POWER CONTROL AND SCHEDULING FOR WIRELESS DATA COMMUNICATIONS

WANG QI

NATIONAL UNIVERSITY OF SINGAPORE

2004

POWER CONTROL AND SCHEDULING FOR WIRELESS DATA COMMUNICATIONS

WANG QI

(M. Eng. USTC)

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF ENGINEERING
DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE
2004

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Dr. Francois Chin and Associate Professor Chi Chung Ko, for their invaluable advice, guidance and tolerance through the years, and most of all, for the freedom they allowed me to pursue the research I want to do.

I am also grateful to Dr. Lei Zhongding, for his invaluable advice and informative discussion with me which make great impact on my research.

My gratitude also goes to National University of Singapore who provides me a perfect research environment and grants the use of the facilities.

I am also grateful to my family and friends for their support, care, encouragement and understanding during the course of this research. They are always the strongest backing what kept me going on during the tough times.

Finally, my thanks also go to all of my friends in Open Source Software Lab. The friendship with them made my study and life in NUS fruitful and enjoyable.

TABLE OF CONTENTS

SUMMARY

Data traffic will play a major role in future wireless communication systems. In this thesis, we focus on the power management issues for wireless data traffic in a code-division multiple-access mobile system and the information dissemination issues among mobile devices in mobile networks.

We first propose a power control algorithm for a direct sequence code-division multiple access radio mobile system under multipath environments. The motivation for this study is two fold: limiting multiple access interference to increase system throughput and reducing power consumption to prolong battery life. We utilize the delay insensitive property of the data traffic to automatically control the transmission based on the channel quality. The proposed algorithm controls the transmit power based on signal-to-interference ratio estimation to achieve an optimal trade-off between the transmit power and delay. Simulation results show significant improvement in terms of throughput and power consumption, in comparison with other power control algorithms.

We then propose a new class of scheduling strategies for data dissemination in mobile networks. We show by analysis and simulation that the network performance depends on both the number of nodes in the network and the number of packets to be disseminated. Our results show significant capacity improvement. Our scheduling strategies also achieve higher packet dissemination success probability. Furthermore, we point out that such distributed scheduling is very simple and induces no additional communication overhead to the mobile networks.

LIST OF FIGURES

LIST OF ABBREVIATION

ABR	Available Bit Rate
BER	Bit Error Rate
BLER	Block Error Rate
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CRC	Cyclic Redundancy Check
DPSK	Differential Phase Shift Keying
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FIFO	First In First Out
GSM	Global system for Mobile Communication
LAN	Local Area Network
MC	Mobile Client
MSS	Mobile Support Station
PBSPC	Power & Backlog Sensitive Power Control
PCMA	Power Controlled Multiple Access
PDA	Personal Digital Assistant
PER	Packet Error Rate
PSC	Probability of Seamless Communication
QoS	Quality of Service
RNC	Radio Network Controller
SIR	Signal-to-Interference Ratio

TDMA	Time Division Multiple Access
VBR	Variable Bit Rate
WAN	Wide Area Network
WCDMA	Wideband CDMA, Code Division Multiple Access

CHAPTER ONE

Introduction

In this chapter, the motivation of our research is first provided. The main contributions we have made in this thesis are then briefed. Finally, we conclude this chapter by describing the organization of this thesis.

1.1 Background

In recent years, a variety of mobile devices such as cellular phones, notebook computers, palmtops, Personal Digital Assistant (PDAs) are becoming smaller and more pervasive and are not only being carried by humans, but also integrated into physical objects (such as cars, electrical appliances). There is a growing demand for access to information, especially local and general news, traffic or weather reports, sports, maps, guide books, music and video packets, by wireless users roaming through metropolitan areas. Many of today's "fixed" applications will continue to make sense in a mobile environment, and a variety of new services are certain to evolve for people on the move. Therefore, wireless networking is rapidly becoming a major component of the modern communication infrastructure.

A fundamental characteristic of mobile wireless networks is the time variation of the channel strength of the underlying communication links. Such time variation occurs at multiple time scales and can be due to multipath fading, path loss, shadowing by obstacles, and interference from other users. The impact of such time variation on the design of wireless networks permeates throughout the layers, ranging from coding and power control at the physical layer to cellular handoff and coverage planning at the networking layer. An important means to cope with the time variation of the channel is the use of *diversity*. Diversity can be obtained over time (interleaving of coded bits), frequency (combining of multipaths in code-division multiple access – CDMA systems), and space (multiple antennas or multiple base stations). The basic idea is to improve performance by creating several independent signal paths between the transmitter and the receiver.

These diversity modes pertain to a point-to-point link. Recent results point to another form of diversity, inherent in a wireless network with multiple users. This *multiuser diversity* is best motivated by an information theoretic result of Knopp and Humblet [1]. They focused on the uplink in single cell, with multiple users communicating to the base station via time-varying fading channels. To maximize the total information theoretic capacity, they showed that the optimal strategy is to schedule at any one time only the user with the best channel to transmit to the base station. Diversity gain arises from the fact that, in a system with many users whose channels vary independently, there is likely to be a user with a very good channel at any one time. Overall system throughput is maximized by allocating at any time the common channel resource to the user that can best exploit it. Similar results can be obtained for the downlink from the base station to the mobile users [2].

