availability of direct communication for **L** is examined again, which checks whether N_B is located within the new $\mathbf{R^L_{FR,UL}}$, i.e.,

$$D_{NR}^{N_B} < R - \frac{\min\{D^I\}}{2} = R - \frac{1}{2}D^{\mathbf{L}}\sqrt[\tau]{\frac{\mathfrak{T}_\Phi}{\Theta}}, \tag{3.14}$$

where $D_{N_R}^{N_B}$ is the distance between N_B and N_R. If $\mathbf{P}_m^{N_R}$ is covered by the new $\mathbf{R^L_{FR,UL}}$, then go to step 4. Otherwise, **L** will not be conducted either in either the DL subframe or the UL subframe. The communication between N_T and N_R will be transferred to the conventional communication.

Step 4: The heuristic scheme is utilized to provide another chance of direct communication for **L**. If the chance results in a failed communication between N_T and N_R, the conventional communication scheme will be provided for N_T and N_R. In Step 4, therefore, **L** will be scheduled in DL subframe if $f^{\mathbf{L}} = 0$, which represents that the previous communication of **L** is success. Otherwise, the direct communication for **L** will not be permitted.

3.5 Predictive Motion and Interference-based Scheduling (PMIS) Algorithm

Since MSs may move around and their actual position information will not always be obtained by the BS, a predication mechanism should be considered to estimate the positions of MSs in the DCS problem. In this section, a flexible motion predication mechanism is given first; while the proposed PMIS algorithm is detailed in Section 3.5.2.

3.5.1 Motion Predication Mechanism

As mentioned in Section 3.2.2, the GMM model is considered in this work. In order to estimate parameters γ_k and ϱ_k at the time instant t_k, both (3.3) and (3.4) can be combined

and rewritten as

$$\underline{y}_k = \mathbf{H}_k \underline{\gamma}_k + \underline{v}_k, \tag{3.15}$$

where

$$\underline{y}_k = \begin{bmatrix} \alpha_k \\ V_k \end{bmatrix}, \qquad \mathbf{H}_k = \begin{bmatrix} \alpha_{k-1} & \bar{\alpha} & 0 & 0 \\ 0 & 0 & V_{k-1} & \bar{V} \end{bmatrix},$$

$$\underline{\gamma}_k = \begin{bmatrix} \gamma_{k-1} \\ 1 - \gamma_{k-1} \\ \varrho_{k-1} \\ 1 - \varrho_{k-1} \end{bmatrix}, \qquad \underline{v}_k = \begin{bmatrix} \sqrt{(1 - \gamma_{k-1}^2)}\Upsilon_{\alpha_{k-1}} \\ \sqrt{(1 - \varrho_{k-1}^2)}\Upsilon_{V_{k-1}} \end{bmatrix}.$$

The state vector $\underline{y}_k$ contains the moving direction α_k and velocity V_k of the MS at time instant t_k. $\mathbf{H}_k$ is the design matrix for parameter estimation, while $\underline{\gamma}_k$ represents the state vector for the time-varying parameters γ_{k-1} and ϱ_{k-1} of the MS at time instant t_k. The vector $\underline{v}_k$ denotes the system noises that are scaled from the random variables $\Upsilon_{\alpha_{k-1}}$ and $\Upsilon_{V_{k-1}}$. Therefore, the parameters γ_k and ϱ_k can be estimated (i.e., $\hat{\underline{\gamma}}_{k+1}$) at the time instance t_k by solving (3.15) using the recursive least square (RLS) estimation [25], [26] as

$$\hat{\underline{\gamma}}_{k+1} = \hat{\underline{\gamma}}_k - \mathbf{G}_{k+1}(\mathbf{H}_{k+1}\hat{\underline{\gamma}}_k - \underline{y}_{k+1}) \tag{3.16}$$

with

$$\begin{aligned}
\mathbf{K}_{k+1} &= \mathbf{G}_{k+1}\mathbf{H}_{k+1}^T, \\
\mathbf{G}_{k+1} &= \frac{1}{\lambda}\left[\mathbf{G}_k - \frac{\mathbf{G}_k\mathbf{H}_{k+1}^T\mathbf{H}_{k+1}\mathbf{G}_k}{\lambda + \mathbf{H}_{k+1}\mathbf{G}_k\mathbf{H}_{k+1}^T}\right],
\end{aligned} \tag{3.17}$$

where the adjustable parameter λ determines the convergence rate of the RLS method. As the parameters γ and ϱ are estimable with online adaptation, the further position of the MS can be forecasted based on the current state information.

It is assumed that the state information of an MS, including its position $\mathbf{P}_c = \{x_c, y_c\}$, moving direction α_c, and velocity V_c, is obtained at a time instant t_c via its position system. With (3.3) and (3.4), the position of the MS at the further time instant t_{c+n} for $n \in \mathbb{N}_1 = \{1, 2, \cdots\}$ can be estimated as

$$\hat{x}_{c+n} = x_{c+n-1} + V_{c+n-1} \triangle t \cos \alpha_{c+n-1},$$
$$\hat{y}_{c+n} = y_{c+n-1} + V_{c+n-1} \triangle t \sin \alpha_{c+n-1}, \tag{3.18}$$

where $\triangle t$ is the sampling interval between the time instants t_{c+n} and t_{c+n-1}.

3.5.2 PMIS Algorithm

Combining the PIS algorithm and the aforementioned motion predication mechanism, the PMIS algorithm is proposed. Similarly, let $\mathbf{L} = \{N_T, N_R\}$ denotes a direct communication pair, where N_T and N_R are the transmitter and receiver respectively. Fig. 3.5 illustrates the flowchart of the PMIS algorithm, which consists of nine important steps. It can be observed that Steps 1 to 4 is the four major steps of the PIS algorithm (as mentioned in Section 3.4), which is utilized to arrange a proper communication period for a pair of MSs that are expected to conduct direct communication. The additional steps, including Steps 5 to 9, of the PMIS algorithm is detailed as follows.

Step 5: In this step, the future positions of N_T and N_R are estimated based on (3.18).

Step 6: Based on the estimated position of N_T, the possibility of direct communication for $\mathbf{L}$ in the next DL subframe is calculated in this step. The DL feasible region for $\mathbf{L}$ is utilized to determine the calculating result. If the result shows that $\mathbf{L}$ can be arranged in the next DL subframe, the algorithm will be complete. Otherwise, it will go to Step 7.

Step 7: This step is utilized to predict whether the direct communication between N_T and N_R can be conduct in the next UL subframe. According to the estimated distances of $D^{\mathbf{L}}$ and
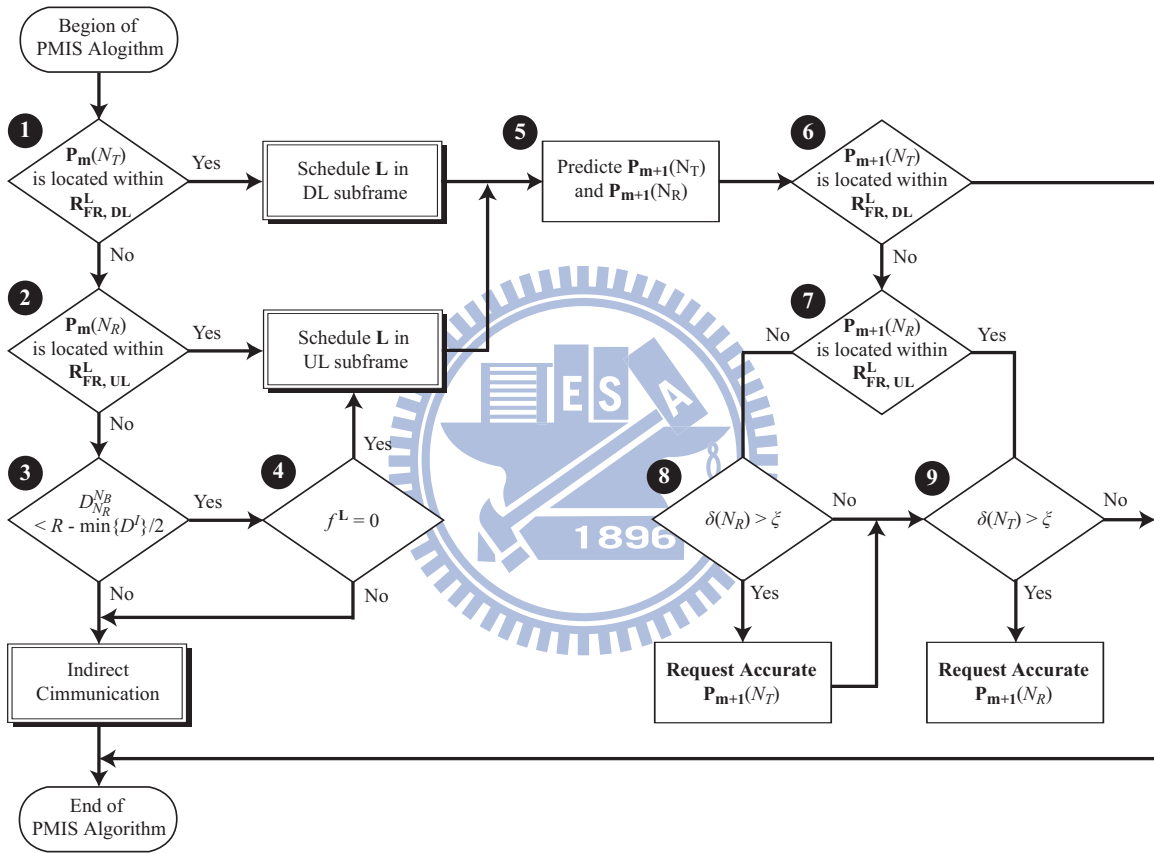
Figure 3.5: Flow diagram of PMIS Algorithm.

$D_{N_R}^{N_B}$, the DL feasible region is acquired to predict the availability of direct communication for **L**. If the communication is permitted, the algorithm will execute Step 9. Otherwise, it will go to Step 8.

Step 8: Let $\delta(N_R)$ denotes the number of times to predict N_R's position and ξ is a predefined threshold value. With the increment of $\delta(N_R)$, the predicted error for N_R's position may also be raised. In this step, therefore, $\delta(N_R)$ is examined. If $\delta(N_R) > \xi$, the BS will request N_R to provide its actual position information. Otherwise, it will go to Step 9.

Step 9: The functionality of this step is similar to that of Step 8, but the targeted MS is N_T instead of N_R. Let $\delta(N_T)$ denotes the number of times to predict N_T's position. If $\delta(N_R) > \xi$, the BS will request N_T to provide its actual position information and complete the algorithm. Otherwise, the algorithm will directly be ended.

3.6 Performance Evaluation

In this section, simulations are conducted to evaluate the performance of the proposed PIS and PMIS algorithms on the adaptive point-to-point communication (APC) approach (as mentioned in Chapter 2). A 19 cell-based wrap around topology [18] is considered as the simulation layout. Each cell consists of a centered BS and various number of uniformly distributed MSs. Both the intra-cell and inter-cell traffic flows are considered in the simulation, wherein each MS generates 20 traffic flows with randomly selected destination. The packet arrival process of each flow is assumed to follow a Poisson process with rate $\lambda = 2$ packet/frame. The size of each packet is selected to follow the exponential distribution with the mean value of 200 bytes. Since the scheduling algorithm is not specified in the IEEE 802.16 standard, the deficit round robin (DRR) [19] and weighted round robin (WRR) [20] algorithms are selected as the BS's DL and UL schedulers respectively. The DRR algorithm is also utilized by the SS to share the UL grants that are provided by the BS for its corresponding UL connections. The simulation is implemented via MATLAB even-driven simulator. Each obtained result is average from 100 simulation runs. The parameters adopted within the simulation are listed in Table 3.1.

Table 3.1: Simulation parameters for PMIS algorithm

Parameter	Value
OFDM symbol duration	34 μs
Maps modulation	BPSK
Data modulation	QPSK, 16-QAM, 64-QAM
Frame duration	10 ms
Number of MSs	10, 20, 30, 40, 50
Mobility model	random waypoint mobility
Mean velocity (V)	0, 5, 10, 15, 20 m/s
Simulation time	1 minute
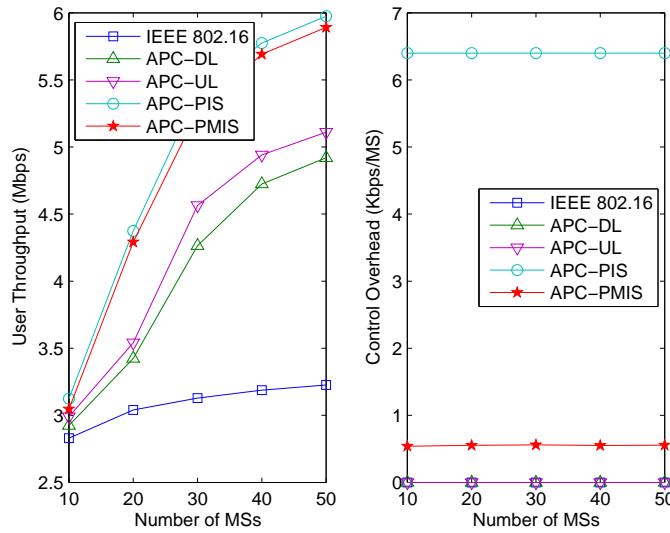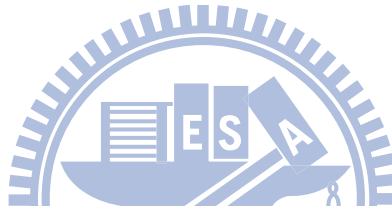
Figure 3.6: Performance comparison of user throughput and control overhead versus number of MSs (with $\bar{V} = 10$).

Fig. 3.6 illustrates the performance of the user throughput and control overhead under various number of MSs for the compared schemes. The mean velocity of 10 m/s is considered for all MSs in this comparison. As can be expected that the user throughput increases as the number of MSs is augmented for all the schemes. Due to the effect of direct communication conducted among MSs, the performance of the APC-based approaches outperform that of the conventional indirect scheme (denoted as IEEE 802.16). For the original APC approach (i.e.,without the consideration of scheduling algorithm for directly communicable pairs), it is observed that conducting direct transmissions in UL subframes (denoted as APC-UL) has better performance than that in DL subframes (denoted as APC-DL). The reason can be contributed to the potential inter-cell interferences to neighbor-cell MSs that are introduced by conducting direct communication in DL subframes, which can consequently decrease the entire system performance. If the direct transmissions are arranged in UL subframes, the conventionally indirect communication pairs will not be interfered since all the MSs are served as transmitters during those UL subframes. With the consideration of scheduling algorithms for directly communicable pairs in the APC approach (i.e., the APC-PIS and APC-PMIS schemes), the enhanced performance of user throughput is acquired. It is because that the directly communicable pairs are properly arranged in either DL subframes or UL subframes without inducing additional interference in the proposed PIS and PMIS algorithms. Comparing the APC-PIS and APC-PMIS approaches, it is observed that the higher user throughput and control overhead are shown in the APC-PIS scheme, which can be attributed to the acquirement of MSs position information. In the APC-PIS scheme, the exact position information of MSs are acquired via the updates of MSs, which results in large amount of control overhead. On the other hand, the position information of MSs are properly estimated by the BS in the APC-PMIS scheme, which efficiently reduces the necessary of position updates by MSs. Consequently, the lower control overhead is shown in the APC-PMIS approach in comparison with the APC-PIS scheme. Noted that the control overhead means the overhead related to the position updates of MSs. Since the position information of MSs is not utilized in the APC-DL, APC-UL, and IEEE 802.16 approaches, there is no control overhead in these
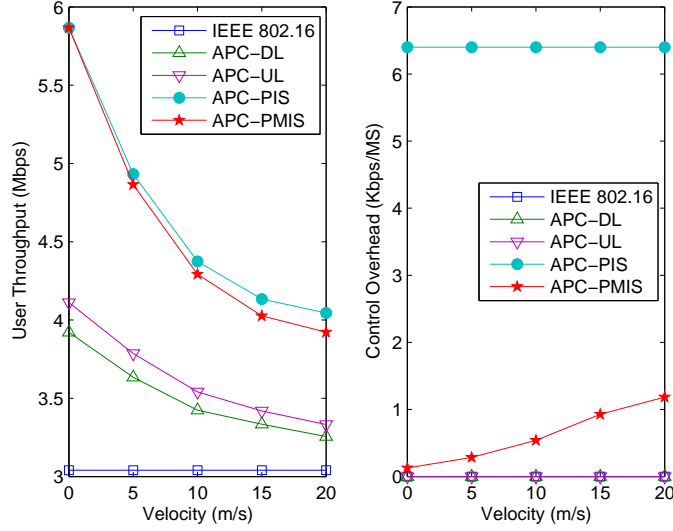
Figure 3.7: Performance comparison of user throughput and control overhead versus mean velocity ($\bar{V}$).

schemes.

Performance comparison with an increasing velocity ranging from 0 to 20 m/s is shown in Fig. 3.7, wherein 20 MSs are considered. As can be expected that the user throughput decreases as the mean velocity of MSs is increased for all the schemes excepted for the IEEE 802.16 approach, which can be attributed to the effect of direct communication. With larger velocity for the MSs, the variation of distances and channel conditions among MSs are changed frequently, which reduces the average life time of direct links. In such a case, the communication operation for the original direct communication pairs are changed from the direct manner to the indirect manner, which results in the decrement of user throughput. Due to the same reasons addressed for Fig. 3.6, the APC-PIS approach has the best performance of user throughput and worst control overhead among the compared schemes.

3.7 Concluding Remarks

In this work, a scheduling algorithm for each pair of MSs that is expected to conduct direct communication is proposed. Both the interference region and feasible region for the pair of

59
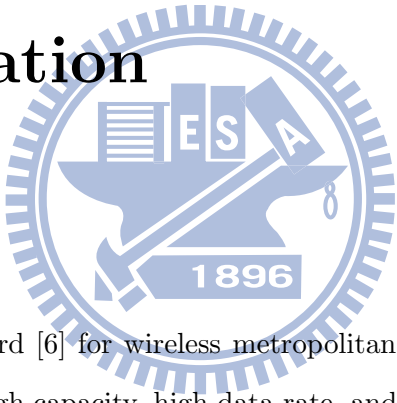
MSs to perform direct communication are studied and calculated. Based on these two types of information, the predictive interference-based scheduling (PIS) algorithm properly arranges the MSs to conduct direct communication in either DL subframe or UL subframe without increasing additional interference for other communications. Furthermore, since MSs may move around, a motion prediction mechanism is proposed. Combining the predication mechanism with the PIS algorithm, a predictive motion and interference-based scheduling (PMIS) algorithm is given to reduce the control overhead regarding the updates of MSs positions. The efficiency of the proposed PIS and PMIS algorithms are evaluated and compared via simulations. Simulation studies show that both of the PIS and PMIS algorithms efficiently enhance the performance of user throughput in comparison with the original adaptive point-to-point (APC) approach. Furthermore, the control overhead resulted from the PIS algorithm is significantly reduced by the PMIS approach.
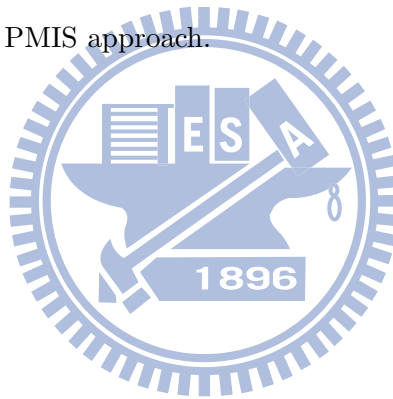
Chapter 4

Adaptive Listening Window Approach for IEEE 802.16m Sleep Mode Operation

4.1 Introduction

The IEEE 802.16-2004 standard [6] for wireless metropolitan area networks is developed to support various demand for high capacity, high data rate, and advanced multimedia services. The IEEE 802.16e amendment [7] enhances the original standard by addressing issues of mobility management and energy conservation for mobile stations (MSs). In order to support future advanced services with higher data rate and higher mobility, the next generation of IEEE 802.16 system is being developed in the IEEE 802.16m task group [3]. Since mobility is considered a key feature in wireless networks, how to prolong the battery lifetime of MSs has been recognized as one of the critical issues.

Several power-saving mechanisms are designed for conserving MSs energy in different wireless mobile networks. In IEEE 802.11 wireless local area networks [27], the access point (AP) periodically broadcasts beacons to notify the associated power-saving MSs regarding their traffic indications. The MS remains in sleep state most of time and wakes up to receive

the beacons with a fixed wake-up interval. If the beacon received by the MS indicates the presence of buffered packets, the MS will keep awake and issue power-saving poll (PS-poll) message to the AP in order to retrieve the buffered packets; otherwise it will return to the sleep state. After completely receiving the buffered packets in the awake state, the MS will go back to the sleep state immediately. On the other hand, the third-generation mobile cellular system universal mobile telecommunications system (UMTS) employs discontinuous reception (DRX) mechanism to conserve the power of MSs [28, 29]. The MS enters the power-saving mode while it has been idle for a period of time without data transportation in the power-active mode. In the power-saving mode, the MS is provided with a series of fixed-length DRX cycles wherein the MS stays in the sleep state. At the end of a DRX cycle, the MS will wake up and listen to the paging channel for upcoming messages. If the paging message indicates that there exists packets buffered at the base station (BS), the MS will terminate the power-saving mode and go back to the power-active mode for packet reception; otherwise it will return to the sleep state until the end of next DRX cycle.

In IEEE 802.16e networks, power-saving mode is named sleep mode that is defined as the state in which the MS conducts per-negotiated periods of absence from the serving BS air interface. In the sleep mode, the MS is provided with a series of sleep cycles which consist of both sleep windows and listening windows. The sleep window is a time period in which the BS shall not transmit data or management messages to the MS. During the listening window, on the other hand, the MS is expected to transmit/receive data or management messages based on the same manner as in the state of normal operations. Three types of sleep mode operations, which correspond to the power-saving classes (PSCs), are specified in the IEEE 802.16e standard to support different types of traffic. It is noted that each PSC is defined as a group of connections that share the common demand properties. Different types of PSCs are differentiated by parameter sets, procedures of activation/deactivation, and policies of MS's availability for data transmission. The difference among these PSCs mainly came from the way for determining the length of sleep windows.

The PSC of type II (PSC II) is defined by fixed period of sleep cycles that consist of fixed-

length listening windows and sleep windows, which is recommended for serving connections of unsolicited grant service (UGS) and real-time variable-rate (RT-VR) types. According to the traffic characteristics of these connections, two potential problems can occur by adopting the PSC II as follows: (*i*) unnecessary energy consumption resulted from unutilized frames within the listening windows, and (*ii*) packet loss due to the worse channel conditions. In this work, an approach with adaptive listening window (ALW) is proposed for the PSC II in IEEE 802.16e networks. According to the delay constraints, the ALW scheme dynamically adjusts the length of listening window based on the number of both buffered arrival packets and retransmission packets. It is worthwhile to mention that the concept of ALW approach has recently been proposed by the authors and is adopted in the IEEE 802.16m standard draft [30]. The performance of the proposed ALW scheme is evaluated and validated via both numerical analysis and simulation studies. The results show that the ALW scheme efficiently reduce the packet loss rate for both UGS and RT-VR connections in comparison with the PSC II mechanism. The energy conservation for RT-VR connections is also enhanced by adopting the ALW algorithm.

The remainder of this chapter is organized as follows. Section 4.2 briefly describes the system model of packet transmission and sleep mode operation for the IEEE 802.16e networks. The proposed ALW approach is explained in Section 4.3; while its numerical analysis is carried in Section 4.4. Both the performance evaluation and the validation of the ALW approach are conducted in Section 4.5. Section 4.6 draws the conclusions.

4.2 Preliminaries

4.2.1 System Model

The Point-to-multipoint (PMP) mode, where the traffic controlled by the BS, is considered the well-adopted network configuration in IEEE 802.16e networks. In this work, the PMP mode with time division duplexing (TDD) for a BS and an MS is considered as the system model to investigate the proposed ALW approach. In the PMP mode with TDD, each frame
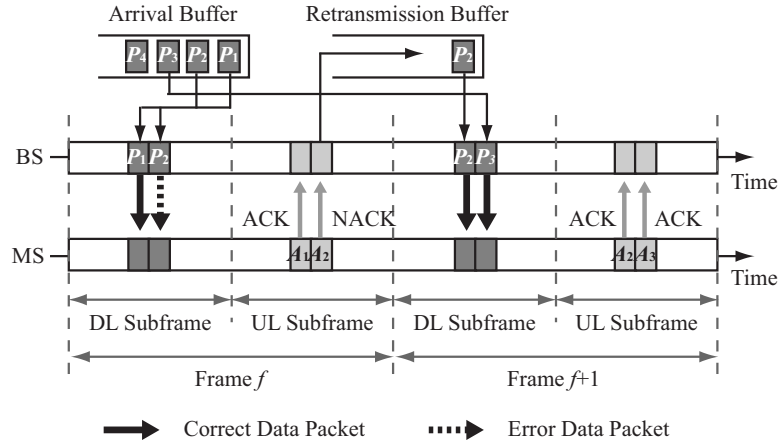
Figure 4.1: Schematic diagram of packet transmission model in IEEE 802.16 PMP mode with TDD.

consists of a downlink (DL) subframe and a uplink (UL) subframe, which are exploited to deliver DL traffic (from BS to MS) and UL traffic (from MS to BS) respectively. Moreover, in order to preserve the quality of service (QoS) for the connections, the corrupted packets are considered to be retransmitted based on the corresponding delay constraint.

An exemplified system model is illustrated in Fig. 4.1, wherein a single DL connection is considered. It is observed that all the arrival packets are queued into the arrival buffer of the BS before conducting its transmission services. The BS transmits the buffered packets to the MS in the DL subframe and arranges a corresponding opportunity for the MS to provide its reception status. If a packet is corrupted, a negative acknowledgement (NACK) message is sent back to the BS and the packet will be considered for retransmission based on the QoS constraint. On the other hand, in the case with successful packet transmission, a positive acknowledgement (ACK) message is returned to the BS and new packets will be consecutively transmitted to the MS. For the purpose of facilitating the explanation, the corrupted packets requesting for retransmission are retained in the retransmission buffer, which are considered to possess higher transmission priority over the packets stored in the arrival buffer. As shown in Fig. 4.1, two packets P_1 and P_2, stored in the arrival buffer, are transmitted in the DL subframe of the fth frame by the BS. The corresponding acknowledgement messages A_1 and A_2 are returned by the MS in the UL subframe of the same frame. Due to the transmission
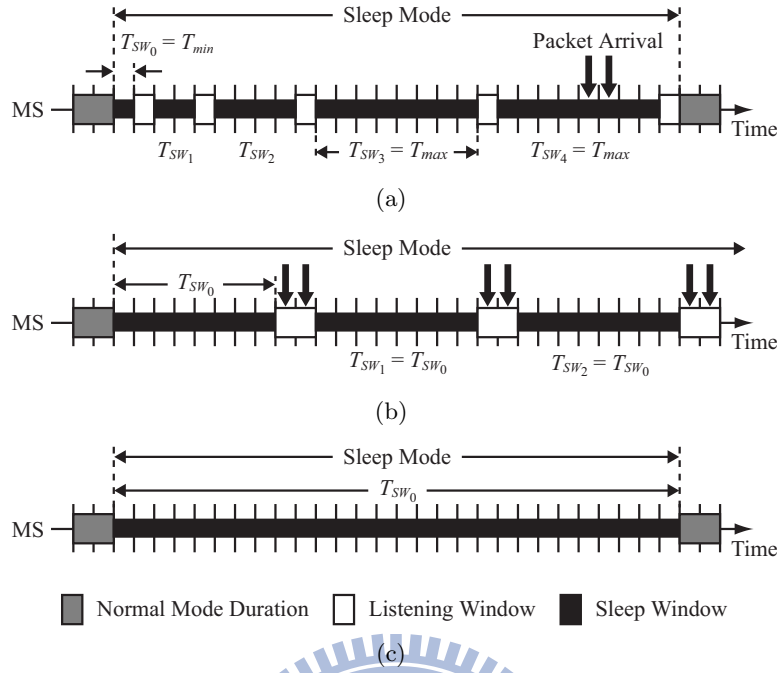
Figure 4.2: Schematic diagram of PSCs in IEEE 802.16 system: (a) PSC I, (b) PSC II, and (c) PSC III.

error reported via A_2, the packet P_2 is retained in the retransmission buffer and will be transmitted immediately in the DL subframe of the $(f+1)$th frame. It is noted that an intuitive and simple ACK/NACK scheme is exploited in this work. However, other types of BS/MS negotiation schemes can also be applied in the considered networks.

4.2.2 IEEE 802.16e Sleep Mode Operations

Fig. 4.2 depicts the operation of PSCs in IEEE 802.16e networks. For PSC of type I (PSC I) that is recommend for non-real-time traffic, the operation is similar to that of UMTS DRX mechanism by detecting the incoming traffic in the sleep mode and receiving the packets in the normal mode. The detection intervals (i.e., sleep windows) with length of binary-exponential increment is considered in the PSC I. As shown in Fig. 4.2(a), the PSC I starts with a sleep window of length $T_{SW_0} = T_{min}$. The length of the next sleep window is double of the previous one if no traffic is addressed to the MS, e.g., $T_{SW_1} = 2 \times T_{SW_0}$. This process is repeated as long as the length of sleep window does not exceed T_{max}. As the length of sleep window

reaches T_{max}, the lengths of subsequent sleep windows will remain constant at T_{max}, e.g., $T_{SW_4} = T_{SW_3} = T_{max}$. It is noted that the parameters T_{min} and T_{max} represent the minimum and maximum sleep windows respectively, which are negotiated by the BS and MS before activating PSC I. On the other hand, as shown in Fig. 4.2(b), alternately fixed-length sleep windows and fixed-length listening windows are defined in PSC II, which is recommended for serving real-time traffic. With the operation of PSC II, the MS keeps staying in the sleep mode and receiving packets within listening windows until an explicit termination of the sleep mode is requested by the MS or BS. The lengths of sleep windows and listening windows are predefined according to the negotiation between the BS and MS. It is noted that the permission of data transportation within the sleep mode is the major difference between PSC II and UMTS DRX mechanism. Fig. 4.2(c) illustrates the configuration of PSC of type III, wherein only one sleep window is considered for multicast or management traffic.

In recent research studies, performance analysis of sleep mode operations in IEEE 802.16e networks have been investigated. The basic sleep mode operation for the PSC I in IEEE 802.16e networks has been analytically modeled and discussed in [31–41]. The works conducted in [31–33] only consider DL connection in the sleep mode; while both DL and UL connections are investigated in [34–37]. The power-saving performance that affected by queue behaviors based on different queueing models are studied in [38–41]. Moreover, enhanced sleep mode mechanisms for PSC I have been proposed and analyzed in [42–46]. The works in [42, 43] dynamically enlarge the length of initial sleep window to reduce the total number of listening windows; while the probabilistic sleep windows is proposed in [44]. The adjustments of T_{mim} and T_{max} are investigated in [45, 46]. On the other hand, the PSC II is analyzed and discussed in [32, 37]. However, enhanced power-saving schemes for PSC II have not been studied and addressed in the existing literature.

4.2.3 Problem Statement

As mentioned in the previous subsection, the PSC II is defined by fixed period of sleep cycles that consist of fixed-length listening windows and sleep windows. It is recommended for

66

serving connections that require guaranteed data rate and delay, including both UGS and RT-VR types. The UGS connection is utilized to support fixed-rate data; while variable bit rates are supported by the RT-VR connection. According to the traffic characteristics of connections, two potential problems can occur by adopting the PSC II which are detailed as follows.

1. *Unnecessary Energy Consumption.* Since the UGS connection is designed to support real-time data stream with constant bit rate, the configuration of PSC II is considered appropriate for the UGS connection to conserve energy. As for the RT-VR connection, however, the length of listening window is generally set to achieve the required maximum bit rate. In such case, all the buffered arrival packets should be completely transmitted from the BS to the MS during the fixed-length listening window. However, with the nature of variable bit rate within the RT-VR connection, the design of fixed-length listening windows will result in the existence of unutilized frames, named *idle frames*, with full power-on mode. According to the current IEEE 802.16e standard, the MS will be kept awake until the end of listening window, which incurs excessive energy consumption for the MS.

2. *Packet Loss.* In order to improve the reliability of wireless links, retransmission mechanisms have been studied for real-time traffic [47, 48]. In general, packet retransmissions happen under the situations with worse channel conditions, which can resulted from either inter-cell interference or MS's mobility. With higher inter-cell interference or higher mobility of the MS, the frequency of packet retransmissions is required to be increased. Based on the predefined fixed length of listening window as described in PSC II, the buffered packets may not be completely transmitted from the BS to the MS. All the remained packets will violate the delay constraint and consequently be dropped. It is also noted that frequent transition between normal and sleep modes is not recommended in PSC II due to excessive handshakes between the BS and the MS. Therefore, with the fixed length of listening window, the packet loss rate for both the UGS and RT-VR connections can be significantly increased as the quality of wireless channel becomes

worse.

In order to overcome the aforementioned disadvantages of PSC II and consequently enhance the performance of sleep mode operation, a flexible and improved power-saving mechanism should be considered.

4.3 Adaptive Listening Window (ALW) Approach

In this section, the proposed ALW approach is presented for the PSC II in IEEE 802.16e networks. According to the standard for an MS in the sleep mode, the packets arrived at the BS during the sleep window of previous sleep cycle will be buffered, and consequently be transmitted within the listening window of current sleep cycle. Due to the characteristics of RT-VR connection, the total number of buffered packets can vary in different sleep cycles. Moreover, retransmissions for error packets that satisfy tolerable delay should also be considered owing to different channel conditions. In the design of proposed ALW approach, the length of listening window will be adaptively adjusted based on the total number of buffered packets that exist in both the arrival buffer and retransmission buffer as well as the delay constraint. It is noted that the arrival buffer is utilized to preserve the arrival packets in the previous sleep cycle; while the retransmission buffer keeps packets that are ready to be retransmitted within the current listening window. Compared with the PSC II, it is worthwhile to mention that the length of listening window will be dynamically increased or decreased for each sleep cycle while adopting the proposed ALW approach, which can consequently reduce both the unnecessary energy consumption and packet loss rate.

Fig. 4.3 illustrates the schematic diagram of an exemplified PSC II with the proposed ALW scheme for an RT-VR connection, where the length of predefined listening window (T_L) is selected as 2 frames. As mentioned above, the packets arrived at the BS during the $(k-1)$th sleep cycle are transmitted within the listening window of the kth sleep cycle. It is observed that all the buffered packets are completely transmitted in the first frame of the listening window within the kth sleep cycle and the second frame of that becomes idle. For the purpose of energy conservation, the MS decreases the length of listening window and
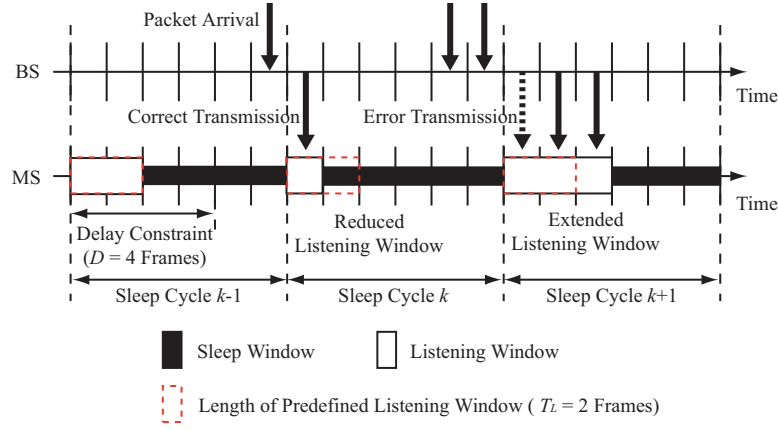
Figure 4.3: Schematic diagram of PSC II with proposed ALW approach for RT-VR connection with $T_L = 2$ frames and $D = 4$ frames.

returns to the sleep state early in the second frame of the kth sleep cycle. On the other hand, the length of listening window in the $(k + 1)$th sleep cycle is extended as 3 frames because of the transmission error occurred in the first frame. Since the delay constraint (D) for the packets is 4 frames in this example, retransmission for the error packets is required in order to preserve the QoS. It is noted that the maximum length for the extended listening window will be equal to D, which is the maximally tolerable packet delay for each connection.

Based on the concept as described above, an event-driven algorithm called *Listening Window Adjustment Algorithm* (refer to Algorithm 1) is proposed for the BS to make an appropriate decision for adjusting the listening window length of an MS. The notations utilized in the algorithm are listed as shown in Table 4.1. The algorithm begins if one of the following events is occurred within the sleep cycle k,

1. $B_A(t) = empty$ and $B_R(t) = empty$: all the buffered packets are completely transmitted. In such case, the activities of the BS and MS can be found in lines 2 and 3 of the Algorithm 1. The BS will send an information to the MS in the $(f_t + 1)$th frame for terminating the listening window, which consequently cause the MS to enter into the sleep state.