This type of strategies incurs additional delay, because packets have to be buffered until the channel becomes strong relative to other users. Therefore, the time scale of channel fluctuations that can be exploited through multiuser diversity is limited by the delay tolerance of the user or application. For example, for voice oriented applications that can tolerate little delays and delay variations, multiuser diversity can not be exploited to provide Qualify of Service (QoS) assurance. On the other hand, for data applications which can tolerate delays on the order of fractions of seconds to several seconds, short time-scale fading due to interference of multiple signal paths can be taken advantage of. In this thesis, we focus on the power control problem for wireless data applications which can tolerate large delays. The idea is to have sufficient incentive for the users to automatically detect channel conditions favorable to a time sharing structure and implement them without using a centralized control or synchronization method. The power control problem is reformulated as a cost minimization problem where the desire for increased transmitted power is weighed against the associated cost.

For applications that are so asynchronous in nature that they can tolerate end-to-end delays of minutes or even hours, even more diversity gain can be obtained because the *network topology* changes significantly over time due to user mobility. Examples of such applications include electronic mail, database synchronization between a mobile terminal and a central database, and certain types of event notification. This leads to the design of wireless networks beyond classical cellular architectures. New mobile networking architectures have been proposed that exploit node mobility to achieve large network capacity. In these networking architectures, nodes are connected intermittently when they are in proximity and networking is brought about by node mobility. In this thesis, we study a specific type of application: the data sharing and dissemination in mobile (not necessarily ad hoc) environments. A class of scheduling strategy is proposed to speed up the data dissemination rate in a system which exploits the node mobility to achieve more diversity gain.

1.2 Our Contributions

The contributions, which are elaborated throughout the thesis, can be briefly listed as follows:

1. Studied the power control schemes for voice oriented traffic in current cellular networks. Explored the possibility and methods to improve the system performance by using power control to achieve a time sharing structure for data oriented traffic in wireless networks.

2. Proposed the Power & Backlog Sensitive Power Control (PBSPC) algorithm for wireless data under a multipath environment. The PBSPC determines the next transmission power based on the estimation of the current channel quality, the queue length of the transmitter and the last transmission power.

3. Evaluated the performance of PBSPC in terms of system throughput, average transmission power and average transmission success probability, and compared it with the existing Fixed Signal-to-Noise Ratio (Fixed SIR) algorithm and Power Controlled Multiple Access algorithm (PCMA) to observe how much improvement the algorithm has achieved over the Fixed SIR power control and PCMA.

4. Studied new mobile networking architectures which exploit node mobility to achieve better system performance. A specific type of application, the data sharing and dissemination in mobile environments is studied in details.

5. Proposed a class of scheduling strategies for disseminating data among mobile devices. Popular data files are divided into packets and cached in the access points. The scheduling strategies schedule the data packets downloading at the access

points. Mobile stations exchange their packets when they encounter each other.

6. Performed simulations to evaluate the performance of the two scheduling strategies and compared them with the random strategy to observe how much improvement they have obtained over the random strategy.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows.

Chapter 2 begins with an introduction to the basics about mobile computing and then introduces the concepts and methods of power control. The data dissemination issue is covered in the final part of this chapter.

In Chapter 3, we first briefed the power control schemes used for current cellular mobile systems. A famous power control algorithm used for voice-oriented traffic is studied and the motivation of designing power control algorithm for data traffic is examined by analysis and a numerical example. We then develop a power & backlog sensitive power control algorithm for data traffic in a CDMA wireless communication system. Finally, the performance of downlink channels using the proposed algorithm is evaluated via computer simulations.

In Chapter 4, the concept of data dissemination in wireless system is first introduced. The performance matrix and design issues are discussed. The related work is briefly reviewed. We then considered a finite linear network with mobility limited mobile nodes and propose a new class of scheduling strategies for the data dissemination in mobile networks. Finally the algorithm is evaluated via simulations.

Lastly, Chapter 5 concludes the thesis and some recommendations for future work are made.

CHAPTER TWO

Basic Issues about Mobile Computing

In this chapter, we first give some background of mobile computing. Then the reference architecture of wireless computing environment and the characteristics of wireless computing environments are discussed. The QoS concept is introduced at last.

2.1 Introduction

We are living in a mobile world where people travel more and more frequently from place to place. At the same time, information has become such an important asset that many people are craving after to meet their information needs as well as to confer upon themselves a competitive advantage. Wireless computing is the phenomenon that enables one to receive and disseminate information even when one moves about – it is mobile computing with untethered communications.

The development of wireless computing systems is made possible by the emergence of lightweight, battery-operated, low-cost and portable devices such as palmtops and PDAs, and the availability and exploitation of wireless networks. Through the wireless networks, portable devices will become an integrated part of existing distributed computing environments, and mobile users can have access to data stored at information servers located at the static portion of the network even when they are on the move.