2. $t = (t_k + T_L d_f)$: the current time is in the end of the predefined listening window. According to the IEEE 802.16e standard, the MS will terminate the listening window

Algorithm 1: *Listening Window Adjustment Algorithm*

    Input: t, t_k, d_f, f_t, D, T_L, $B_A(t)$, and $B_R(t)$

    Output: $A_{BS}(f_t + 1)$ and $A_{MS}(f_t + 1)$

1 **if** $(t_k \leq t < (t_k + T_L d_f))$ *or* $((t_k + T_L d_f) < t < (t_k + D d_f))$ **then**

2 $A_{BS}(f_t + 1) \leftarrow$ send a termination information to the MS

3 $A_{MS}(f_t + 1) \leftarrow$ return to sleep state

4 **else if** $t = (t_k + T_L d_f)$ **then**

5   **if** $(B_A(t) = empty)$ *and* $(B_R(t) = empty)$ **then**

6 $A_{BS}(f_t + 1) \leftarrow$ do nothing

7 $A_{MS}(f_t + 1) \leftarrow$ return to sleep state

8   **else**

9 $A_{BS}(f_t + 1) \leftarrow$ send a extension information to the MS

10 $A_{MS}(f_t + 1) \leftarrow$ keep in the awake state

11   **end**

12 **else**

13   // $t = (t_k + D d_f)$

14   **if** $(B_A(t) = empty)$ *and* $(B_R(t) = empty)$ **then**

15 $A_{BS}(f_t + 1) \leftarrow$ send a termination information to the MS

16   **else**

17 $A_{BS}(f_t + 1) \leftarrow$ send a termination information to the MS and drop all the buffered packets

18   **end**

19 $A_{MS}(f_t + 1) \leftarrow$ return to sleep state

20 **end**

Table 4.1: Notations for Listening Window Adjustment Algorithm

Notation	Description
t	Initial time of algorithm execution
t_k	Initial time of the kth sleep cycle
d_f	Duration for a single frame
f_t	Frame where time t is located
D	Delay constraint in unit of frame
T_L	Length of the predefined listening window in unit of frame
$B_A(t)$	Status of the arrival buffer at time t
$B_R(t)$	Status of the retransmission buffer at time t
$A_{BS}(f_t + 1)$	Activity of the BS in $(f_t + 1)$th frame
$A_{MS}(f_t + 1)$	Activity of the MS in $(f_t + 1)$th frame

and return to the sleep state at $t = t_k + T_L d_f$. For the purpose of reducing packet loss rate while still satisfying the delay constraint, the statuses of both arrival buffer and retransmission buffer will be examined at this time instant. If both buffers are empty, the BS will do nothing and the MS will return to the sleep state (i.e., lines 6 and 7 in Algorithm 1), which follows the original operations as specified in the standard. Otherwise, the BS will send an information to the MS in the $(f_t + 1)$th frame for increasing the length of listening window, and the MS will keep awake until it receives a termination information from the BS (i.e., lines 9 and 10 in Algorithm 1).

3. $t = (t_k + D d_f)$: the current time is in the end of the delay constraint. Following the QoS requirement, the BS will drop all the buffered packets while the delay constraint is violated. In such case, a termination information will be sent by the BS to the MS in the $(f_t + 1)$th frame and the MS will return to the sleep state. The activities of the BS and MS can be found in line 14 to 19 in the Algorithm 1.

Since the complexity of the algorithm is $O(1)$, the decision for adjusting the listening window length can be made by the BS immediately. Therefore, it is expected that the proposed ALW approach, which is based on Algorithm 1, can be implemented in practical systems.

4.4 Numerical Analysis

In this section, a mathematical model is presented to analyze the energy consumption of the proposed ALW approach as well as the PSC II mechanism for an UGS/RT-VR connection. As mentioned previously, the packet arrived during the previous sleep cycle are stored in the arrival buffer at the BS. For the purpose of facilitating the analysis, it is assumed that the size of packets is constant for both UGS and RT-VR connections. Let the number of packets arrived during the previous sleep cycle denoted as N_A. For the UGS connection, N_A with the
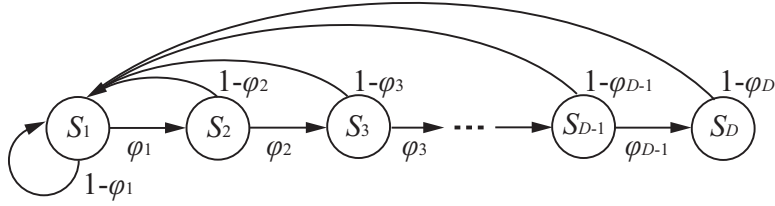
Figure 4.4: State transition model of the proposed ALW approach.

probability of n packets can be obtained as

$$P_r^{UGS}(N_A = n) = \begin{cases} 1 & \text{if } n = \lambda T_C \\ 0 & \text{otherwise} \end{cases}, \tag{4.1}$$

where λ is the packet arrival rate in unit of of packet per frame; while T_C denotes the length of sleep cycle in unit of frame. On the other hand, the packet arrival process of the RT-VR connection is assumed to follow a Poisson process with rate λ, where the probability of N_A with n packets can be expressed as

$$P_r^{RT-VR}(N_A = n) = \frac{(\lambda \cdot T_C)^n}{n!} e^{-(\lambda \cdot T_C)}. \tag{4.2}$$

All the buffered packets are considered to be completely transmitted to the MS within the current listening window while the delay constraint is satisfied. In general, the value of the delay constraint is considered between the length of listening window and the length of sleep cycle, i.e., $T_L \leq D \leq T_C$.

4.4.1 Analytical Model

Fig. 4.4 shows the state transmission model of the ALW scheme for an MS. S_k (for $k \in \{1, 2, \ldots, D\}$) denotes the state that all the buffered packets are completely transmitted in successive k frames, named *busy frames*, within a sleep cycle. The transition probability from S_i to S_{i+1}, denoted as φ_i for $1 \leq i \leq D-1$, will be obtained based on the probabilities $P_r(\mathcal{C}_1)$, $P_r(\mathcal{C}_2)$, and $P_r(\mathcal{C}_3)$ that are defined as follows. Let δ indicates the service rate in unit of packet per frame for the connection. The state transition from S_i to S_{i+1} can occur

while the number of packets stored in the arrival buffer is more than $i\delta$, where the probability $P_r(\mathcal{C}_1)$ can be acquired as

$$P_r(\mathcal{C}_1) \triangleq P_r(N_A > i\delta) = 1 - P_r(N_A \leq i\delta) = 1 - \sum_{n=1}^{i\delta} P_r^\varsigma(N_A = n), \qquad (4.3)$$

with $\varsigma \in \{UGS, RT - VR\}$ as the index to represent the probability from either the UGS connection or RT-VR connection. Furthermore, considering the scenario that the number of packets stored in the arrival buffer is less than $i\delta$, two other conditions will also cause the state transition to happen owing to the packet retransmissions. One of the conditions is that the number of packets from the arrival buffer and the error packets from the retransmission buffer counted from the previous $i - 1$ frames are observed to be more than $i\delta$. For the purpose of facilitating the analysis, it is assumed that the error probability for each transmitted packet is equal to p. The corresponding probability $P_r(\mathcal{C}_2)$ can be obtained as

$$\begin{aligned} P_r(\mathcal{C}_2) &\triangleq P_r(N_A \leq i\delta, \ N_A + N_R(i) - r_i > i\delta) \\ &= \sum_{n=1}^{i\delta} P_r^\varsigma(N_A = n) \cdot P_r(N_A + N_R(i) - r_i > i\delta | N_A = n), \end{aligned} \qquad (4.4)$$

where r_i denotes the number of packets with transmission error in the ith frame; while $N_R(i) = \sum_{j=1}^{i} r_j$ indicates the total number of error packets from the 1st to the ith frame in the current sleep cycle. The conditional probability $P_r(N_A + N_R(i) - r_i > i\delta | N_A = n)$ is denoted by $\Delta^{(i)}(n)$, which can be calculated as

$$\Delta^{(i)}(n) = \begin{cases} 0 & \text{for } 1 \leq n \leq \delta, \ 1 \leq i \leq D - 1 \\ \sum_{j=2\delta-n+1}^{\delta} \mathbb{C}_j^\delta & \text{for } \delta < n \leq 2\delta, \ i = 2 \\ \sum_{j=\max\{0,2\delta-n+1\}}^{\delta} \mathbb{C}_j^\delta \cdot \Delta^{(i-1)}(j + n - \delta) & \text{for } \delta < n \leq (i-1)\delta, \ 3 \leq i \leq D - 1 \\ \sum_{j=0}^{i\delta-n} \mathbb{C}_j^\delta \cdot \Delta^{(i-1)}(j + n - \delta) + \sum_{l=i\delta-n+1}^{\delta} \mathbb{C}_l^\delta & \text{for } (i-1)\delta < n \leq i\delta, \ 3 \leq i \leq D - 1 \end{cases},$$

$$(4.5)$$

where $\mathbb{C}_j^\delta \triangleq C_j^\delta p^j (1-p)^{\delta-j}$ is defined as the error probability of δ successively transmitted packets with j packets that are corrupted. The other condition happens while the total number of buffered packets is less than $i\delta$, however, there exists transmission errors in the ith frame. The resulting probability $P_r(\mathcal{C}_3)$ can be acquired as

$$P_r(\mathcal{C}_3) \triangleq P_r(N_A \leq i\delta, \ N_A + N_R(i) - r_i \leq i\delta, \ r_i \geq 1)$$

$$= \sum_{n=1}^{i\delta} P_r^\varsigma(N_A = n) \cdot P_r(N_A + N_R(i) - r_i \leq i\delta, \ r_i \geq 1 | N = n), \quad (4.6)$$

where the conditional probability $P_r(N_A + N_R(i) - r_i \leq i\delta, \ r_i \geq 1 | N = n)$ denoted by $\Gamma^{(i)}(n)$ is obtained as

$$\Gamma^{(i)}(n) = \begin{cases} \sum_{j=1}^{n} \mathbb{C}_j^n & \text{for } 1 \leq n \leq \delta, \ i = 1 \\ \sum_{j=1}^{n} \mathbb{C}_j^n \cdot \Gamma^{(i-1)}(j) & \text{for } 1 \leq n \leq \delta, \ 2 \leq i \leq D-1 \\ \sum_{j=0}^{\min\{\delta, i\delta-n\}} \mathbb{C}_j^\delta \cdot \Gamma^{(i-1)}(j+n-\delta) & \text{for } \delta < n \leq i\delta, \ 2 \leq i \leq D-1 \end{cases} \quad (4.7)$$

With (4.3), (4.4), and (4.6), the transition probability from S_i to S_{i+1} can be obtained as

$$\varphi_i \triangleq Pr(S_{i+1}|S_i) \quad (4.8)$$

$$= Pr(\mathcal{C}_1) + Pr(\mathcal{C}_2) + Pr(\mathcal{C}_3)$$

$$= 1 - \sum_{n=1}^{i\delta} P_r^\varsigma(N_A = n)(1 - \Delta^{(i)}(n) - \Gamma^{(i)}(n)), \quad (4.9)$$

where $1 \leq i \leq D-1$. Let π_k be the steady-state probability of state S_k. Based on (4.9), the steady-state probability of state S_j is expressed as

$$\pi_j = \pi_1 \prod_{i=1}^{j-1} \varphi_i, \quad 2 \leq j \leq D, \quad (4.10)$$

where π_1 is obtained by adopting the normalized condition $\sum_{l=1}^{D} \pi_l = 1$ as

$$\pi_1 = \frac{1}{1 + \sum\limits_{j=2}^{D} \prod\limits_{i=1}^{j-1} \varphi_i}. \tag{4.11}$$

Therefore, the average number of busy frames (T_B) can be acquired as

$$E[T_B] = \sum_{i=1}^{m} \pi_i \cdot i. \tag{4.12}$$

It is worthwhile to mention that the proposed analytical model can also be applied to analyze the PSC II by replacing the number of state in Fig. 4.4 with T_L. In other words, the maximum number of busy frames is limited to the value of T_L in PSC II. It is because that all the buffered packets are dropped in the end of the predefined listening window regardless of the potential longer delay constraint.

4.4.2 Energy Consumption

Let ε_B, ε_I, and ε_S denote the energy consumption per frame under busy state, idle state, and sleep state respectively. As a result, the energy consumption for a connection can be obtained as

$$E_C = \varepsilon_B E[T_B] + \varepsilon_I E[T_I] + \varepsilon_S E[T_S], \tag{4.13}$$

where T_I and T_S denote the number of idle frames and sleep frames respectively. For the proposed ALW approach, $E[T_I] = 0$ and $E[T_S] = T_C - E[T_B]$ since the MS immediately returns to the sleep state after the busy frames. The energy consumption of the ALW scheme (E_C^{ALW}) can be expressed as

$$E_C^{ALW} = (\varepsilon_B - \varepsilon_S)E[T_B] + \varepsilon_S T_C. \tag{4.14}$$

For the PSC II, on the other hand, the average number of idle frames and sleep frames can be acquired as $E[T_I] = T_L - E[T_B]$ and $E[T_S] = T_C - T_L$ respectively. Therefore, the energy consumption of PSC II (E_C^{PSC}) can be obtained as

$$E_C^{PSC} = (\varepsilon_B - \varepsilon_I)E[T_B^*] + (\varepsilon_I - \varepsilon_S)T_L + \varepsilon_S T_C, \qquad (4.15)$$

where $E[T_B^*]$ denotes the average number of busy frames for the PSC II.

4.5 Performance Evaluation

In this section, the energy consumption of the PSC II and the proposed ALW approach derived from the analytical model are validated via simulations. Furthermore, extensive simulations with different traffic models are conducted to evaluate the performance of the proposed ALW approach in comparison with the PSC II mechanism for UGS connection and RT-VR connection respectively. Single BS and one MS with a DL connection is considered in the simulation. In the PSC II, the length of the predefined listening window is generally set to achieve the transmission with required maximum bit rate. For the UGS connection, therefore, its value can be chosen as $T_L^U = \lceil (\lambda T_C)/\delta \rceil$; while $T_L^R = \lceil (2\lambda T_C)/\delta \rceil$ is considered for the RT-VR connection. The simulation is implemented via MATLAB even-driven simulator. Each obtained result is average from 100 simulation runs. The parameters adopted within the simulation are listed in Table 4.2, where the energy consumption parameters are acquired from an industrial manufactured mobile WiMAX chip [49].

Fig. 4.5 shows both simulation and analytical results of energy consumption over packet arrival rate for the proposed ALW approach and the PSC II mechanism. As illustrated in the figure, the analytical results are observed to coincide with the simulation outcomes. The detailed performance comparisons for the UGS connection and RT-VR connection are described as follows respectively.

Table 4.2: Simulation parameters for ALW approach

Parameter	Value
Frame duration	5 ms
Energy consumption of Busy frame	280 mW
Energy consumption of Idle frame	120 mW
Energy consumption of Sleep frame	10 mW
Packet arrival rate (λ)	0.1, 0.2, 0.3, 0.4, 0.5 packets/frame
Packet service rate (δ)	1, 2, 3, 4, 5 packets/frame
Packet error probability (p)	0.01, 0.03, 0.05, 0.07, 0.09
Length of Sleep Cycle (T_C)	30 frames
Simulation time	10 minutes
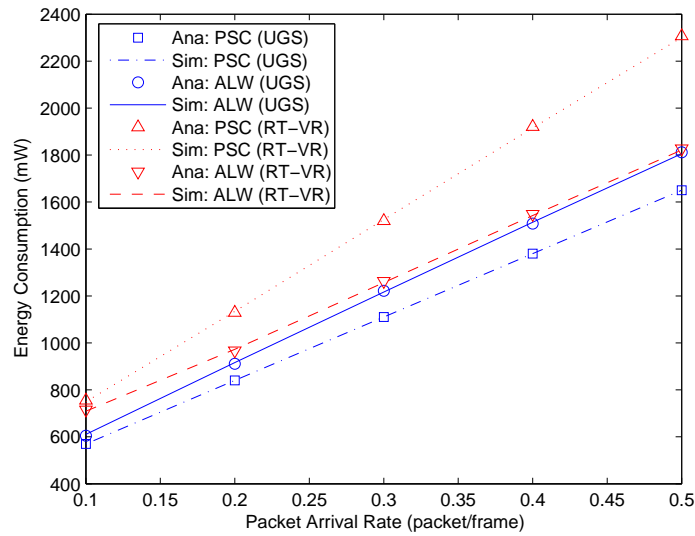
Figure 4.5: Validation of analytical results: energy consumption versus packet arrival rate.
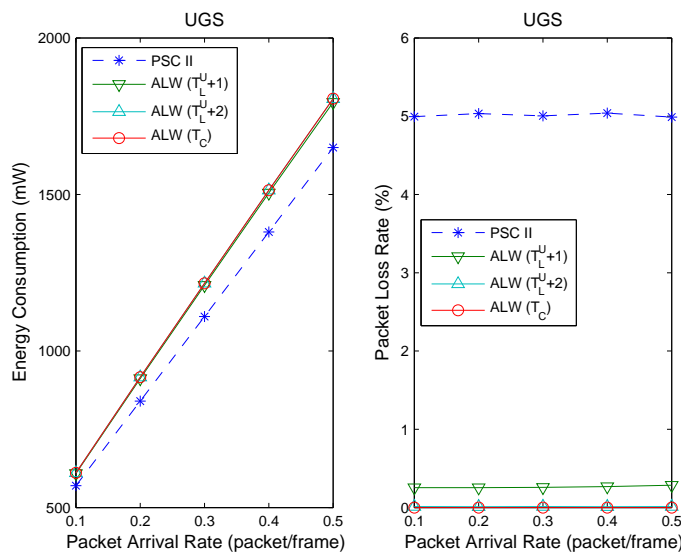
Figure 4.6: Performance comparison of energy consumption and packet loss rate versus packet arrival rate (λ) of UGS connection (with $p = 0.05$ and $\delta = 3$).

4.5.1 Simulation Results for UGS Connection

Since the fixed-rate data is supported by UGS connections, the number of packet arrivals during each sleep cycle can be considered in the simulations as a deterministic value of λT_C. Figs. 4.6 to 4.8 show the performance comparisons of the UGS connection between the proposed ALW approach and the PSC II mechanism. The delay constraint for the UGS connection is shown within the parenthesis of each scheme. For example, $\text{ALW}(T_L^U + 1)$ represents the ALW scheme for the UGS connection with delay constraint as the length of predefined listening window plus one frame. It is noted that same performance is obtained for the PSC II under different delay constraints due to the fixed-length listening window $T_L \leq D$. Therefore, the description of delay constraint for the PSC II is omitted.

Fig. 4.6 illustrates the performance comparison for both energy consumption and packet loss rate over various packet arrival rate (λ). The energy consumption increases as the value of λ is augmented in both schemes. It is observed that the energy consumption of the proposed ALW scheme is slightly higher than that of the PSC II, which can be attributed to the lower packet loss rate of the ALW approach. In the PSC II, the length of predefined listening window

Figure 4.7: Performance comparison of energy consumption and packet loss rate versus packet error probability (p) of UGS connection (with $\lambda = 0.3$ and $\delta = 3$).

is set to consist with constant bit rate for the UGS connection. In case the transmission error is occurred, the corrupted packets will be dropped without consideration for retransmission, which leads to the high packet loss rate. On the other hand, the ALW approach increases the length of listening window based on the number of buffered packets as well as the delay constraint. Therefore, the energy consumption is slightly increased but the packet loss rate is significantly reduced. Furthermore, with the exploitation of the ALW scheme, lowered packet loss rate can be achieved if the delay constraint is increased.