2.2 Reference Architecture of Wireless Computing Environments

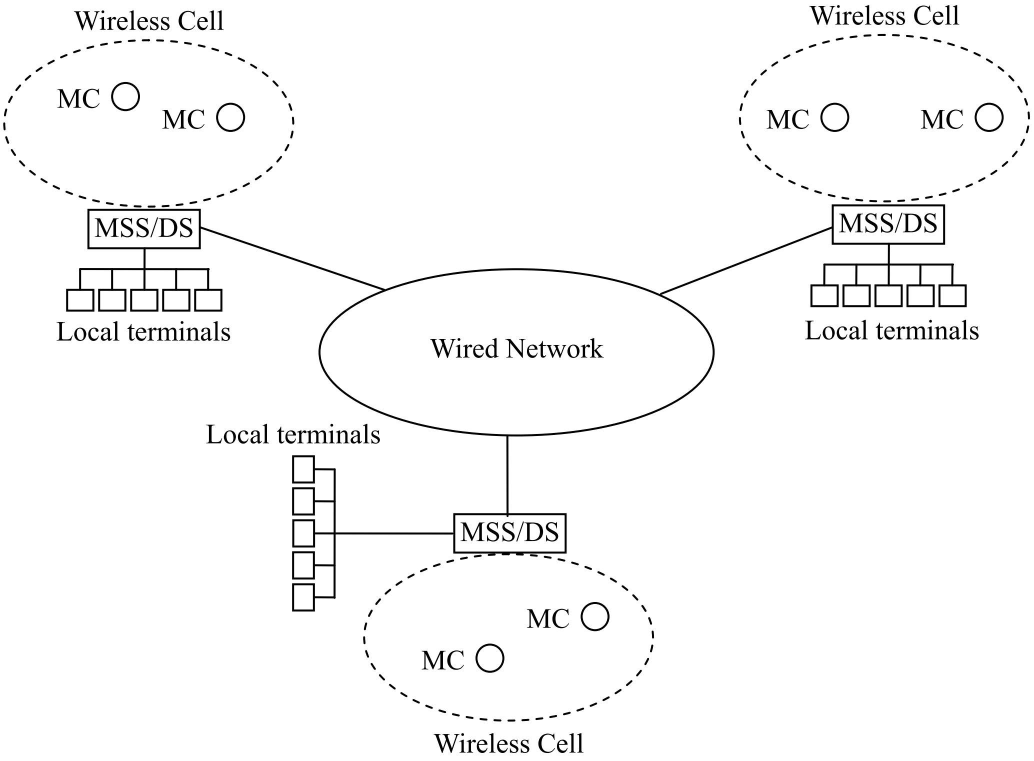


Figure 2-1 Reference architecture of a wireless computing environment.

Figure 2-1 shows a typical architecture of a wireless computing environment [3]. The wireless environment consists of two distinct sets of entities: a larger number of mobile devices/clients (MC) and relatively fewer, but more powerful fixed hosts or database servers (DS). The fixed hosts are connected through a wired network, and may also be serving local terminals. Some of the fixed hosts, called mobile support stations (MSS), are equipped with wireless communication capability. The MSSs and MCs communicate through a wireless communication channel called a wireless cell defined over the radio coverage area of the MSSs. The diameter of a cell, as well as the available bandwidth, may vary according to the specific wireless technology which can either be a cellular connection or a wireless local network. For example, the diameter of a cell may span a few meters for infrared technology to 1 or 2 miles for radio or satellite networks. On the other hand, the bandwidth of local area networks (LANs) using infrared technology can be of the order of 10Mbps, whereas wide area networks (WANs) have much lower transfer rates. We note that the fixed network is static, whereas the wireless network is mobile since MCs may move from cell to cell.

MCs are expected to be portable devices that are low-cost, lightweight and powered by batteries. MCs can be dumb terminals or walkstations [3]. The former are diskless hosts (for instance, palmtops) with reduced memory and computing capabilities and has no transmit capability. These MCs will only listen to the channel and select the information requested by users from the data broadcast by the servers. Walkstations, on the contrary, are comparable to the classic workstations, and can both receive and send messages on the wireless network. Most of the portable walkstations can also disconnect from the server by operation in a doze mode or a power-off mode in order to conserve energy.

MSSs can provide both information and commonly used application software for mobile users. They manage and service on-demand requests from mobile clients within the coverage of their cells. Based on the requests, the data/software are retrieved and sent via the wireless channel to the mobile clients. Mobile users can either process the data on their portable devices or tap on the more powerful MSS to perform tasks on their behalf remotely. Each MC is associated with a specific MSS, called home MSS. A home MSS for a MC maintains specific information about the MC itself, such as the user profile, logic files, access rights, and user private files. Additionally, a user may register as a visitor under some other MSSs. Thus a MSS may also need to keep track of the addresses of users who are currently residing in the cell supervised by the MSS itself.

The wireless channel is logically separated into two subchannels: an uplink channel which is used by clients to submit queries to the server via MSS, and a downlink channel which is used by MSS to pass the data from MSS to the intended clients. In general, MSS periodically broadcasts frequently demanded information to MCs within its cell who will listen and filter the necessary information. For information that is less frequently accessed, MCs will have to explicitly request for these information.

2.3 Characteristics of Wireless Computing Environments

There are several characteristics that are typical of wireless computing environments:

1. *Limited Wireless Channel Bandwidth*: The bandwidth of wireless channel is very limited. It can vary from 10Kbps for wide coverage area (e.g. Mobitex 8 Kbps link or Global system for Mobile Communication - GSM 9.6 Kbps link), through 19.2 Kbps (e.g. Cellular Digital Packet Data) to about 10 Mbps for wireless LAN (e.g. infrared link).

2. *Susceptibility of Wireless Communications to Interference*: Wireless transmission is error-prone. Many factors contribute to this. Signal may be lost due to external interference such as electromagnetic noises or bad weather. Wireless communications are also terrain-sensitive – signal may be affected by tall buildings and trees, hills and valleys. Signals may also be weakened due to the clients' distance from the server or the high speed at which the clients are moving. Data may also be delayed due to congested network traffic.

3. *Asymmetric Communication Environments*: Wireless computing is characterized by its inherent communications asymmetry. First, the data volume from servers to

client (downstream direction) is much larger than that from clients to server (upstream direction). Second, asymmetry arises as a result of the difference in network bandwidth in the downstream and upstream channel. Third, there is usually only a small number of servers that provide information to a large number of clients. These servers can easily be swamped by an unexpected larger number of requests.