In order to reflect the effect from various channel conditions resulted from either inter-cell interference or MS mobility, different packet error probabilities are considered in the simulations. The performance comparisons with an increasing probability of packet error (p) is shown in Fig. 4.7. As can be expected that the energy consumption of the PSC II is constant over various values of p, which is due to the fixed-length listening window. However, the packet loss rate of the PSC II increases significantly as the value of p is augmented. Comparing with the PSC II, the ALW approach incurs higher energy consumption, which primarily results from its lowered packet loss rate. Fig. 4.8 illustrates the performance comparisons with various values of service rate (δ). It can be observed that the energy consumption of the
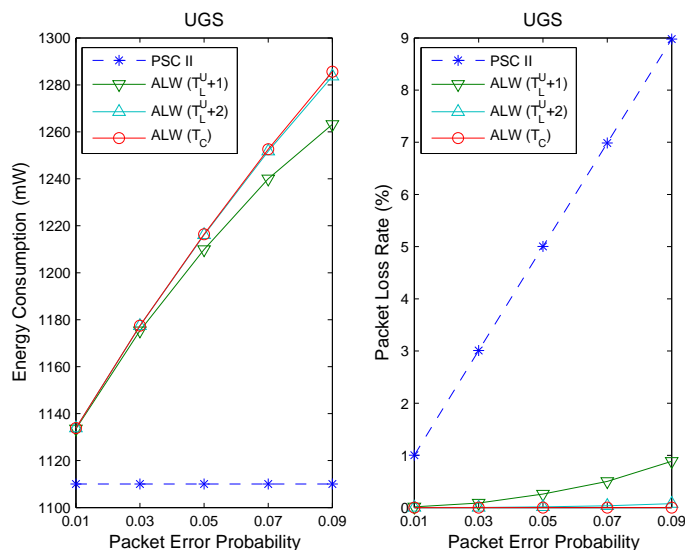
Figure 4.8: Performance comparison of energy consumption and packet loss rate versus packet service rate (δ) of UGS connection (with $\lambda = 0.3$ and $p = 0.05$).

PSC II either decreases or remains constant as the value of λ is increased. The reason can be attributed to the length of predefined listening window $T_L^U = \lceil (\lambda T_C)/\delta \rceil$ determined by λ, T_C, and δ. Given $\lambda = 0.3$ and $T_C = 30$ in the simulation, for example, $T_L^U = 5, 3, 3$ can be derived under the cases of $\delta = 2, 3, 4$ respectively, which results in the same energy consumption of the PSC II for $\delta = 3$ and 4. On the other hand, more opportunities for packet transmission will lead to lower packet loss rate for a constant number of packets. Based on the example mentioned above, transmission opportunities $\delta T_L^U = 10, 9, 12$ are provided for the entire 9 buffered packets under the case of $\delta = 2, 3, 4$ respectively. Therefore, the packet loss rate of PSC II with $\lambda = 2$ is lower than that with $\lambda = 3$, and $\lambda = 4$ has the lowest one as shown in Fig. 4.8. As for the ALW scheme, it is expected that the energy consumption is slightly increased but the packet loss rate is significantly reduced in comparison with the PSC II.

4.5.2 Simulation Results for RT-VR Connection with Poisson Distribution

In this subsection, the packet arrival process of the RT-VR connection is assumed to follow a Poisson process with rate λ. The performance comparisons between the proposed ALW approach and the PSC II mechanism are given in Figs. 4.9 to 4.11. The performance com-

Figure 4.9: Performance comparison of energy consumption and packet loss rate versus packet arrival rate (λ) of RT-VR connection (with $p = 0.05$ and $\delta = 3$).

Figure 4.10: Performance comparison of energy consumption and packet loss rate versus packet error probability (p) of RT-VR connection (with $\lambda = 0.3$ and $\delta = 3$).
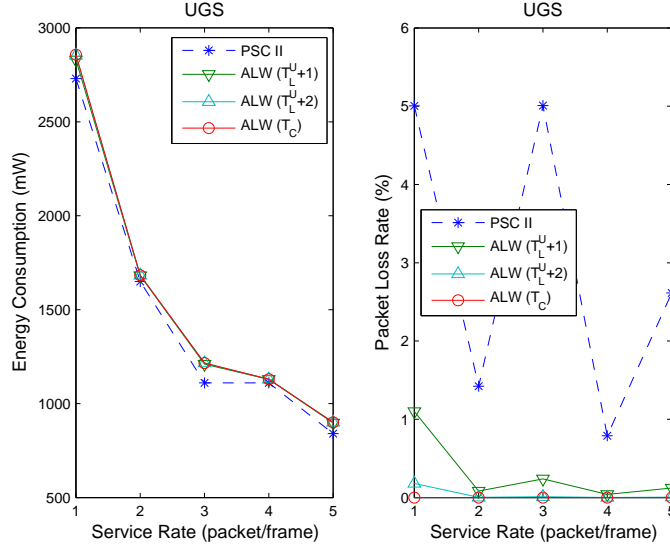
Figure 4.11: Performance comparison of energy consumption and packet loss rate versus packet service rate (δ) of RT-VR connection (with $\lambda = 0.3$ and $p = 0.05$).

parison with an increasing value of λ is shown in Fig. 4.9. In the PSC II, the length of predefined listening window is set to achieve the transmission with required maximum bit rate for the RT-VR connection. Although it can provide opportunities for packet retransmission, the idle frames with unnecessary energy consumption is also produced in most cases. Moreover, packet loss can happen while the number of arrived packets is equal to the number of packets that can be served within the predefined listening window. On the other hand, the proposed ALW scheme will adaptively adjusts the length of listening window based on the states of buffers as well as delay constraint. Comparing with the PSC II, lower energy consumption and lower packet loss rate are achieved by the ALW scheme. Figs. 4.10 and 4.11 illustrates the performance comparisons with various packet error probabilities and service rates respectively. It can be expected that all the performance curves of the proposed ALW scheme outperform that of the PSC II mechanism, which can be explained by similar reasons as discussed in the previous subsection. The merits of proposed ALW scheme can therefore be observed.
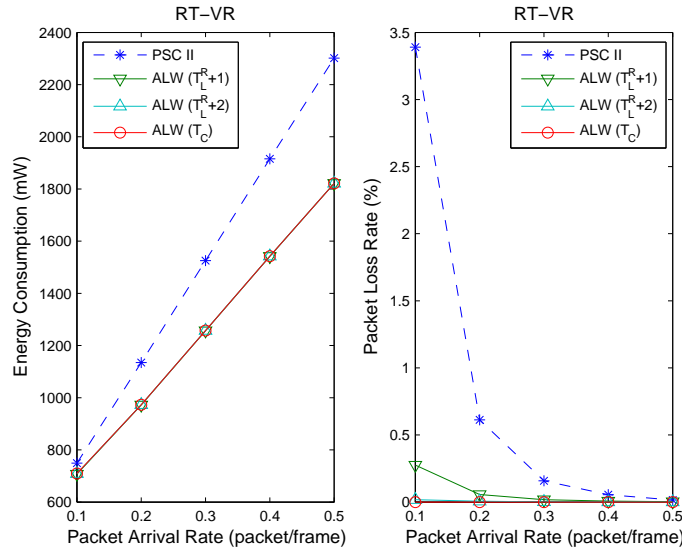
Figure 4.12: Performance comparison of energy consumption and packet loss rate versus packet arrival rate (λ) of RT-VR connection (with $\alpha = 0.5$, $p = 0.05$, and $\delta = 3$).

4.5.3 Simulation Results for RT-VR Connection with Gamma Distribution

The effect of different traffic models on the performance of PSC II and ALW approach is discussed in this subsection. The Gamma distribution is considered as the distribution of packet inter-arrival time for the RT-VR connection since it has been well-adopted to model different types of random variables [50]. Let t_a denotes the variable of packet inter-arrival time, whose probability density function with parameters $\alpha > 0$ and $\beta > 0$ can be defined as

$$f_{t_a}(t_a) = \frac{t_a^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-t_a/\beta} \quad \text{for } t_a \geq 0, \tag{4.16}$$

where $\Gamma(\alpha)$ is the gamma function with shape parameter α, and β is a scale parameter to determine the mean packet inter-arrival time, i.e., $1/\lambda = \alpha\beta$. It is noted that the Gamma distribution with $\alpha = 1$ presents the exponential distribution for packet inter-arrival time, which corresponds to the Poisson process for packet arrivals considered in the previous subsection.

Figs. 4.12 to 4.14 show the performance comparisons of the RT-VR connection with Gamma distribution between the proposed ALW approach and the PSC II mechanism. The performance comparison with the case of $\alpha = 0.5$ over various packet arrival rates and packet
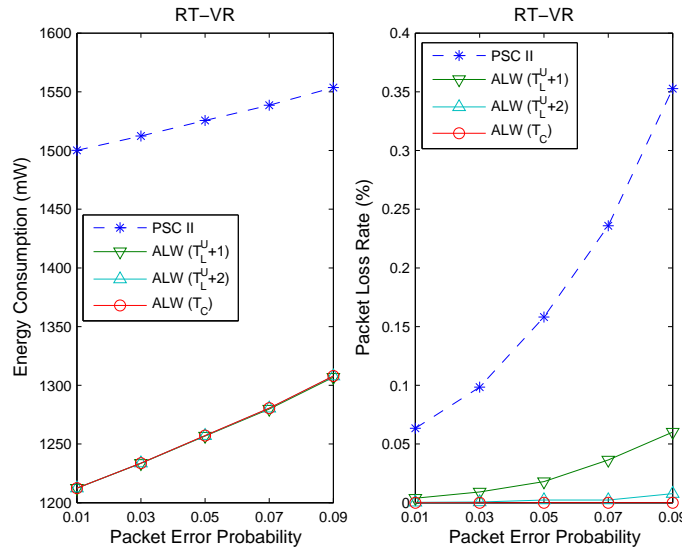
Figure 4.13: Performance comparison of energy consumption and packet loss rate versus packet error probability (p) of RT-VR connection (with $\alpha = 0.5$, $\lambda = 0.3$, and $\delta = 3$).
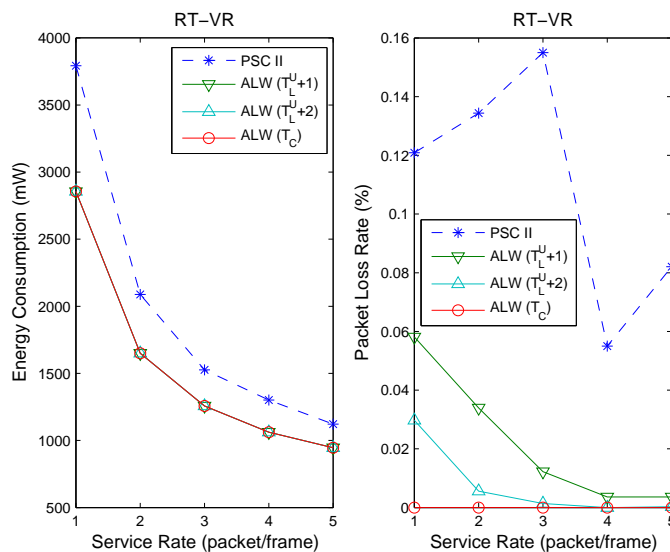

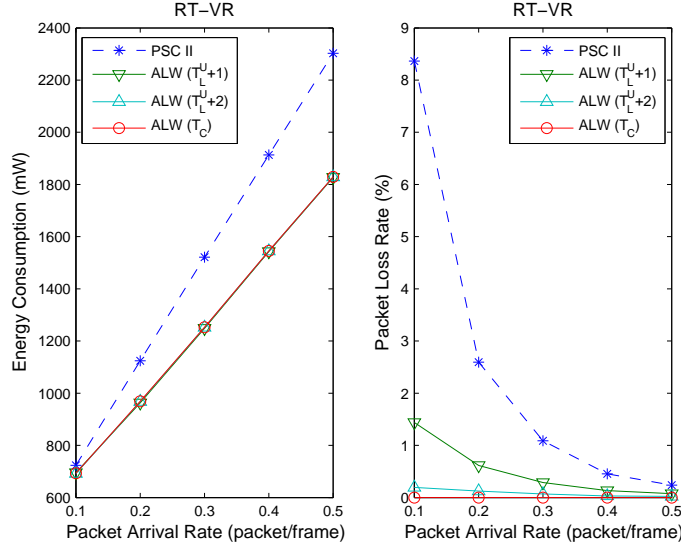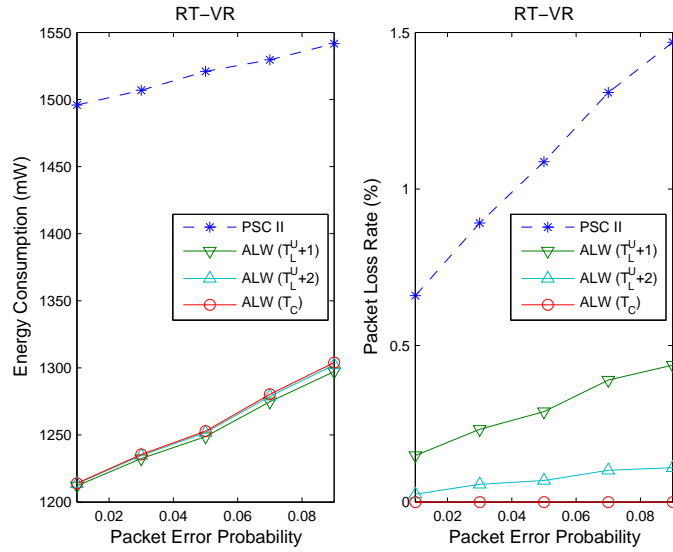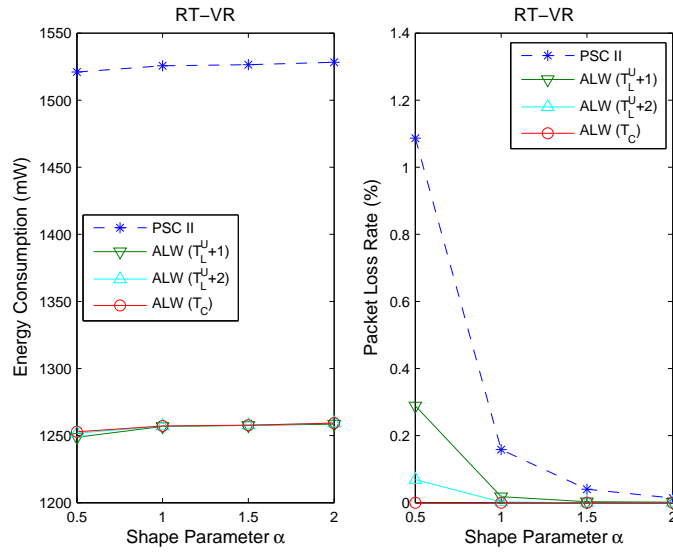
Figure 4.14: Performance comparison of energy consumption and packet loss rate versus shape parameter (α) of RT-VR connection (with $\lambda = 0.3$, $p = 0.05$, and $\delta = 3$).

error probabilities are given in Figs. 4.12 and 4.13, respectively. As can be expected that all the cases by adopting the proposed ALW approach outperform the PSC II mechanism, which can be explained by similar reasons as discussed in the previous subsection. Fig. 4.14 depicts the performance comparisons with various α values. By observing the different cases using the ALW scheme, the energy consumption of $\alpha = 0.5$ is slightly lower than that from the other α values. However, relatively higher packet loss rate is obtained with the $\alpha = 0.5$ case, which can be attributed to the characteristics of Gamma distribution. For $\alpha < 1$, the traffic becomes non-smooth since the standard deviation of t_a is larger than its mean value. In this situation, larger variance in number of aggregated packets will be obtained from each sleep cycle. Since the maximum length of listening window is considered constant in each case, less energy consumption for the $\alpha = 0.5$ case will be acquired due to its potential small number of aggregated packets. On the other hand, higher packet loss rate will be obtained owing to the constant listening window size which can not accommodate the larger number of aggregated packets. For larger α values with the condition of $\alpha > 1$, the ratio of the standard deviation to the mean approaches zero, which results in the situation that t_a approaches a deterministic value $\alpha\beta$ and comparatively high energy consumption and low packet loss rate will be obtained. Finally, it can be observed that the performance of the ALW approach outperform that of the PSC II mechanism under different values of α. The merits of proposed ALW scheme can therefore be observed.

4.6   Concluding Remarks

In this chapter, a power-saving mechanism with adaptive listening window (ALW) is proposed for the power-saving classes of type II (PSC II) in IEEE 802.16e networks. The ALW scheme adaptively adjusts the length of listening window based on both the buffered packets and delay constraint. Moreover, the concept of the ALW approach has recently been proposed by the authors and is adopted in the IEEE 802.16m standard draft. The energy consumption of the ALW and PSC II for UGS/RT-VR connection is evaluated and validated via both the numerical analysis and simulations. Simulation results show that the ALW scheme outperforms

the PSC II for the RT-VR connection in terms of energy conservation. Furthermore, significant enhancement of packet loss rate for both the UGS and RT-VR connection is achieved by adopting the proposed ALW scheme.
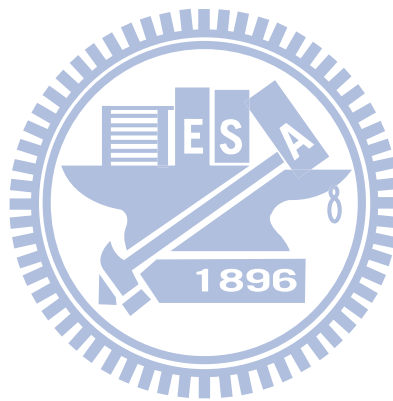
Chapter 5

Statistical Sleep Window Control Approach for IEEE 802.16m Sleep Mode Operation

5.1 Introduction

For a mobile station (MS) with non-real-time downlink traffic, an alternation of normal mode and sleep mode is provided for the MS to conserve its energy. The MS enters into the sleep mode while it has been idle for a pre-specified period of time in the normal mode without data transmission. The sleep mode operation of the MS is recommended to follow power-saving class of type I (PSC I), wherein the length of sleep window starts with a predefined initial-window value and increases exponentially until a predefined final-window value is reached. If the arrival of packets is indicated by the base station (BS) within the fixed-length listening window, the MS deactivates the sleep mode and returns to the normal mode for data reception. The performance analyses of PSC I have been investigated in recent research studies. The works in [31, 34] analytically modeled the basic operation of PSC I with the consideration of sleep mode only; while a system consists of normal mode and sleep mode is analyzed in [32, 33]. The complex model for multiple downlink and uplink traffic is modeled and is investigated

in [37]. On the other hand, enhanced sleep mode mechanisms for PSC I have been proposed and analyzed in [42, 43, 45, 46]. The works in [42, 43] dynamically enlarge the length of initial sleep window to reduce the total number of listening windows; while the adjustment of the initial and final sleep windows are investigated in [45, 46]. From the analytical results of these studies, it can be perceived that the inefficiency of PSC I comes from the configuration of sleep mode operation and the mechanism of binary-exponential traffic detection. For the purpose of enhancing the energy efficiency of MSs, several notions are included in the IEEE 802.16m system description document (SDD) [51] and an improved sleep mode operation is required by IEEE 802.16m system [52].

In this work, a statistical sleep window control (SSWC) approach, following the notions of IEEE 802.16m sleep mode operation, is proposed to maximize the energy conservation for an MS with non-real-time downlink traffic. The proposed SSWC approach determines the length of each sleep window that is appropriate for present traffic state with the consideration of tolerable delay and/or queue size. Since the states of non-real-time traffic are difficult to be obtained, a traffic model construction (TMC) procedure and a traffic state estimation (TSE) process are presented in the SSWC approach. The TMC procedure models a discrete-time Markov-modulated Poisson process (dMMPP) based on the trace of non-real-time traffic; while a partially observable Markov decision process (POMDP) is exploited to conjecture the present traffic state in the TSE process. Furthermore, performance evaluations of window selection made in each traffic state are conducted to indicate the proper length of each sleep window. Based on the properties of POMDP, two suboptimal policies for sleep window selection are exploited in the SSWC approach, including the sleep ratio-based (SR) and energy cost-based (EC) policies. The efficiency of proposed SSWC approach is evaluated and compared via simulations. In terms of energy efficiency and packet transmission delay, simulation results show that the proposed SSWC approach outperforms both the conventional PSC I and the evolutional PSC I from IEEE 802.16m system.

The remainder of this chapter is organized as follows. Section 5.2 briefly describes the sleep mode operation of IEEE 802.16e and IEEE 802.16m systems, associated with the for-
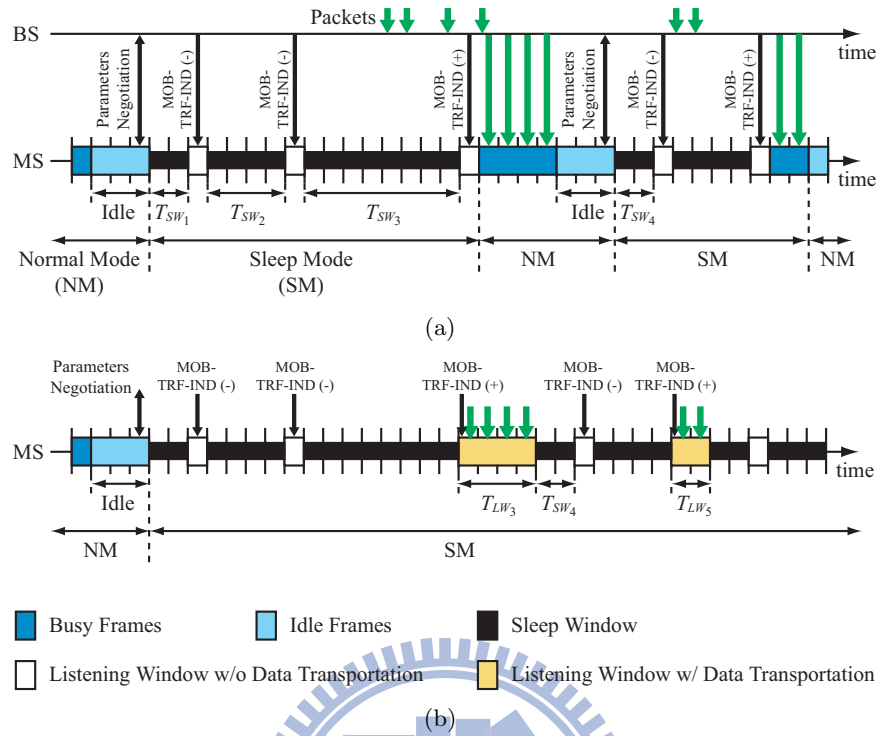
Figure 5.1: Schematic diagram of sleep mode operation in IEEE 802.16 systems: (a) PSC I in IEEE 802.16e system and (b) evolutional PSC I in IEEE 802.16m system.

mulation of the target problem. The traffic model construction and state estimation within the proposed SSWC approach, including the TMC and TSE procedures, are explained in Section 5.3. Section 5.4 describes two selection policies, i.e., the sleep ratio-based and energy cost-based policies, of the SSWC approach with the considerations of delay and/or queue size. Performance evaluation of proposed SSWC approach is illustrated in Section 5.5. Section 5.6 draws the conclusions.

5.2 Preliminaries

5.2.1 IEEE 802.16e Sleep Mode Operation with Non-real-time Traffic

In IEEE 802.16e system, non-real-time traffic is supported by PSC I wherein three major parameters, including length of initial-sleep window (T_{min}), length of final-sleep window (T_{max}), and length of listening window, are negotiated between the BS and MS in the normal mode.

Fig. 5.1(a) illustrates the sleep mode operation of PSC I in IEEE 802.16e system. The MS enters the sleep mode while it has been idle for a period without data transportation in the normal mode. This period is normally standardized and serves as a timer in practical implementation. When PSC I is activated, it starts with a sleep window of length $T_{SW_1} = T_{min}$. After the sleep window, a listening window is provided for the MS to receive traffic indication message MOB-TRF-IND, which is broadcasted from the BS. This message indicates whether there has been traffic addressed to the MS. If negative indication is acquired from the MOB-TRF-IND message, the MS will continue in the sleep mode after the listening window; otherwise it will return to the normal mode for data reception. In the case that the MS continues in the sleep mode, the length of the next sleep window is doubled from the previous one, e.g., $T_{SW_2} = 2 \times T_{SW_1}$. This process is repeated as long as the length of sleep window does not exceed T_{max}. As the length of sleep window reaches T_{max}, the lengths of subsequent sleep windows will remain constant at T_{max}. On the other hand, in the case that the sleep mode of the MS is reactivated in the normal mode, the length of first sleep window is reset as T_{min}, e.g., $T_{SW_4} = T_{min}$ as shown in Fig. 5.1(a).

As can be observed in Fig. 5.1(a), a series of alternate sleep modes and normal modes are provided for the MS by adopting PSC I in IEEE 802.16e system. It is perceived that the inefficiency of PSC I comes from two parts as follows: (a) frequent state transitions and (b) under-utilized listening windows. According to the operations of PSC I, the MS deactivates the sleep mode and returns to the normal mode for data reception. The sleep mode is reactivated while an idle period without data transmission occurs in the normal mode. This configuration results in unnecessary energy waste of the MS during the idle periods. On the other hand, a binary-exponential mechanism for increasing the length of sleep window is exploited in PSC I in order to detect incoming traffic. This mechanism resets the length of sleep window as T_{min} while the PSC I is reactivated, which consequently introduces a large number of under-utilized listening windows. All the listening windows are utilized to receive traffic indication messages only, which reduces energy efficiency of the MS.

5.2.2 Notions of IEEE 802.16m Sleep Mode Operation

The IEEE 802.16m standard for next generation mobile BWNs is designed to support advanced services with higher data rate and higher mobility. In the meanwhile, an improved power-saving mechanism is designed to enhance the energy conservation of MSs. In order to contend with aforementioned disadvantages of PSC I and to provide a flexible sleep mode operation, the IEEE 802.16m SDD specifies two significant notions, which are described as follows:

1. **Listening window adjustment.** During listening windows, the MS is expected to transmit/receive data or management messages based on the same manner as in the state of normal operations. The length of each listening window can be dynamically adjusted based on traffic availability or control signaling.

2. **Sleep mode parameter update.** The BS or MS may dynamically update the parameters of sleep mode operation (e.g., length of sleep window) based on the change of traffic patterns. The updating procedure may be executed without deactivating the sleep mode.

The first notion successfully eliminates the energy waste resulted from the idle periods since the data transportation is permitted during the adjustable listening windows in the sleep mode. As for the drawback of too many under-utilized listening windows, the second notion provides a flexible platform to dynamically change the configuration of sleep mode operation, which consequently reduces the number of under-utilized listening windows.

Following the notions described above, the PSC I with slight modification, named as evolutional PSC I, can be applied to IEEE 802.16m system. Fig. 5.1(b) illustrates the operation of evolutional PSC I in IEEE 802.16m system. It can be observed that all the processes of the evolutional PSC I are the same as that of the conventional PSC I in IEEE 802.16e system (as shown in Fig. 5.1(a)) except for the period of data transportation. In the evolutional PSC I, the arriving packets will be received by the MS during the adjustable listening windows in the sleep mode (e.g., T_{LW_3} and T_{LW_5} in Fig. 5.1(b)) instead of returning

to the normal mode. After the completion of data transportation, the MS terminates the listening window and returns to sleep state immediately. The length of the followed sleep window is reset to the predefined initial-window value, e.g., $T_{SW_4} = T_{min}$. Therefore, the MS keeps staying in the sleep mode and the process is repeated until an explicit termination of the sleep mode is requested by the MS or the BS. The evolutional PSC I successfully overcomes the problem of frequent state transitions; nevertheless, additional numbers of unutilized listening windows are still existed within the sleep mode operation. The performance of evolutional PSC I will be evaluated and be described in Section 5.5.

5.2.3 Problem Statement

Based on the aforementioned notions of IEEE 802.16m system, it is motivated that a flexible sleep mode operation should be provided to improve the energy consumption of MSs. The sleep mode operation for an MS can be modeled as a sleep windows selection (SWS) problem that is composed of multiple operations of control cycles. The definition of control cycle and the problem statement are described as follows:

Definition 3 (Control Cycle). *Given a BS and an MS that expects to enter the sleep mode or has stayed in the sleep mode, a control cycle C_i is defined as a time duration consisting of a decision epoch d_i, a sleep window SW_i, and a listening window LW_i. The BS determines the length of the sleep window SW_i at the decision epoch d_i. The MS stays in the sleep mode during the sleep window SW_i and wakes up for data transmission in the listening window LW_i.*

Problem 3 (Sleep Windows Selection Problem). *Given a non-real-time downlink traffic and a sequence of control cycles, how to select a proper length of sleep window under the tolerable packet delay δ for each control cycle in order to maximize the energy efficiency for the MS?*

Fig. 5.2 depicts the time sequence of ideal sleep mode operation for an MS with non-real-time downlink traffic, wherein all the control signals are ignored for the convenience of
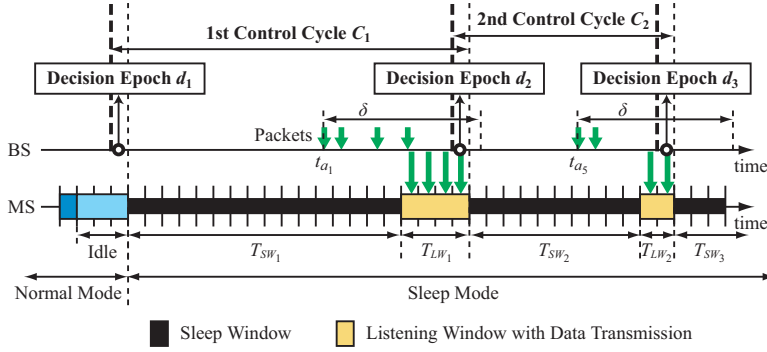
Figure 5.2: Schematic diagram of ideal sleep mode operation for an MS with non-real-time downlink traffic.

description. It can be observed that each control cycle C_i is overlapped with the adjacent control cycles C_{i-1} and C_{i+1} excepted for the first control cycle. The first control cycle C_1 starts at the tail of Ms's idle period in the normal mode. The remainder control cycles are individually executed in the end of the listening window within the previous control cycle, e.g., the cycle C_2 begins at the last frame of the listening window LW_1.

5.3 Statistical Sleep Window Control (SSWC) Approach

In order to resolve the SWS problem, a statistical sleep window control (SSWC) approach is proposed, which determines the length of each sleep window that is appropriate for present traffic state and delay constraint. As shown in Fig. 5.2, the first packet arrives at the BS at time t_{a_1} and the tolerable delay is δ. Therefore, the termination of sleep window T_{SW_1} that is determined at the decision epoch d_1 should fall within the time interval ranging from t_{a_1} to $t_{a_1} + \delta$. Similarly, at the decision epoch d_2, the length of second sleep window T_{SW_2} is determined according to the packet arrival time t_{a_2} and the delay constraint δ. It is intuitive that the length selected for each sleep window is dominated by the knowledge of the present traffic patterns, e.g., packet arrival time and packet arrival rate. However, these types of information is considered difficult to be obtained for non-real-time traffic. Only the number of buffered packets that arrived during the previous control cycle can provide information for potential estimation of the current traffic state.
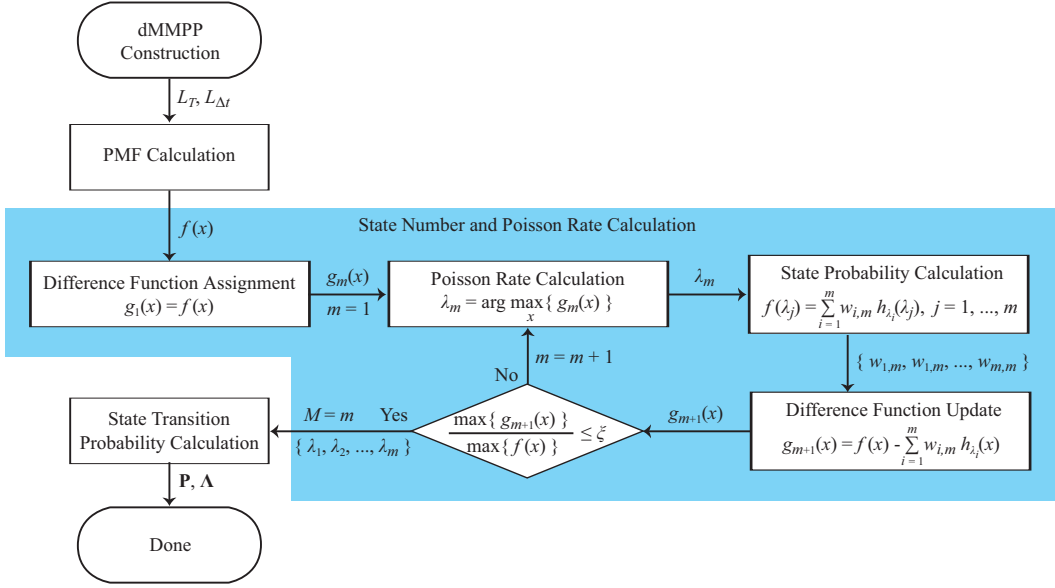
Figure 5.3: Flow diagram of TMC procedure.

In the proposed SSWC approach, according to the information acquired from previous states, a POMDP [53, 54] is exploited to conjecture the present state of non-real-time traffic at each decision epoch. In the following two subsections, two procedures of the proposed SSWC approach will be presented including the TMC and TSE processes. The TMC procedure models a dMMPP based on the trace information from non-real-time traffic; while the POMDP formulation is conducted in the TSE procedure in order to estimate the current traffic state. Based on the present traffic state estimated via these two procedures, two sub-optimal policies for solving the SWS problem are designed within the SSWC approach, which will be explained in Section 5.4.

5.3.1 Traffic Model Construction (TMC) Procedure

In the proposed SSWC approach, the non-real-time downlink traffic is modeled as a dMMPP, which is considered more generic than the conventional Poisson model and is able to capture the characteristics of the Internet traffic [55–57]. While the notion of dMMPP construction presented in [57] is exploited to capture the characteristics of self-similar traffic at multiple time scales, it can also be applied to calculate the states of the non-real-time traffic. Therefore,

94

the objective of TMC procedure is to establish an inferred dMMPP model for non-real-time traffic, which will be utilized in the TSE process. It is noted that the dMMPP is equivalent to the MMPP in the continuous time domain, however, it is a more natural model for data processing owing to the number of data arrivals that happen in sampling intervals [57].

The dMMPP can be regarded as a Markov random walk where the increments in each interval have a Poisson distribution. The parameter within the Poisson distribution is considered a function of the state in the Markov chain. Thus the M-state dMMPP can be represented by a transition probability matrix $\mathbf{P}$ of the Markov chain and a matrix $\mathbf{\Lambda}$ of Poisson arrival rate. These matrices are defined as

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,M} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,M} \end{bmatrix} \text{ and } \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_M \end{bmatrix}, \tag{5.1}$$

where $p_{i,j}$ represents the transition probability from state s_i to state s_j, and $\sum_{\forall j} p_{i,j} = 1$ for $i,j \in \{1,2,\cdots,M\}$. The parameter λ_i is the mean packet arrival rate in state s_i. All the parameters of the dMMPP are inferred from the probability mass function (PMF) of the number of arriving packets in a trace duration. Let L_T and $L_{\Delta t}$ denote the length of trace duration and sampling interval respectively. Fig. 5.3 depicts the flowchart of the TMC process, which is detailed in the following paragraphs.

The procedure starts by computing the data sequence $\Upsilon = \{\varrho_1, \varrho_2, \cdots, \varrho_N\}$ wherein each component ϱ_i represents the number of arriving packets in the trace duration L_T and $N = L_T/L_{\Delta t}$. Based on the data sequence Υ, the PMF of x packets arrived in time interval $L_{\Delta t}$ can be calculated as

$$f(x) = \frac{\sum_{\forall \varrho_i = x} \varrho_i}{xN}, \quad i = 1, 2, \cdots, N. \tag{5.2}$$
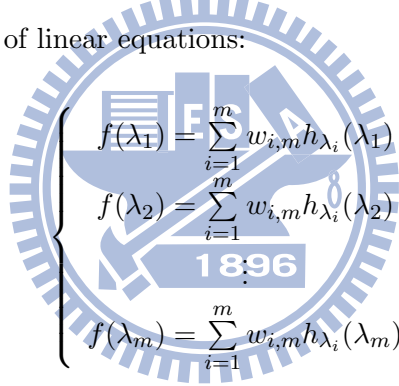
The PMF $f(x)$ represents the distribution of the traffic, which can be treated as the weighted sum of all the Poisson arrival rates of the dMMPP model. Therefore, the states and the

corresponding Poisson rates are obtained by iteratively subtracting a Poisson probability function from $f(x)$. Let $g_m(x)$ denotes a function of the difference between $f(x)$ and the weighted sum of Poisson probability function at the mth iteration, where $g_1(x) = f(x)$. In the mth iteration, the following three steps are executed.

1. *Poisson rate calculation.* The mth state of dMMPP is acquired in the mth iteration, wherein the corresponding Poisson rate is obtained as

$$\lambda_m = \arg\max_x \{g_m(x)\}. \tag{5.3}$$

2. *State probability calculation.* Based on the acquired Poisson rates $\{\lambda_1, \lambda_2, \cdots, \lambda_m\}$, the weights of each Poisson probability function $\{w_{1,m}, w_{2,m}, \cdots, w_{m,m}\}$ is derived via solving the following set of linear equations:

$$\begin{cases} f(\lambda_1) = \sum_{i=1}^{m} w_{i,m} h_{\lambda_i}(\lambda_1) \\ f(\lambda_2) = \sum_{i=1}^{m} w_{i,m} h_{\lambda_i}(\lambda_2) \\ \quad \vdots \\ f(\lambda_m) = \sum_{i=1}^{m} w_{i,m} h_{\lambda_i}(\lambda_m) \end{cases}, \tag{5.4}$$

where $h_{\lambda_i}(x)$ denotes the Poisson probability function with rate λ_i.