4. *Limited Effective Battery Lifespan*: Users of wireless computing systems are expected to be carrying portable devices. Batteries are the lifeblood of these devices, and the battery life is affected by virtually all facets of system operation. The complexity and the amount of data processed directly affect the energy consumption. The amount and distance of data transmitted also have a direct effect on battery life. Unfortunately, as most current battery research does not predict a substantial change in the available energy in a consumer battery, it is important that wireless devices be designed to be energy-conserving.

5. *Threat to Security*: Since signals travel through the open atmosphere, they can be readily intercepted by anyone. Thus, wireless networks pose more security requirements than its wired counterparts.

6. *High Cost*: It is generally more expensive to send data over wireless networks. In some networks (e.g. in cellular phones), network access is charged based on connection time while in others (such as those in radio packets) the rate is based on the number of messages (packets) transmitted.

2.4 Quality of Service (QoS)

During the past few years, the importance of quality of service (QoS) has grown considerably in the telecommunication industry. QoS is the ability of a network (wireless or wireline) element to have some level of assurance that its traffic and service requirements can be satisfied. The typical QoS metrics are end-to-end delay, packet loss, BER (Bit Error Rate), throughput, delay variation etc.

Different applications need different QoS assurance. For instance, data applications require throughput assurance; on the other hand, real-time audio applications need delay variation assurance. To support different services, the wireless network should provide service differentiation, in other words, to provide different levels of assurance (QoS) to different applications.

Traditional QoS constraints are usually guaranteed by circuit-switching concepts, while future wireless networks are typically realized by packet-switching technologies. Meanwhile, the main difficulty in packet switching is how to guarantee QoS in wireless access due to the nature of wireless transmission and networking.

First, typical wireless links suffer from severe fading, shortage of bandwidth, propagation loss, interference, and distortion [5], [6]. This unreliability may manifest as high BER. Forward error correction (FEC) and power control techniques may be used to guard against high BER, but there are tradeoffs to be made [7], [8]. FEC improves the reliability of the link, at the cost of reduced effective throughput. Similarly, power control is used to combat interference. Thus, it is inevitable that some data are delayed unpredictably due to retransmissions, or lost.

Second, wireless networks allow stations/nodes in the networks to move around. It induces a special feature in wireless networks known as mobility management. In other words, network topology is dynamically changing as stations move in/out cell coverage or turn on/off the transmission. Any flow to the old access point must be redirected to the new, and until this is accomplished, data in transit must be stored at the old access point and forwarded to the new, or dropped [9]. There is no guarantee that the same level of resources will be available under a new access point to which the mobile station moves. This implies that there many be service disruption with the mobility of the mobile stations.

In [10], wireless access to broadband networks with end-to-end QoS constraints is able to serve three kinds of mobile traffic sources: available-bit rate (ABR), constant-bit-rate (CBR), and variable-bit-rate (VBR) sources in a wireless multimedia-networking environment. And the QoS requirements depend on the traffic types: ABR traffic, which has the characteristics of conventional data traffic, requires stringent packet error rate (PER) and is insensitive to delay and jitter. CBR (e.g. voice) and VBR (e.g. video) traffic requires moderate packet error rate and is usually very sensitive to delay and jitter.

Now, some QoS measures have been proposed for wireless environments. For example, there are two QoS parameters that a wireless host may specify: the loss profile and the probability of seamless communication (PSC) [11]. Assuming that data loss is inevitable, the loss profile specifies which packets in a stream may be discarded. Presumably, certain patterns of data loss are acceptable to the application whereas others are not. The PSC specifies the probability that there would not be a break in service (i.e. data loss) during hand-off. The PSC parameter is used by the network to decide whether data should be multicast to neighboring access points, and if so to which ones. The idea is that the availability of data before hand-off ensures prompt delivery. Clearly, seamless communications imply that wireless resources must also be reserved ahead of time in adjacent cells, Similar mechanisms are also used in the proposals described in [12] – [14].

CHAPTER THREE

Power Control for Wireless Data

In this chapter, we first briefed the power control schemes used in current cellular mobile systems. A famous power control algorithm used for voice-oriented traffic is studied and the motivation of designing power control algorithm for data traffic is examined by analysis and a numerical example. We then develop a power & backlog sensitive power control algorithm for data traffic in a CDMA wireless communication system. Finally, the performance of downlink channels using this algorithm is evaluated by computer simulations.

3.1 Power Control

Power control is an essential technique implemented not only in most cellular networks but also in future 3G systems to overcome channel gain fluctuations, prolong the battery duration of handsets and improve the system capacity. Power control is a method which adjusts the transmit power level so that the received signal-to-interference ratio (SIR) adapts to the interference and channel conditions. It has meanwhile been designed in the CDMA cellular system to resist the near-far-effect and reduce the co-channel interference since all the users in the same cell occupy the same frequency band.

Power control is of fundamental importance to the operation of wireless networks for a number of reasons. First of all, by adjusting transmitter power to maintain an acceptable receiving signal quality at each mobile using the least possible power, it is feasible and favorable to minimize power consumption and prolong the battery life of mobile nodes. Moreover, efficient power control schemes can significantly mitigate the near-in and far-out interferences, improve the spatial reuse of channel resources and increase the network capacity, which is particularly beneficial to cellular CDMA systems because by controlling the power of each user accessing a cell, resources can be shared equitably among all users and the capacity can be maximized [15] – [17]. Power control can also be used to implement various basic dynamic network operations online, such as admission control, link QoS maintenance, channel selection and switching, resource allocation, and handoff control, etc [18].

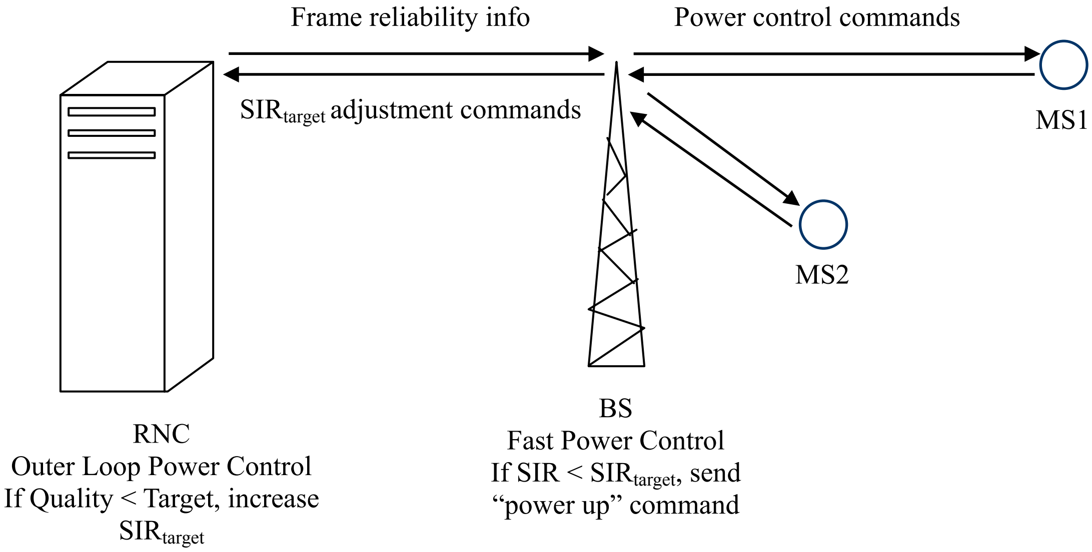


Figure 3-1 Power control loops.

In cellular networks, power control is needed for both forward/downlink channels (base station to mobile stations) and reverse/uplink channels (mobile stations to base station). As there is a pilot channel on the forward link, power control is considerably less complicated to implement on forward links than on reverse links. The base station just generates the power control bit and inserts it on every forward channel to instruct individual mobile stations to adjust their transmission power level.

Power control is more crucial in CDMA, in particular on the uplink. Without it, a single high powered mobile could block a whole cell. As seen in Figure 3-1 Mobile stations MS1 and MS2 operate within the same frequency, separable at the base station only by their respective spreading codes. It may happen that MS1 at the cell edge suffers a path loss, say 70 dB above that of MS2 which is near the base station BS. If there were no mechanism for MS1 and MS2 to be power-controlled to the same level at the base station, MS2 could easily block MS1, giving rise to the so-called near-far problem of CDMA. The optimum strategy in the sense of maximizing capacity is to equalize the received power per bit of all mobile stations at all times.

Power control schemes in mobile cellular network can be either centralized or distributed. A centralized mechanism is capable of achieving optimal performance, although it is more costly and unscalable for large networks. A distributed mechanism, on the contrary, does not provide the best solution, but it is less expensive to operate.

According to the information measured to determine whether to raise or lower the transmission power of a mobile station, power control can be divided into two categories: strength-based and SIR-based. In a strength-based scheme, the strength of a signal arriving at the base station from a mobile station is measured to determine proper power control action, whereas in a SIR-based design, the quantity measured is SIR, with the interference consisting of both channel noise and multiuser interference. Obviously, the SIR-based scheme is more accurate to indicate signal quality and therefore provides better performance than that of the strength-based scheme.

Power control in cellular networks can employ either an open or a closed loop. In closed-loop power control, the base station uses the quality of the signal received from a mobile station to issue a command to tell the mobile station how to adjust its transmitted power. In open-loop power control, each mobile station adjusts its transmitted power assuming that path loss in both directions, forward and reverse, are symmetric. Open-loop power control can only compensate for the loss due to long-term fading, and it can not make up for the loss due to short-term fading. The two types of power control are shown in Figure 3-1.

3.1.1 Open-Loop Power Control

The open-loop power control estimates the path loss to the channel by measuring the received signal level, assuming that path loss in both directions, forward and reverse, are symmetric, and adjusts the transmitter power accordingly to meet the desired target. The commanded adjustments are transmitted to the mobile station in the overhead messages. They are used to adjust the open-loop power control for different sized cells and receiver sensitivities.

One can see that open-loop power control mechanisms attempt to make a rough estimate of path loss by means of a downlink beacon signal. Such a method would be too inaccurate without considering fast fading since the fast fading is essentially uncorrelated between uplink and downlink, especially in a frequency division duplex (FDD) system where there is a large frequency separation between uplink and downlink frequency bands. Open-loop power control is used only to provide a coarse initial power setting of the mobile station at the beginning of a connection. It is indeed true that the open loop estimate is often in error by several dB without closed-loop power control.