3. *Difference function update.* The final step in the mth iteration is to update the difference function as follows,

$$g_{m+1}(x) = f(x) - \sum_{i=1}^{m} w_{i,m} h_{\lambda_i}(x). \tag{5.5}$$

The iterative procedure will be terminated when the ratio of the maximum of $g_{m+1}(x)$ to the maximum of $f(x)$ is lower than a predefined error threshold ξ, i.e., $\max\{g_{m+1}(x)\}/\max\{f(x)\} \leq \xi$. Consequently, the number of states in the dMMPP is obtained as $M = m$ and the Poisson rates in each state s_i is λ_i, where $i \in \{1, 2, \cdots, M\}$.

The state transition probabilities of the constructed dMMPP are derived by counting the number of transitions between each pair of states. Therefore, each sampling interval will first be associated with one of the dMMPP states. Let $\psi_j(x)$ denotes the probability of x packets that arrived in time interval $L_{\Delta t}$ originated from state s_j, which can be obtained as

$$\psi_j(x) = \frac{h_{\lambda_j}(x)}{\sum_{m=1}^{M} h_{\lambda_m}(x)}. \tag{5.6}$$

The state assignment for the ith sampling interval is randomly calculated according to the probability set $\{\psi_1(\varrho_i), \psi_2(\varrho_i), \cdots, \psi_M(\varrho_i)\}$. According to the state transition within the trace duration L_T, the transition probability of the dMMPP can be obtained as

$$p_{i,j} = \frac{n_{i,j}}{\sum_{m=1}^{M} n_{i,m}} \tag{5.7}$$

where $n_{i,j}$ represents the number of transitions from state s_i to state s_j. Based on the aforementioned construction process, the matrices $\mathbf{P}$ and $\mathbf{\Lambda}$ of dMMPP can be acquired, which will be applied to the POMDP model in the following subsection.

5.3.2 Traffic State Estimation (TSE) Procedure

The TSE procedure exploits a POMDP model to conjecture the present traffic state at each decision epoch. A POMDP model is formally described as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{Z}, \mathcal{O}, \mathcal{R} \rangle$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, $\mathcal{T}$ is a set of state transition probabilities, $\mathcal{Z}$ is a set of observations, $\mathcal{O}$ is a set of observation probabilities, and $\mathcal{R}$ is a set of immediate rewards. For the proposed SSWC approach, $\mathcal{S}$ and $\mathcal{T}$ correspond to the set of dMMPP states and its state transition probability matrix $\mathbf{P}$, respectively, which are defined in the previous subsection. Since the objective of the SSWC approach is to resolve the SWS problem, the set of actions is defined as $\mathcal{A} = \{a_1, a_2, \cdots, a_N\}$ where a_i represents the action of selecting a sleep window of length T_{a_i}.

Considering a sequence of control cycle $\{C_1, C_2, \cdots, C_T\}$ in the proposed SSWC approach, the set of corresponding decision epoches is defined as $\mathcal{D} = \{d_1, d_2, \cdots, d_T\}$. Fig. 5.4 depicts
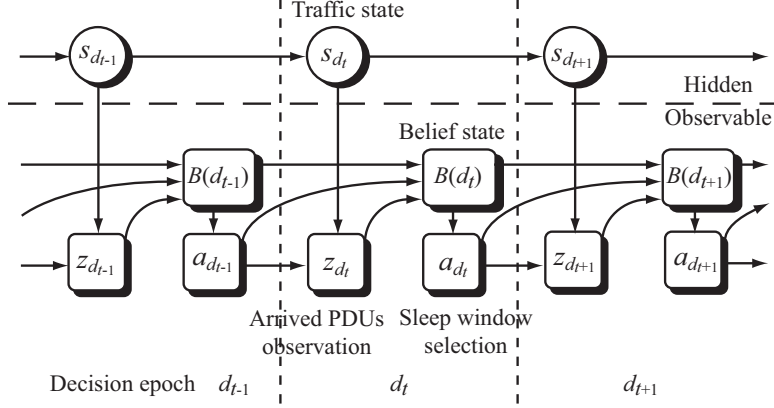
Figure 5.4: Schematic diagram of POMDP model for SSWC approach.

the schematic diagram of the POMDP model for the proposed SSWC approach. At each decision epoch $d_t \in \mathcal{D}$, the traffic state $s_{d_t} \in \mathcal{S}$ is considered unobservable. However, the number of packets that arrived during the previous control cycle C_{t-1} can be observed and acquired. Thus the set of observations is written as $\mathcal{Z} = \{z_1, z_2, \cdots, z_Q\}$ where z_i denotes the number of packets arrived in the interval between two consecutive decision epoches. The observation probability can be expressed as

$$o(a_{d_{t-1}}, s_{d_t}, z_{d_t}) \triangleq Pr(z_{d_t} | a_{d_{t-1}}, s_{d_t}) = \frac{\lambda_{d_t} T_{a_{d_{t-1}}}}{z_{d_t}} e^{-(\lambda_{d_t} T_{a_{d_{t-1}}})}, \tag{5.8}$$

which is a conditional probability of an observation $z_{d_t} \in \mathcal{Z}$ at decision epoch d_t given the action $a_{d_{t-1}} \in \mathcal{A}$ made at d_{t-1} and the present traffic state $s_{d_t} \in \mathcal{S}$ at d_t. It is intuitive that the length of sleep window can be determined based on the latest observation of arriving packets. However, it has been shown in [53] that non-optimal solution is in general obtained within the POMDP model only based on the most recent observation.

In order to acquire the optimal policy, a belief state is introduced in the POMDP model for obtaining the history of previous observations. Given a decision epoch $d_t \in \mathcal{D}$, the set of belief state is defined as $\mathcal{B}(d_t) = \{b(s_1^{d_t}), b(s_2^{d_t}), \cdots, b(s_M^{d_t})\}$, which represents the estimated probability distribution over the set of traffic states $\mathcal{S} = \{s_1, s_2, \cdots, s_M\}$. Each element $b(s_i^{d_t})$ denotes the probability of traffic state s_i at decision epoch d_t. It is noted that $b(s_i^{d_t}) \leq 0, \forall s_i \in \mathcal{S}$ and $\sum_{\forall s_i} b(s_i^{d_t}) = 1, \forall d_t \in \mathcal{D}$ since the traffic must reside in one of the states within the

$\mathcal{S}$ set at any given decision epoch. As shown in Fig. 5.4, the belief state $\mathcal{B}(d_{t+1})$ is updated at decision epoch $d_{t+1} \in \mathcal{D}$ by exploiting previous action $a_{d_t} \in \mathcal{A}$ and the corresponding observation $z_{d_{t+1}} \in \mathcal{Z}$. Thus each element $b(s_i^{d_{t+1}})$ of the belief state $\mathcal{B}(d_{t+1})$ can be derived using the Bayes rule as

$$
\begin{aligned}
b(s_i^{d_{t+1}}) &= Pr(s_i^{d_{t+1}}|\mathcal{B}(d_t), a_{d_t}, z_{d_{t+1}}) \\
&= \frac{Pr(z_{d_{t+1}}|s_i^{d_{t+1}}, a_{d_t}, \mathcal{B}(d_t))Pr(s_i^{d_{t+1}}|a_{d_t}, \mathcal{B}(d_t))}{Pr(z_{d_{t+1}}|a_{d_t}, \mathcal{B}(d_t))} \\
&= \frac{Pr(z_{d_{t+1}}|s_i^{d_{t+1}}, a_{d_t}) \sum_{s_j^{d_t} \in \mathcal{S}} Pr(s_i^{d_{t+1}}|a_{d_t}, \mathcal{B}(d_t), s_j^{d_t})Pr(s_j^{d_t}|a_{d_t}, \mathcal{B}(d_t))}{Pr(z_{d_{t+1}}|a_{d_t}, \mathcal{B}(d_t))} \\
&= \frac{Pr(z_{d_{t+1}}|s_i^{d_{t+1}}, a_{d_t}) \sum_{s_j^{d_t} \in \mathcal{S}} Pr(s_i^{d_{t+1}}|a_{d_t}, s_j^{d_t})Pr(s_j^{d_t}|\mathcal{B}(d_t))}{Pr(z_{d_{t+1}}|a_{d_t}, \mathcal{B}(d_t))} \\
&= \frac{o(s_i^{d_{t+1}}, a_{d_t}, z_{d_{t+1}}) \sum_{s_j^{d_t} \in \mathcal{S}} p_{i,j} b(s_j^{d_t})}{\sigma(\mathcal{B}(d_t), a_{d_t}, z_{d_{t+1}})},
\end{aligned}
\tag{5.9}
$$

where

$$
\sigma(\mathcal{B}(d_t), a_{d_t}, z_{d_{t+1}}) \triangleq Pr(z_{d_{t+1}}|a_{d_t}, \mathcal{B}(d_t)) = \sum_{s_j^{d_t} \in \mathcal{S}} \sum_{s_i^{d_{t+1}} \in \mathcal{S}} b(s_j^{d_t}) p_{i,j} o(s_i^{d_{t+1}}, a_{d_t}, z_{d_{t+1}}). \tag{5.10}
$$

It can be observed that the belief state is a summed statistics for the entire history of the process, which progressively incorporates the action and the corresponding observation at each decision epoch. Based on the belief states, more precise traffic states can be estimated by exploiting the proposed SSWC approach.

5.4   Sleep Window Selection Policy

In this section, two window selection policies of the SSWC approach are proposed and explained. Since the SSWC approach determines the length of each sleep window according to present traffic state, the system performance of each state/action pair will first be investigated via different evaluation metrics in subsection A. Based on the evaluation results as well as the considerations of tolerable delay and/or queue size, the suboptimal policies for resolving the

SWS problem are proposed in the SSWC approach as shown in subsection B.

5.4.1 Evaluation Metrics

Three evaluation metrics, including sleep ratio, energy cost, and packet delay, of each state/action pair (s_i, a_k), $\forall s_i \in \mathcal{S}$ and $\forall a_k \in \mathcal{A}$, are investigated in the proposed SSWC approach. All of these metrics are explained as follows.

Sleep Ratio

In order to measure the proportion of sleep windows to entire operation time for an MS in the sleep mode, the *sleep ratio* is defined as the length of a sleep window compared with the corresponding control cycle duration. Thus the expected sleep ratio of the (s_i, a_k) pair can be expressed as

$$\overline{R}(s_i, a_k) = \frac{E[T_{SW}(s_i, a_k)]}{E[T_{SW}(s_i, a_k)] + E[T_{LW}(s_i, a_k)]},\tag{5.11}$$

where $E[T_{SW}(s_i, a_k)]$ and $E[T_{LW}(s_i, a_k)]$ represent the expected length of the sleep window and the followed listening window, respectively. Since the action a_k is selected in state s_i, the length of sleep window $E[T_{SW}(s_i, a_k)] = T_{a_k}$; while $E[T_{LW}(s_i, a_k)]$ can be derived as

$$E[T_{LW}(s_i, a_k)] = (N_{\lambda_i, a_k}^{SW} + N_{\lambda_i, a_k}^{LW})E[U],\tag{5.12}$$

where N_{λ_i, a_k}^{SW} and N_{λ_i, a_k}^{LW} represent the average number of arriving packets in the sleep window and the listening window, respectively. Let U be the random variable of service time for each packet and $E[U] = 1/\mu$ is the mean service time. By applying the Little's theorem [58], N_{λ_i, a_k}^{SW} and N_{λ_i, a_k}^{LW} can be expressed as

$$N_{\lambda_i, a_k}^{SW} = \lambda_i T_{a_k},$$
$$N_{\lambda_i, a_k}^{LW} = \lambda_i E[T_{LW}(s_i, a_k)].\tag{5.13}$$

By substituting (5.13) and $E[U] = 1/\mu$ into (5.12), the length of listening window based on an action a_k made in state s_i can be obtained as

$$E[T_{LW}(s_i, a_k)] = \frac{\lambda_i T_{a_k}}{\mu - \lambda_i}. \tag{5.14}$$

With $E[T_{SW}(s_i, a_k)] = T_{a_k}$ and (5.14), the value of (5.11) can be acquired.

Energy Cost

To evaluate the power consumption for an MS in the sleep mode, the *energy cost* is defined as the average energy consumption per frame during a control cycle. Let ε_{SW} and ε_{LW} denote the energy consumption per frame within the sleep window and listening window, respectively. Moreover, the energy consumption of switching between listening window and sleep window is considered as ε_S. The expected energy cost of the (s_i, a_k) pair can be intuitively expressed as

$$\overline{E}(s_i, a_k) = \frac{2\varepsilon_S + \varepsilon_{SW} E[T_{SW}(s_i, a_k)] + \varepsilon_{LW} E[T_{LW}(s_i, a_k)]}{E[T_{SW}(s_i, a_k)] + E[T_{LW}(s_i, a_k)]}. \tag{5.15}$$

Packet Delay

Within the sleep window, the arriving packets are buffered at the BS and will be transmitted to the MS within the subsequent listening window. For the purpose of satisfying the QoS requirement in terms of delay constraint, the *packet delay* is defined as the time duration between the instant of packet arrived at the medium access control (MAC) layer in the BS and the instant of the packet completely transmitted to the MS. Since the packet arrival follows the Poisson distribution in each state and the service rate is assumed as general distribution, the $M/G/1$ queueing model is utilized to describe the packet arrival and departure. The expected packet delay of the (s_i, a_k) pair can be acquired as

$$\overline{D}(s_i, a_k) = E[D_{SW}(s_i, a_k)] + E[D_{LW}(s_i, a_k)] + E[U], \tag{5.16}$$

where $E[D_{SW}(s_i, a_k)]$ represents the average remaining length of sleep window for the packet; while the expected waiting time of the packet during the listening window is denoted as $E[D_{LW}(s_i, a_k)]$. According to $M/G/1$ with server vacation [59], $E[D_{SW}(s_i, a_k)]$ can be obtained as

$$E[D_{SW}(s_i, a_k)] = \frac{E[T_{SW}^2(s_i, a_k)]}{2E[T_{SW}(s_i, a_k)]} = \frac{T_{a_k}}{2}. \qquad (5.17)$$

On the other hand, based on Pollaczek-Khintchine mean value formula [58] and Little's theorem, $E[D_{LW}(s_i, a_k)]$ can be derived as

$$E[D_{LW}(s_i, a_k)] = \frac{\lambda_i E[U^2]}{2(1 - \rho_i)}, \qquad (5.18)$$

where $\rho_i \triangleq \lambda_i/\mu$ denotes the traffic intensity. By substituting (5.17) and (5.18) into (5.16), the expected packet delay $\overline{D}(s_i, a_k)$ can therefore be obtained.

5.4.2 Suboptimal Selection Policies

According to the aforementioned performance evaluations, two selection policies, including sleep ratio-based (SR) policy and energy cost-based (EC) policy, are proposed in the SSWC approach. For the SR policy, the immediate reward of each state/action pair is defined as the sleep ratio mentioned above; while the energy cost is utilized as the immediate reward in the EC policy. The reward set $\mathcal{R}$ of the POMDP is calculated via the *Reward Assignment Algorithm* as illustrated in Algorithm 2, where Φ indicates the type of policy, i.e, $\Phi = SR$ and $\Phi = EC$ denote the SR policy and EC policy respectively. The algorithm assigns the reward for each (s_i, a_k), $\forall s_i \in \mathcal{S}$ and $\forall a_k \in \mathcal{A}$, with the considerations of tolerable delay δ and/or queue size Q, which are explained as follows:

1. ***Tolerable delay consideration*** $(\phi = \tilde{d})$. With the consideration of delay constraint, the expected packet delay derived from (5.16) is utilized to exclude the infeasible action. As can be seen in lines 4 to 10 of Algorithm 2, if the expected delay of (s_i, a_k) satisfies the tolerable delay δ, the immediate reward $r(s_i, a_k)$ is assigned according to either

102

(5.11) or (5.15). Otherwise, a pre-defined value $\overline{R}_{min}$ or $\overline{E}_{max}$ is assigned to represent the cost for SR or EC policy, respectively.

2. **Queue size consideration** $(\phi = \tilde{q})$. On the other hand, the reward assigned with the consideration of queue size can be found in lines 13 to 19 of Algorithm 2. The average number of arriving packets obtained from (5.13) is considered as a criterion in this case, while all the other processes of reward assignment are the same as that mentioned in the case of tolerable delay consideration.

3. **Both of the delay and queue considerations** $(\phi = \tilde{b})$. In the case of considering both the tolerable delay and queue size, all the aforementioned processes are conducted to assign proper rewards for all state/action pairs, i.e., lines 3 to 20 in Algorithm 2.

Algorithm 2: *Reward Assignment Algorithm*

Input: $\mathcal{S}$, $\mathcal{A}$, tolerable delay δ, queue size Q, policy type Φ, and consideration type ϕ
Output: set of immediate rewards $\mathcal{R}(\mathcal{S}, \mathcal{A})$

1 **foreach** $s_i \in \mathcal{S}$ **do**
2   **foreach** $a_k \in \mathcal{A}$ **do**
3   **if** $(\phi = \tilde{d})$ *or* $(\phi = \tilde{b})$ **then**
4   **if** $\overline{D}(s_i, a_k) \leq \delta$ **then**
5                  $r(s_i, a_k) \leftarrow \overline{R}(s_i, a_k)$ for $\Phi = SR$
6   (or $r(s_i, a_k) \leftarrow \overline{E}(s_i, a_k)$ for $\Phi = EC$)
7   **else**
8                  $r(s_i, a_k) \leftarrow \overline{R}_{min}$ for $\Phi = SR$
9   (or $r(s_i, a_k) \leftarrow \overline{E}_{max}$ for $\Phi = EC$)
10   **end**
11   **end**
12   **if** $(\phi = \tilde{q})$ *or* $(\phi = \tilde{b})$ **then**
13   **if** $N_{\lambda_i, a_k}^{SW} + N_{\lambda_i, a_k}^{LW} \leq Q$ **then**
14                  $r(s_i, a_k) \leftarrow \overline{R}(s_i, a_k)$ for $\Phi = SR$
15   (or $r(s_i, a_k) \leftarrow \overline{E}(s_i, a_k)$ for $\Phi = EC$)
16   **else**
17                  $r(s_i, a_k) \leftarrow \overline{R}_{min}$ for $\Phi = SR$
18   (or $r(s_i, a_k) \leftarrow \overline{E}_{max}$ for $\Phi = EC$)
19   **end**
20   **end**
21   **end**
22 **end**

The optimal policy of the SWS problem is unavailable owing to the reason of uncertain traffic states. However, thanks to the belief states of POMDP model, the uncertain traffic states can be estimated. Moreover, the reward of an action made in a given traffic state is provided by the immediate reward set $\mathcal{R}$. Based on these two types of information, the sub-optimal policies can be acquired via adopting a T-step value function in the SSWC approach. For the SR policy, the decision of window selection made at decision epoch $d_t \in \mathcal{D}$ can be acquired as

$$D_{d_t}^{SR}(b(s_i^{d_t})) = \arg V_{d_t}^{SR}(b(s_i^{d_t})) = \arg \max_{a_k^{d_t} \in \mathcal{A}} [\Gamma_{d_t}(b(s_i^{d_t}))]. \tag{5.19}$$

where

$$\Gamma_{d_t}(b(s_i^{d_t})) = \sum_{s_i^{d_t} \in \mathcal{S}} b(s_i^{d_t}) r(s_i^{d_t}, a_k^{d_t}) + \gamma^{d_{t+1}} \sum_{z_j^{d_{t+1}} \in \mathcal{Z}} \sigma(\mathcal{B}(d_t), a_k^{d_t}, z_j^{d_{t+1}}) V_{d_{t+1}}(b(s_i^{d_{t+1}})), \tag{5.20}$$

with $r(s_i^{d_t}, a_k^{d_t})$ selected as either $\overline{R}(s_i, a_k)$ or $\overline{R}_{min}$ according to Algorithm 2 for the SR scheme. The function $V_{d_t}^{SR}(b(s_i^{d_t}))$ in (5.19) is defined as the T-step value function for the SR policy at a decision epoch d_t which starts at d_t, and there are $T-1$ decision steps remaining. The first item of (5.20) denotes the reward for a belief state $b(s_i^{d_t}) \in \mathcal{B}(d_t)$. The expected reward of the future belief state $b(s_i^{d_{t+1}}) \in \mathcal{B}(d_{t+1})$ is represented in the second item, where $\sigma(\mathcal{B}(d_t), a_k^{d_t}, z_j^{d_{t+1}})$ can be obtained from (5.10). The parameter $\gamma^{d_{t+1}}$ is denoted as a discount factor of the d_{t+1}-step for convergence control of the value function. In other wards, the value function $V_{d_t}^{SR}(b(s_i^{d_t}))$ intends to find the maximum sleep ratio of an action according to the currently estimated traffic state and the expected sleep ratio of future actions made in the successive states. On the other hand, for the EC policy, the decision made at d_t and its corresponding T-step value function can be obtained as

$$D_{d_t}^{EC}(b(s_i^{d_t})) = \arg V_{d_t}^{EC}(b(s_i^{d_t})) = \arg \min_{a_k^{d_t} \in \mathcal{A}} [\Gamma_{d_t}(b(s_i^{d_t}))]. \tag{5.21}$$

where the function $\Gamma_{d_t}(b(s_i^{d_t}))$ is defined in (5.20) with $r(s_i^{d_t}, a_k^{d_t})$ selected as either $\overline{E}(s_i, a_k)$ or $\overline{E}_{max}$ based on Algorithm 2 for the EC scheme. Unlike that for the SR policy, minimum reward is utilized in the value function $V_{d_t}^{EC}(b(s_i^{d_t}))$ for the EC case in order to minimize the energy consumption of the MS.