3.1.2 Closed-loop Power Control

In closed-loop power control in the uplink, the base station performs frequent estimation of the received SIR and compares it to a target SIR. If the measured SIR is higher than the target SIR, the base station will command the mobile station to lower the power; if it is too low it will command the mobile station to increase its power. This measure-command-react cycle is executed at a rate of 1500 times per second (1.5 kHz) in Wideband Code-Division Multiple Access (WCDMA) or each mobile station and thus operates faster than any significant change in path loss that could possibly happen and, indeed, even faster than the speed of fast Rayleigh fading for low to moderate mobile speeds. Thus closed-loop power control will prevent any power imbalance among all the uplink signals received at the base station.

The same closed-loop power control technique is also used on downlink, though here the motivation is different: on the downlink there is no near-far problem due to the one-to-many scenario. All the signals within one cell originate from the one base station to all mobiles. It is, however, desirable to provide a marginal amount of additional power to mobile stations at the cell edge, as they suffer from increased other-cell interference. Also on the downlink, a method of enhancing weak signals caused by Rayleigh fading with additional power is needed at low speeds when other error-correction methods based on interleaving and error correction codes do not yet work effectively.

3.1.3 Outer-Loop Power Control

There is one more related control loop connected with closed-loop power control: outer-loop power control. As shown in Figure 3-1, outer-loop power control adjusts the target SIR value in the base station according to the needs of the individual radio link and aims at a constant quality, usually defined as a certain target BER or block error rate (BLER). The reason for the need for outer-loop power control is because the required SIR, say, BLER = 1% depends on the mobile speed and the multipath profile. Now, if one were to set the target SIR for the worst case, i.e. high mobile speeds, one would waste much capacity for those connections at low speeds. Thus, the best strategy is to let the target SIR float around the minimum value that just fulfils the required target quality.

Outer-Loop power control is typically implemented by having the base station tag each uplink user data frame with a frame reliability indicator, such as a Cyclic Redundancy Check (CRC) result obtained during decoding of that particular user data frame. Should the frame quality indicator indicate to the Radio Network controller (RNC) that the transmission quality is decreasing, the RNC in turn will command the base station to increase the target SIR set point by a certain amount. The reason for having outer loop power control residing in the RNC is that this function should be performed after a possible soft handover combining.

3.2 Motivation on Power Control for Wireless Data

Power control research has attracted substantial attention in the past few years, focusing mostly on voice-oriented "continuous" traffic, which is dominant in current generation wireless networks. Next generation wireless data networks will support packet switched traffic for delivering sophisticated information services over wireless channels (e.g. wireless web browsing and mobile computing). The power control problem in packet switched wireless environments is not well understood. Preliminary studies [19] have only partially addressed it.

In this section, we briefly explain the conventional, Fixed SIR power control method [20], and show by a numerical example how an alternative method can lead to substantial reduction in overall energy consumption.

3.2.1 Fixed SIR Power Control

Consider a network with N mobile transmitters, each communicating with another node or nodes over wireless channels, with the channels separated by codes, space, and/or frequency. For i and j in $\{1, \dots ,N\}$ let $g_{i,j}$ represent the interfering-link gain from transmitter j to i's receiver when $j \neq i$, and the intended-link gain (including path loss and signal processing gain) between transmitter i and i's receiver when $j = i$, so a signal transmitted by j with power p_j is received by i's receiver with power $g_{i,j}p_j$. Define the $N \times N$ gain matrix $\mathbf{G} = [g_{i,j}]$. In addition, denote by η_i the noise power at i's receiver, and Φ_i the set of admissible power levels for transmitter i.

Given such a network and some power control objective, we call the objective achievable if and only if it is satisfied for a set of power vectors $\{p_i \in \Phi_i \colon i = 1, \dots , N\}$. For instance, the objective may be to attain a set of N specified SIRs γ_i, one for each transmitter/receiver pair. If each Φ_i is the set of nonnegative real-valued constant values, we have a Fixed SIR power controlled system. Define $\mathbf{p}$ as the transmitter power vector, γ as the target SIR vector, then in vector/matrix notation, we desire a solution to the equation

$$\mathbf{p} = \hat{\mathbf{Z}}\mathbf{p} + \hat{\boldsymbol{\eta}}, \tag{3.1}$$

where

$$\hat{\mathbf{Z}} = \hat{\mathbf{G}} - diag(\boldsymbol{\gamma}), \quad \hat{G}_{ij} = \frac{\gamma_i g_{ij}}{g_{ii}}, \quad \hat{\eta}_i = \frac{\gamma_i \eta_i}{g_{ii}}. \tag{3.2}$$

The objective SIR vector $(\gamma_1, \dots, \gamma_N)$ is achievable if and only if the determinant

$|\mathbf{I}-\hat{\mathbf{Z}}|$ is positive, in which case the solution is given by

$$\mathbf{p}=(\mathbf{I}-\hat{\mathbf{Z}})^{-1}\hat{\boldsymbol{\eta}}. \tag{3.3}$$

For instance, the solution exists for $N = 2$ if

$$\gamma_1\gamma_2\frac{g_{12}g_{21}}{g_{11}g_{22}}<1. \tag{3.4}$$

When we can confirm that a network objective is achievable, our attention turns to whether there exists an algorithm to realize the desired state starting from a given initial condition. In particular, we are interested in algorithms for selecting transmitter powers p_i^{k+1} at times $k = 0, 1, \ldots$, and starting from p_i^0. Let us say that a network, together with an algorithm (or collection of algorithms), is power-stable if the algorithm generates a sequence of admissible power vectors $\{p^k: p_i^k \in \Phi_i, k = 1, 2, \ldots\}$ such that for all i,

$$p_i^k \to p_i \in \Phi_i, \quad k\to\infty, \tag{3.5}$$

and the objective is satisfied for the limiting powers $\{p_i\}$. For example, let us write γ_i^T for the target SIR on channel i and γ_i^k for the SIR at time k on channel i. Then Foschini and Miljanic [20] have shown that whenever a fixed SIR power control system with constant noise power vector η has an achievable vector γ, and the spectral radius of $\hat{\mathbf{Z}}$ is less than one, the algorithm

$$p_i^{k+1}=\frac{\gamma_i^T}{\gamma_i^k}p_i^k \tag{3.6}$$

is power-stable. That is, the power vector converges to a value which yields the target

SIR vector.