5.5 Performance Evaluation

In this section, simulations are conducted to evaluate the performance of the proposed SSWC approach in comparison with the conventional PSC I and the evolutional PSC I. A single BS/MS pair with non-real-time downlink traffic is considered as the simulation scenario. The simulation is implemented via MATLAB event-driven simulator. All the required procedures and functions for sleep mode operation have been implemented in the simulator. The parameters adopted within the simulation are listed in Table 5.1, where the energy consumption parameters are acquired from the industrial manufactured mobile WiMAX chip [49]. Each simulation run is executed with 13 minutes, in which the MS stays in the normal mode during the first 3 minutes and starts the sleep mode operation at the 4th minute. Each obtained results is average from 100 simulation runs. For the proposed SSWC approach, the traffic trace of the first 3 minutes within the simulation are utilized to infer the number of traffic states as described in Section 5.3.1. Moreover, the 3-step value function is considered in the simulation, wherein the discount factor γ for the future ith step is selected as $(0.5)^i$ for $i \in \{1, 2\}$. It is represented that the rewards calculated later in time will have less value than an equivalent reward received closer to the present.

Fig. 5.5 illustrates the sleep mode operations of the conventional PSC I (denoted as IEEE 802.16e), the evolutional PSC I (denoted as IEEE 802.16m), the SSWC approach with SR policy (denoted as SSWC-SR), and the SSWC approach with EC policy (denoted as SSWC-EC) under traffic state of $\lambda = 0.1$. The top subplot shows the arriving packets in frames $[37400, 37700]$ of the simulation; while the corresponding operations of all the schemes are depicted in the remainder subplots. For each approach, the state of 1 denotes that the MS is awake in the normal mode or during the listening window of the sleep mode; while the MS

Table 5.1: Simulation parameters for SSWC approach

| Parameter	Value
Frame duration	5 ms
Idle period (IEEE 802.16e)	4 frame
Default Listening Window length (IEEE 802.16e/m)	1 frame
Initial-Sleep Window length (IEEE 802.16e/m)	1 frame
Final-Sleep Window length (IEEE 802.16e/m)	256 frame
Sleep Window length (SSWC)	[1, 256] frame
Energy consumption of busy frame	280 mW
Energy consumption of idle frame	120 mW
Energy consumption of sleep frame	10 mW
Energy consumption of state switch	1 mW
Mean packet arrival rate (λ)	0.1, 0.2, 0.3, 0.4, 0.5 packets/frame
Mean service rate (μ)	3 packets/frame
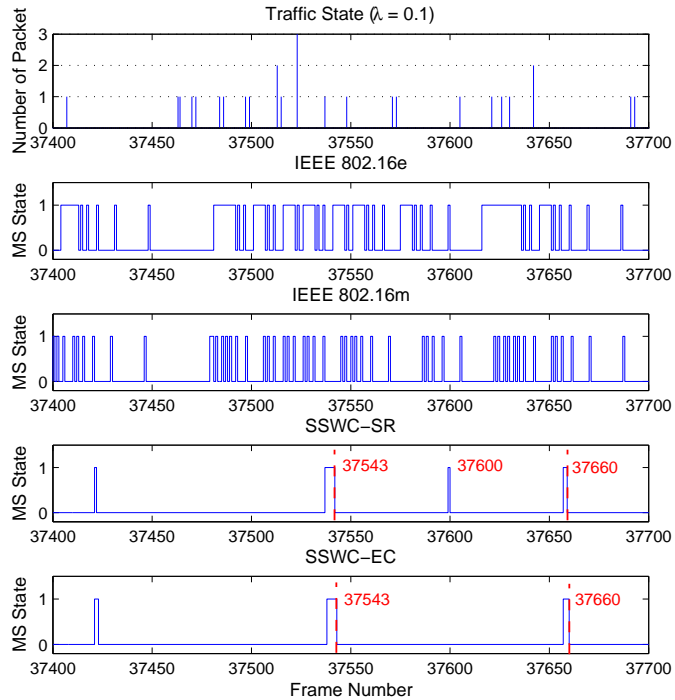
Figure 5.5: An exemplified sleep mode operation among IEEE 802.16e, IEEE 802.16m, SSWC-SR, and SSWC-EC approaches under traffic state of $\lambda = 0.1$.
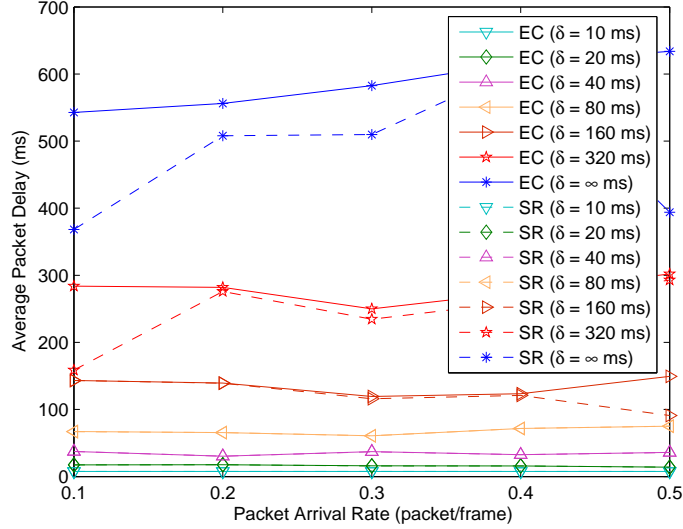
Figure 5.6: Performance comparison of average packet delay between the two proposed SSWC approaches (i.e., EC and SR) with different delay constraints (δ).

stays in the sleep state is represented by the stat of 0. It can be observed that the MS spends lots of time in the awake state while the binary-exponential power-saving mechanism of IEEE 802.16e or IEEE 802.16m is adopted. With the proposed SSWC approach, on the other hand, the MS almost remains in the sleep state and only wakes up to receive the incoming traffic since each sleep period is determined according to the present traffic state estimated via the POMDP. The performance comparison among these schemes are discussed in the following subsections.

5.5.1 Effect of Delay Constraints

The effect of delay constraints is discussed via the performance comparisons as shown in Figs. 5.6 to 5.8. The performance of average packet delay between the two proposed SSWC approaches are illustrated in Fig. 5.6, where EC and SR denote the approaches of SSWC-EC and SSWC-SR respectively. Due to the proposed reward assignment and accurate state estimate, the performance of all the schemes satisfy the respective delay constraints over various mean packet arrival rates. As for the cases that do not consider any delay constraints, the average packet delay is bounded by the maximum length of sleep window, i.e., 640 ms

for the window of length 256 frames in the simulation. The SSWC-EC approach has relative higher packet delay than the SSWC-SR scheme resulted from the different definition of reward functions. For the reward that is defined as the sleep ratio (i.e., the SSWC-SR approach), the length of sleep window that results in the highest reward value will be selected; while the window with lowest reward value is chosen in the SSWC-EC approach in order to minimize the energy cost. These two policies result in different sleep mode operations even if they can possess same ratio of listening frames to sleep frames. For example, considering the duration of frames $[37543, 37660]$ as shown in Fig. 5.5, there are 3 listening frames and 115 sleep frames that appeared in both schemes. In the SSWC-EC approach, one listening frame occurs at frame number 37600 while two happen at numbers 37659 and 37660. The SSWC-SR scheme, on the other hand, has all its three listening frames occur at frame numbers 37558, 37559, and 37660. Therefore, it can be observed that both schemes result in the same sleep ratio of $115/118 = 0.975$ during this considered time interval. It is intuitive to see that the SSWC-EC approach intends to minimize the energy consumption such as to accumulate all the three listening frames in order to reduce the switching cost ε_S between listening and sleep windows. Nevertheless, the target of the SSWC-SR approach is to minimize the sleep ratio which results in the occurrence of first listening frame at frame 37600. The sleep ratio becomes $57/58 = 0.983$ with the length of sleep window as 57 frames and one frame for listening window, which is higher than 0.975 by adopting the SSWC-EC scheme. Consequently, as shown in Fig. 5.6, the performance with lower packet delay can be obtained in the various cases by exploiting the SSWC-SR approach.

Fig. 5.7 depicts the performance comparison of average energy consumption between the two proposed SSWC approaches. It is expected that the average energy consumption increases as the value of mean packet arrival rate is augmented in all the different cases. With lower tolerable delay, more energy consumption is observed in both of the proposed SSWC approaches. This can be attributed to the reason that the length of sleep window is bounded by the delay constraint. In the sleep mode, the MS is provided with a series of alternated sleep windows and listening windows. For the purpose of satisfying tight tolerable delay, the shorter
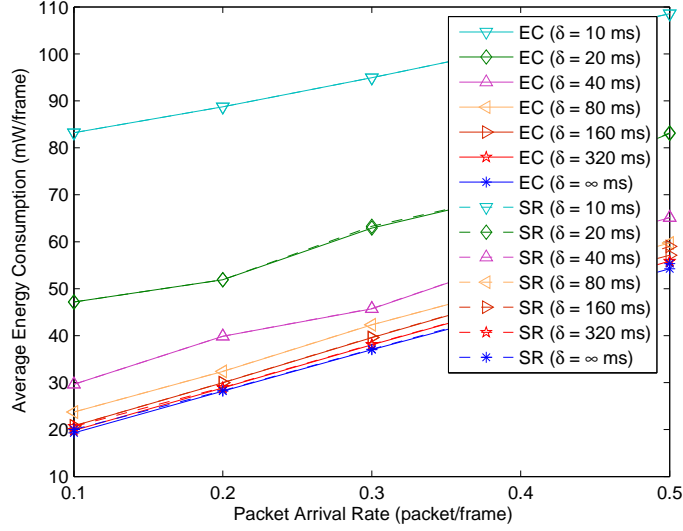
Figure 5.7: Performance comparison of average energy consumption between the two proposed SSWC approaches (i.e., EC and SR) with different delay constraints (δ).

sleep windows will be selected which incurs more number of listening windows during the simulation time, and consequently increases the energy consumption. Comparing the SSWC-EC and SSWC-SR approaches, almost the same level of energy consumption is obtained since similar ratio of listening frames to sleep frames exists in both schemes. However, the shorter sleep windows selected in the SSWC-SR scheme incurs more number of state transition, i.e., from sleep state to awake state and vice versa. Therefore, slightly higher energy consumption is shown in the SSWC-SR approach compared to the SSWC-EC scheme.

The performance comparisons among IEEE 802.16e, IEEE 802.16m, and the proposed SSWC approach over various packet arrival rates is shown in Fig. 5.8. Since similar performances are exhibited in the proposed SSWC-EC and SSWC-SR schemes with different packet delay consideration, the SSWC-EC approach is utilized on behalf of the proposed SSWC scheme to compare with the convnetional binary-exponential power-saving mechanisms. For the IEEE 802.16e and IEEE 802.16m schemes, it is expected that the average packet delay will be decreased as the packet arrival rate is augmented; meanwhile, the energy consumption is raised. The IEEE 802.16m has slightly higher delay but significant decrement of energy consumption compared to the IEEE 802.16e, which can be attributed to the elimination of
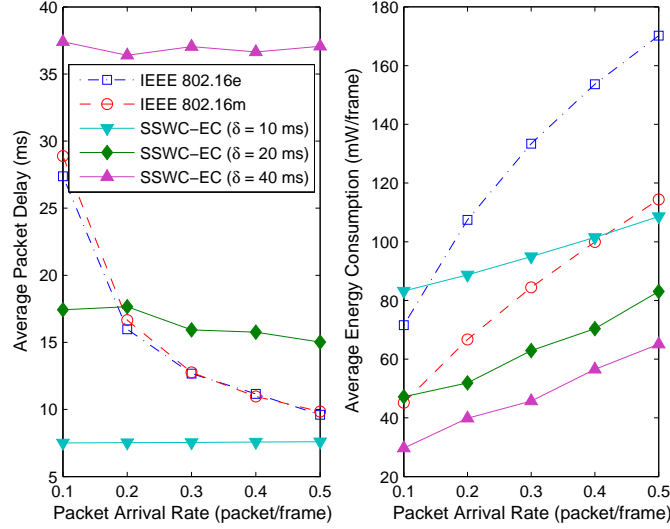
109

Figure 5.8: Performance comparison among IEEE 802.16e, IEEE 802.16m, and SSWC-EC approaches with different delay constraints (δ).

frequent state transitions. As mentioned in Section 5.2.2, in the IEEE 802.16m, the MS terminates the listening window and returns to the sleep state immediately after the completion of data transportation. Without the existence of idle periods in the IEEE 802.16m, energy conservation can be preserved. However, the opportunities to promptly receive the incoming traffic are dissolved and consequently incurs packet transmission delay. Comparing between the proposed SSWC-EC approach and the IEEE 802.16m scheme, relative better performance can be observed in the SSWC-EC approach. It can be seen that the energy consumption of the SSWC-EC ($\delta = 20$ ms) approach is similar to that of the IEEE 802.16m at $\lambda = 0.1$, but the SSWC-EC ($\delta = 20$ ms) approach can provide significantly lower packet delay. It is owing to the reason that the SSWC approach determines the length of each sleep window according to the traffic state and delay constraint. Due to the same reason, as similar packet delays existed in both schemes, the SSWC approach will result in lower energy consumption compared to the IEEE 802.16m scheme, e.g., the case with the SSWC-EC ($\delta = 20$ ms) scheme at $\lambda = 0.2$.
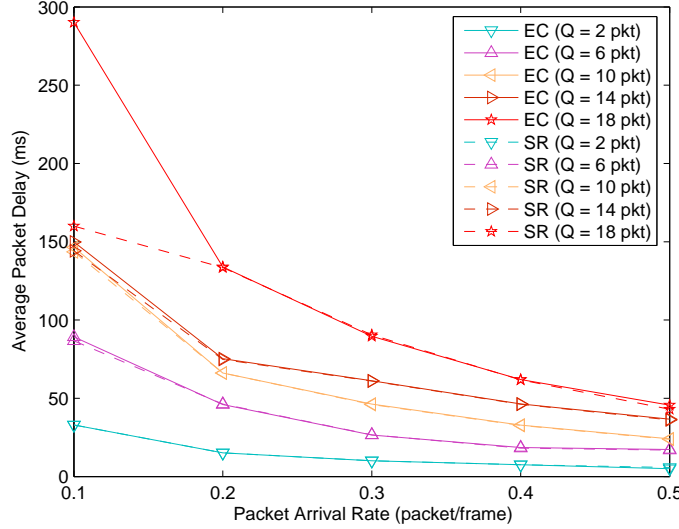
Figure 5.9: Performance comparison of average packet delay between the two proposed SSWC approaches (i.e., EC and SR) with different queue size considerations (Q).

5.5.2 Effect of Queue Length Considerations

Figs. 5.9 to 5.12 illustrate the effect of queue size considerations among the power-saving mechanisms. The performance comparisons of average packet delay among the proposed SSWC approaches are shown in Fig. 5.9. Given a traffic state, the larger queue size results in higher packet delay since the longer sleep window is selected either to maximize the sleep ratio in the SSWC-SR approach or to minimize the energy consumption in the SSWC-EC scheme. Since the number of arriving packets is increased with the augmentation of packet arrival rate, shorter sleep window is considered to satisfy constant size of queue, which consequently reduces the average packet delay. The SSWC-EC approach possesses higher packet delay than the SSWC-SR scheme, which can be attributed to the same reason as explained in Section 5.5.1.

Fig. 5.10 shows the performance of average energy consumption between the two proposed SSWC approaches. It is expected that the average energy consumption increases as the value of mean packet arrival rate is augmented in all the cases. Under the situations with smaller queue size, more energy consumption is required in the proposed SSWC approaches
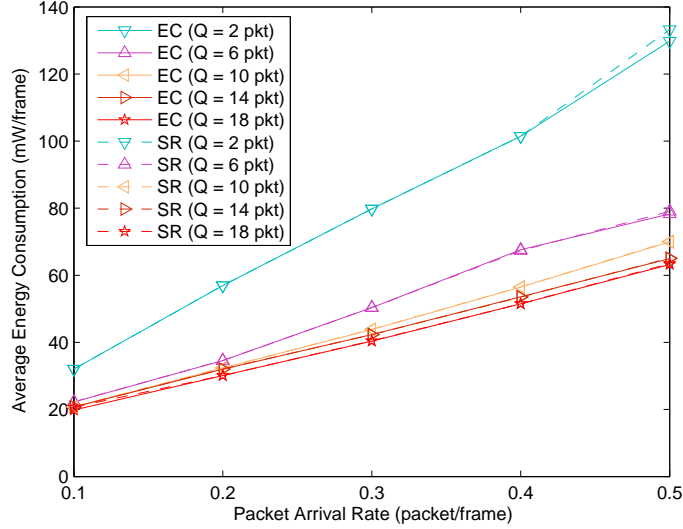
Figure 5.10: Performance comparison of average energy consumption between the two proposed SSWC approaches (i.e., EC and SR) with different queue size considerations (Q).

since the length of sleep window will be limited by both the traffic state and queue size. Considering the scenario of constant traffic state, shorter sleep windows will be selected with the consideration of smaller queue size, which results in additional number of listening windows during the simulation time and hence increases the energy consumption. Owing to similar ratio of listening frames to sleep frames, both the SSWC-EC and SSWC-SR approaches possess similar energy consumption as shown in Fig. 5.10.