3.2.2 Solution for 2-Nodes

From now on let us focus on the case $N = 2$. If the i^{th} channel's receiver noise is η_i^k at time k, then the SIR on channel i is

$$\gamma_i^k = \frac{g_{ii} p_i^k}{g_{ij} p_j^k + \eta_i^k}. \tag{3.7}$$

Given the error rate as a function of SIR, we can calculate the successful data transmission rate as a function of these parameters. Typical BERs for a number of common modulation schemes and environments are well known, such as binary differential phase shift keying (DPSK) in a non-fading environment (BER = $a \cdot \exp\{-b\gamma\}$ for SIR γ and, e.g. $a = b = 1$) or binary noncoherent frequency shift keying in the presence of fading (BER = $1/(\gamma + 2)$) [34]. The function

$$e(\gamma) = \frac{1}{1+\gamma} \tag{3.8}$$

represents a reasonable upper bound on the BER relative to BERs obtained with many conventional modulation schemes. Assume for now that equation (3.8) holds. In order to transmit at respective rates or throughputs r_1 and r_2 ($0 \le r_i < 1$) the transmitter powers would have to be sufficient to attain at the receiver SIRs of $\gamma_1 = r_1/(1-r_1)$ and $\gamma_2 = r_2/(1-r_2)$, respectively. Solving for p_1 and p_2 simultaneously in terms of rates and gains we obtain

$$p_1 = \frac{r_1(g_{12} r_2 \eta_2 + g_{22}(1-r_2)\eta_1)}{(g_{11}g_{22} - g_{12}g_{21}) r_1 r_2 + g_{11}g_{22}(1 - r_1 - r_2)} \tag{3.9}$$

and

$$p_2 = \frac{r_2(g_{21}r_1\eta_2 + g_{22}(1-r_2)\eta_1)}{(g_{11}g_{22} - g_{12}g_{21})r_1r_2 + g_{11}g_{22}(1-r_1-r_2)}, \tag{3.10}$$

as long as the rates r_i are achievable, i.e.

$$\frac{r_1}{1-r_1}\frac{r_2}{1-r_2}\frac{g_{12}g_{21}}{g_{11}g_{22}} < 1. \tag{3.11}$$

Using this Fixed SIR transmission scheme, the total energy used to obtain rates r_1 and r_2 over a unit time interval is therefore

$$E = \frac{r_1\eta_1(g_{21}r_2 + g_{22}(1-r_2)) + r_2\eta_2(g_{12}r_1 + g_{11}(1-r_1))}{(g_{11}g_{22} - g_{12}g_{21})r_1r_2 + g_{11}g_{22}(1-r_1-r_2)}. \tag{3.12}$$

This is the best result can be achieved under Fixed SIR power control scheme because the solution to (3.1) is unique for non-zero receiver noise powers. However, given certain values of the gain matrix G and rates r_i, it is possible to save energy by switching to a time division-like access protocol while keeping the same rates.

3.2.3 Numerical Example

Consider the following typical parameters in a cellular code-division multiple access (CDMA) system: $r_1 = r_2 = 1/3$, $g_{11} = g_{21} = 1$, $g_{22} = g_{12} = 1/16$, $\eta_1 = \eta_2 = 1$. The values obtained using (3.9) and (3.10) are

$$p_1 = 1 \text{ and } p_2 = 16. \tag{3.13}$$

If the time period of interest is the unit interval, then the transmitters will use 1 and 16 units of energy, respectively, for a total of $E = 17$.

On the other hand, if the two transmitters are permitted to transmit using distinct time segments – only transmitter one during $(0, \tau]$, only transmitter two during $(\tau, 1]$, then it would be possible to lower the total energy used. For instance, if $\tau = 2/5$ then

$$p_1(t) = \begin{cases} 5, & 0 < t \le 2/5, \\ 0, & 2/5 < t \le 1, \end{cases} \tag{3.14}$$

and

$$p_2(t) = \begin{cases} 0, & 0 < t \le 2/5, \\ 20, & 2/5 < t \le 1, \end{cases} \tag{3.15}$$

would also yield rates $r_1 = r_2 = 1/3$, but would consume only $E = 14$ units of energy, versus 17 under Fixed SIR power control scheme.

In the preceding example, there is a nearly twenty percent reduction in the energy usage under the time division scheme. Moreover, the lower total transmitter power means lower interference to other users, and hence leads to potential benefits like improved network throughput and stability.

The drawbacks are, first, one of the transmitters is called on to increase its energy consumption from 1 to 2 in switching to the more globally efficient time division scheme; this is likely to be unsatisfactory if the user is not eventually compensated. Second, the asynchronous static scheme is replaced by a synchronized system wherein each transmitter must know when to turn on and off with increased system complexity.