Fig. 5.11 illustrates the performance of packet overflow under various queue size considerations. The *packet overflow* is defined as the number of packets exceeding the pre-defined queue size compared to the total number of packets in the simulation. It is observed that both the IEEE 802.16e and IEEE 802.16m schemes result in comparatively high packet overflow under the consideration of queue size less than 5 packets. This can be attributed to the reason that the bursty packets frequently incur the overflowing packets while the queue size is small. The packet overflow decreases as the increment of queue size in both the IEEE 802.16e and IEEE 802.16m schemes since the large queue size provides the capability for enduring the bursty situation. On the other hand, the SSWC approaches possess relative lower packet overflow owing to the reason that the length of each sleep window is determined according to
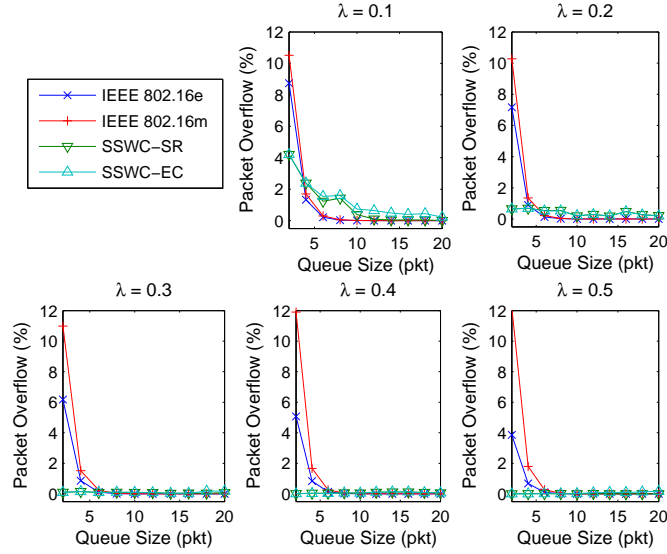
Figure 5.11: Performance comparison of packet overflow among IEEE 802.16e, IEEE 802.16m, SSWC-SR, and SSWC-EC approaches under various queue size considerations (Q).

the traffic state and considered queue size.

Performance comparisons among IEEE 802.16e, IEEE 802.16m, and the proposed SSWC-EC approach over various packet arrival rates is shown in Fig. 5.12. The SSWC-EC approach is adopted on behalf of the proposed SSWC scheme since both the EC and SR policies result in similar performance under the consideration of queue size. Comparing to the IEEE 802.16e and IEEE 802.16m schemes, lower energy consumption can be acquired by adopting the SSWC approach. The SSWC-EC approaches result in comparatively high packet delay since the length of each sleep window is selected according to the traffic state and considered queue size. The MS wakes up as the number of buffered packet is close to the considered queue length, which consequently increases the packet delay. Nevertheless, the primary objective of proposed SSWC approach to reduce energy consumption of the MS can still be achieved. The merits of the proposed schemes can be observed.
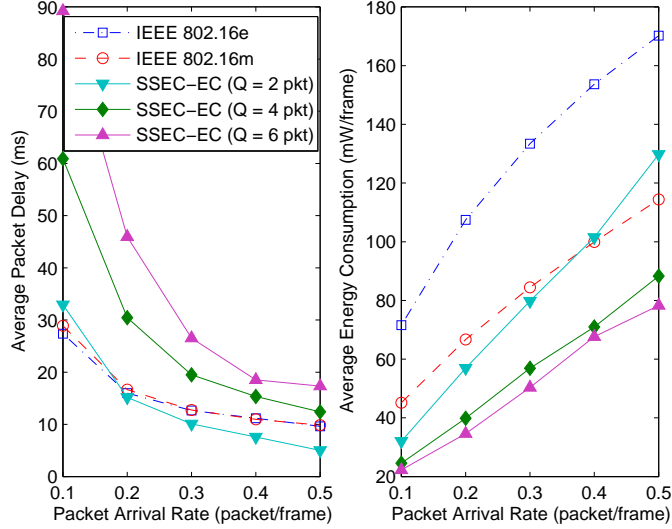
Figure 5.12: Performance comparison among IEEE 802.16e, IEEE 802.16m, and SSWC-EC approaches with different queue size considerations (Q).

5.6 Concluding Remarks

In this chapter, a statistical sleep window control (SSWC) approach for sleep mode operation is proposed to maximize the energy efficiency of an MS with non-real-time downlink traffic. Following the notions of IEEE 802.16m system, the sleep mode operation is formulated as a sleep window selection (SWS) problem. The SSWC approach selects the length of each sleep window based on present traffic state and the considerations of tolerable delay and/or queue size. The traffic model construction (TMC) procedure of SSWC approach formulates a discrete-time Markov-modulated Poisson process (dMMPP) for non-real-time downlink traffic via analyzing the traffic trace, which infers all the states of the traffic. Based on the constructed dMMPP, a partially observable Markov decision process (POMDP) is adopted in the traffic state estimation (TSE) procedure of SSWC approach to conjecture the present traffic state at each decision epoch. By exploiting the properties of the POMDP, two suboptimal policies, including the sleep ratio-based (SR) policy and energy cost-based (EC) policy, with the considerations of tolerable delay and/or queue size are proposed within the SSWC approach to resolve the SWS problem. The efficiency of proposed SSWC approach is evaluated

and compared via simulations. Simulation studies show that the proposed SSWC approach outperforms the conventional IEEE 802.16e PSC I and the evolutional PSC I of IEEE 802.16m system in terms of both energy efficiency and packet delay while the delay constraint is considered. On the other hand, the SSWC approach aggregates the buffered packets as much as possible while satisfying the pre-defined queue size, which maximizes the energy conservation of the MS. With the considerations of both tolerable delay and queue size, an efficient sleep mode operation for mobile broadband wireless networks can be acquired by adopting the proposed SSWC approach.

Chapter 6

Conclusions

In this dissertation, four topics on IEEE 802.16-based mobile broadband wireless networks (MBWNs) are presented. The contributions of this dissertation involve developments of flexible and contention-free adaptive communication approach and its corresponding scheduling algorithm for mobile stations (MSs), and two power-saving mechanisms for real-time and non-real-time traffic respectively. The contributions for each of the four focused areas are summarized as follows:

- **Chapter 2:** A flexible and contention-free adaptive point-to-point communication (APC) approach is proposed to achieve direct transmission between SSs within IEEE 802.16 PMP networks. According to the relative locations and channel conditions among the BS and SSs, the packet transmission operation is switched between the direct link and indirect links in the APC approach. A comprehensive architectural design associated with the extended frame structure for the proposed APC approach is developed to be backward-compatible with the IEEE 802.16 standards. The advantages of exploiting the PDC approach is that both the required bandwidth for packet transmission and packet-rerouting delay for intra-cell traffic can be significantly reduced.

- **Chapter 3:** A scheduling algorithm for each pair of MSs that is expected to conduct direct communication is proposed. Both the interference region and feasible region for the pair of MSs to perform direct communication are studied and calculated. Based

116

on these two types of information, the predictive interference-based scheduling (PIS) algorithm properly arranges the MSs to conduct direct communication in either DL subframe or UL subframe without increasing additional interference for other communications. Furthermore, a prediction mechanism to estimate the further positions of MSs is proposed. Combining the predication mechanism with the PIS algorithm, a predictive motion and interference-based scheduling (PMIS) algorithm is given to reduce the control overheads regarding the position updates of MSs.

- **Chapter 4:** A cycle-based power-saving mechanism with adaptive listening window (ALW) is proposed for IEEE 802.16m sleep mode operation with real-time traffic. The ALW scheme adaptively adjusts the ratio of listening window to sleep window based on both the buffered packets and delay constraints, which results in that both the energy consumption and packet loss rate for real-time traffic are significant reduced. Moreover, the concept of the proposed ALW approach has been approved and included within the new standard draft by the IEEE 802.16m Task Group.

- **Chapter 5:** A statistical sleep window control (SSWC) approach for sleep mode operation is proposed to maximize the energy efficiency of an MS with non-real-time downlink traffic. The proposed approach constructs a discrete-time Markov-modulated Poisson process for representing the states of the non-real-time traffic. Furthermore, a partially observable Markov decision process is exploited to conjecture the present traffic state. Based on the estimated traffic state and the considerations of tolerable delay and/or queue size, two suboptimal policies for maximizing energy conservation are proposed within the approach, which results in an efficient sleep mode operation for MBWNs.

Bibliography

[1] R. Resnick, "WiMAX is fast mobile and it's here now!" *Presented at WiMAX Forum Congress Global*, Jun. 2008.

[2] W. Kim, "Mobile WiMAX, the leader of the mobile Internet era," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 11–12, Jun. 2009.

[3] K. Etemad, "Overview of mobile WiMAX technology and evolution," *IEEE Commun. Mag.*, vol. 46, no. 10, pp. 31–40, Oct. 2008.

[4] *IEEE Standard for Local and Metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Stdandard 802.16-2001, Apr. 2002.

[5] *IEEE Standard for Local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz*, IEEE Standard 802.16a-2003, Apr. 2003.

[6] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interference for Fixed Broadband Wireless Access Systems*, IEEE Standard 802.16-2004, Oct. 2004.

[7] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interference for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, IEEE Standard 802.16e-2005, Feb. 2006.

[8] *IEEE Standard for Local and metropolitan area networks - Part 16: Air Interface for Broadband Wireless Access Systems*, IEEE Standard 802.16-2009, May 2009.

[9] Z. Abichar, Y. Peng, and J. M. Chang, "WiMAX: The emergence of wireless broadband," *IEEE IT Prof.*, vol. 8, no. 4, pp. 44–48, Jul.-Aug. 2006.

[10] *Draft Standard for Information Technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, Amendment 6: Extensions to Direct Link Setup (DLS)*, IEEE P802.11z^{TM}/D5.0, Jun. 2009.

[11] W. Zhang, H. Zhu, and G. Cao, "Improving Bluetooth network performance through a time-slot leasing approach," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Orlando, FL, 2002, pp. 592–596.

[12] C. Cordeiro, S. Abhyankar, and D. P. Agrawal, "A dynamic slot assignment scheme for slave-to-slave and multicast-like communication in Bluetooth personal area networks," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, San Francisco, CA, 2003, pp. 4127–4132.

[13] *Specification of the Bluetooth system*, Bluetooth SIG, 2004.

[14] H.-C. Yin, Y.-G. Jan, S.-T. Sheu, H.-I. Hsu, Y.-H. Tsai, C.-C. Hsieh, T.-H. Lo, and F. C.-D. Tsai, "Virtual direct link access for IEEE 802.16 wireless metropolitan area networks (WMANs)," in *Proc. IEEE Vehicular Technology Conf. (VTC-Spring)*, Dublin, Ireland, 2007, pp. 2981–2984.

[15] *Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interference for Fixed Broadband Wireless Access Systems: Multihop Relay Specification*, IEEE P802.16j/D3, 2008.

[16] L.-C. Chu, P.-H. Tseng, and K.-T. Feng, "GDOP-assisted location estimation algorithms

in wireless location systems," in *Proc. IEEE Global Communication Conf. (GLOBE-COM)*, New Orleans, LA, 2008, pp. 1–5.

[17] P.-H. Tseng and K.-T. Feng, "Location tracking assisted handover algorithms for broadband wireless networks," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, Cannes, France, 2008, pp. 1–5.

[18] *IEEE 802.16m Evaluation Methodology Document (EMD)*, IEEE C802.16m-08/004r5, Jan. 2009.

[19] M. Shreedhar and G. Varghese, "Efficient fair queueing using deficit round robin," *IEEE/ACM Trans. Netw.*, vol. 4, no. 3, pp. 375–385, Jun. 1996.

[20] M. Katevenis, S. Sidiropoulos, and C. Courcoubetis, "Weighted round-robin cell multiplexing in a general-purpose atm switch chip," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 8, pp. 1265–1279, Oct. 1991.

[21] A. Goldsmith, *Wireless Communication*. New York, NY: Cambridge University Press, 2005.

[22] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA: The MIT. Press., 1974.

[23] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proc. ACM Int. Workshop Modeling, Analysis and Simulation of Wireless and Mobility Systems (MSWiM)*, Seattle, WA, 1999, pp. 53–60.

[24] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for multidimensional PCS networks," vol. 11, no. 5, pp. 718–732, Oct. 2003.

[25] S. L. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[26] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken, NJ: John Wiley & Sons Inc., 2003.

[27] *IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11-2007, Jun.

[28] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UE Procedures in Idle Mode and Procedures for Cell Reslection in Connected Mode*, Technical Specification 3GPP TS 25.304, version 5.1.0, 2002.

[29] S.-R. Yang, S.-Y. Yan, and H.-N. Hung, "Modeling UMTS power saving with bursty packet data traffic," *IEEE Trans. Mobile Comput.*, vol. 6, no. 12, pp. 1398–1409, Dec. 2007.

[30] *Power Saving Management for IEEE 802.16m*, IEEE C802.16m-08/567r1, 2008.

[31] Y. Xiao, "Energy saving mechanism in the IEEE 802.16e wireless MAN," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 595–597, Jul. 2005.

[32] L. Kong and D. H. K. Tsang, "Performance study of power saving classes of type I and II in IEEE 802.16e," in *Proc. IEEE Conf. Local Computer Networks (LCN)*, Tampa, FL, 2006, pp. 20–27.

[33] K. Han and S. Choi, "Performance analysis of sleep mode operation in IEEE 802.16e mobile broadband wireless access systems," in *Proc. Vehicular Technology Conf. (VTC-Spring)*, Melbourne, AU, 2006, pp. 1141–1145.

[34] Y. Zhang and M. Fujise, "Energy management in the IEEE 802.16e MAC," *IEEE Commun. Lett.*, vol. 10, no. 4, pp. 311–313, Apr. 2006.

[35] Y. Xiao, "Performance analysis of an energy saving mechanism in the IEEE 802.16e wireless MAN," in *Proc. IEEE Consumer Communications and Networking Conf. (CCNC)*, Las Vegas, NV, 2006, pp. 406–410.

[36] Y. Zhang, "Performance modeling of energy management mechanism in IEEE 802.16e mobile WiMAX," in *Proc. Wireless Communications and Networking Conf. (WCNC)*, Hong Kong, CN, 2007, pp. 3205–3209.

[37] Y.-P. Hsu and K.-T. Feng, "Performance modeling of power saving classes with multiple connections for broadband wireless networks," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Las Vegas, NV, 2008, pp. 1477–1482.

[38] J.-B. Seo, S. Lee, N.-H. Park, H.-W. Lee, and C.-H. Cho, "Performance analysis of sleep mode operation in IEEE 802.16e," in *Proc. IEEE Vehicular Technology Conf. (VTC-Fall)*, Los Angeles, CA, 2004, pp. 1169–1173.

[39] Y. Park and G. U. Hwang, "Performance modelling and analysis of the sleep-mode in IEEE802.16e WMAN," in *Proc. IEEE Vehicular Technology Conf. (VTC-Spring)*, Dublin, IE, 2007, pp. 2801–2806.

[40] Z. Huo, W. Yue, S. Jin, and N. Tian, "Modeling and performance evaluation for the sleep mode in the IEEE 802.16e wireless networks," in *Proc. IEEE Int. Conf. Communications Systems (ICCS)*, Guangzhou, CN, 2008, pp. 1140–1144.

[41] S. Alouf, E. Altman, and A. P. Azad, "Analysis of an M/G/1 queue with repeated inhomogeneous vacations with application to IEEE 802.16e power saving mechanism," in *Proc. Int. Conf. Quantitative Evaluation Systems (QEST)*, St Malo, FR, 2008, pp. 27–36.

[42] J. Xiao, S. Zou, B. Ren, and S. Cheng, "An enhanced energy saving mechanism in IEEE 802.16e," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, San Francisco, CA, 2006, pp. 1–5.

[43] M. Peng and W. Wang, "An adaptive energy saving mechanism in the wireless packet access network," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Las Vegas, NV, 2008, pp. 1536–1540.

[44] J.-R. Lee and D.-H. Cho, "Performance evaluation of energy-saving mechanism based on probabilistic sleep interval decision algorithm in IEEE 802.16e," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1773–1780, Jul. 2007.

[45] M.-G. Kim, M. Kang, and J. Y. Choi, "Remaining energy-aware power management mechanism in the 802.16e MAC," in *Proc. IEEE Consumer Communications and Networking Conf. (CCNC)*, Las Vegas, NV, 2008, pp. 222–226.

[46] M.-G. Kim, J. Y. Choi, and M. Kang, "Adaptive power saving mechanism considering the request period of each initiation of awakening in the IEEE 802.16e system," *IEEE Commun. Lett.*, vol. 12, no. 2, pp. 106–108, Feb. 2008.

[47] S. Perera and H. Sirisena, "Analysis of packet loss for real-time traffic in wireless mobile networks with ARQ feedback," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Atlanta, GE, 2004, pp. 417–422.

[48] Z. Quan and J.-M. Chung, "Contention based negative feedback ARQ for VoIP services in IEEE 802.16 networks," in *Proc. IEEE Internal Conf. on Networks (ICON)*, Singapore, 2006, pp. 1–6.

[49] S. Communications, "Datasheet: SQN1130 System-on-Chip (SoC) for WiMAX mmobile stations."

[50] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY: McGraw-Hill, 2002.

[51] *IEEE 802.16m System Description Document*, IEEE C802.16m-08/003r8, Apr. 2009.

[52] *IEEE 802.16m System Requirements*, IEEE C802.16m-08/002r8, Jan. 2009.

[53] A. C. Cassandra, "Exact and approximate algorithms for partially observable Markov decision processes," Ph.D. dissertation, Brown Univ., Providence, RI, May 1998.

[54] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially

observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1-2, pp. 99–134, May 1998.

[55] A. Andersen and B. Nielsen, "A Markovian approach for modeling packet traffic with long-range dependence," *IEEE J. Sel. Areas Commun.*, vol. 12, no. 5, pp. 719–732, Jun. 1998.

[56] T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process," vol. 17, no. 1-2, pp. 185–211, 2001.

[57] A. Nogueira, P. Salvador, R. Valadas, and A. Pacheco, "Fitting self-similar traffic by a superposition of MMPPs modeling the distribution at multiple time scales," vol. E87-B, no. 3, pp. 678–688, Mar. 2004.

[58] H. Kobayashi and B. L. Mark, *System Modeling and Analysis: Foundations of System Performance Evaluation.* Upper Saddle River, NJ: Prentice Hall, 2009.

[59] C.-H. Ng and B.-H. Soong, *Queueing Modelling Fundamentals: with Applications in Communication Networks*, 2nd ed. Hoboken, NJ: John Wiley and Sons, 2008.