While the first drawback is inherent to the protocol, the second can be at least partially avoided. In fact, it may be possible to achieve much of the benefit usually associated with Time Division Multiple Access (TDMA) without using central control or synchronization to force users into time slots. The idea is to have sufficient incentive for the users to automatically detect conditions favorable to a time division structure and implement them without the intervention of a base station or other mediator. This will be addressed in more detail in Section 3.4.

3.2.4 The Power vs. Delay Tradeoff

Although achieving satisfactory quality of service (QoS) is important for users, they may not be willing to achieve it at arbitrarily high power level, because power is itself a valuable commodity, especially for mobile users. On the other hand, one user using high power level to transmit will cause high interference to others which drives them to increase their power as well, and consequently increase the total interference of the system. This observation motivates a reformulation of the whole problem where the desire for increased transmitted power is weighed against the associated cost.

When observing high interference in the channel, a transmitter realizes that it will have to use high power to overcome the interference and transmit a packet successfully to the receiver. Therefore, it might consider backing off, buffering the incoming traffic, and waiting for the interference to decrease before it transmits again, trading delay cost for power benefit. When it has backed off, its buffer starts filling up with new arriving packets, pushing the backlog/delay cost higher and putting pressure on the transmitter to transmit aggressively at high power in order to reduce its backlog fast, and tradeoff power cost against delay benefit. With many transmitters sharing the channel, the issue becomes more complicated.

3.3 Power Control Algorithm

3.3.1 Introduction

Along with the proliferation of cellular networks in 1990s, as an important means for providing good transmission quality and expanding wireless network capacity, power control design has been extensively studied around the world in the past few years, especially for cellular mobile code division multiple access (CDMA) systems [20] – [25]. Those studies have mainly been used to reduce co-channel interference and to guarantee the SIR of the connections, resulting in a higher utilization and/or better quality of service (QoS). Most of the research, however, has focused on voice-oriented “continuous” traffic, which is dominant in current generation wireless networks. One of the most well known algorithms was originally proposed by Foschini & Miljanic [20] for voice-oriented continuous traffic in wireless networks, and later has been further studied extensively to cover several other aspects and shown to have several optimality properties for continuous traffic by Mitra [22], and Bambos, Chen, and Pottie [21], [23]. The algorithm aims to keep a constant SIR during the transmission and a corresponding constant transmission success probability, when it has a packet to transmit. Such schemes assume that data can not wait in the buffer and have to be transmitted immediately. Thus, current power control algorithms are not optimal for future data traffic.

Unlike delay sensitive voice traffic, which allows only minimal buffering, data traffic is more tolerant to delay than the voice traffic. Taking advantage of this property, a natural scheme is to transmit packets when there is low noise or interference, and not transmit packets or transmit at very low power when the interference is high. The idea has been studied and an algorithm called PCMA-1 has been introduced by Nicholas Bambos & Sunil Kandukuri [26]. The authors first considered the problem of transmission in an extraneous, statistically stationary interference environment, and found optimal policies (power levels as a function of interference) subject to various restrictions. This algorithm is trying to achieve a variable SIR target which is different from those constant SIR algorithms.

However, the above method only takes into account the effect of the interference from the other transmitters. It usually fails to compensate for the channel variation due to path loss via distance attenuation and multipath fading. The successful receiving of a packet not only depends on the interference at the receiver, but also depends on the received signal power which is affected by the transmission path loss. Therefore, a new power control method was proposed to overcome this disadvantage. As we will see later, the power control method improves the system performance and outperforms the power control algorithm in [26].

3.3.2 System Model

We will consider a power-controlled CDMA cellular system where the transmission power is continuously tunable. The system consists of N cells and M active mobile stations. Each cell is assumed to be identical, with base stations located at the center of the cells and mobile stations randomly distributed in the cells. Each mobile station is associated with the base station in the cell. The transmission quality is assumed to be dependent only on the received signal-to-interference ratio (SIR). We only consider the downlink transmission here, because the uplink case can be treated similarly [27], [28].

Each communication link is assumed in a Rayleigh fading channel and the characteristic of each link is assumed to be independent and identical. The transmission signals experience attenuation by three independent parts; path loss, shadowing (long-term fading), and multipath and Doppler shift (short-term fading). The total channel effect denoted by L can be expressed as

$$L_{(dB)} = L_{path} + L_{shadowing} + L_{fading} . \tag{3.16}$$

The path loss, denoted by L_{path}, is proportional to d^{-n}, d being the distance between the transmitter and the receiver, and n is the path loss exponent which ranges from 3 to 5 for typical shadowed urban cellular. We use $n = 4$ in this study.

The shadowing, denoted by $L_{shadowing}$, is observed to be log-normal distribution [29]. It can be expressed as

$$L_{shadowing} = 10^{X/10} \tag{3.17}$$

where X is a Gaussian distributed random variable with mean μ and standard deviation $\sigma_{s,}$ which is affected by the configuration of terrain and rages from 5 to 12, with 8 as a typical value [29].

The short-term fading, denoted by L_{fading}, is observed to be Rayleigh distributed (on strength) when no line-of-sight path exists. The probability density function (PDF) for amplitude r of the received signal envelope is given by

$$R(r) = \begin{cases} \dfrac{r}{\sigma^2} \exp\left(-\dfrac{r^2}{2\sigma^2} \right) & 0 \le r \le \infty \\ 0 & r < 0 \end{cases} \tag{3.18}$$

where σ^2 is the time-average power of the received signal before envelope detection.

In our model, we assume that there are P paths for each link, but no line-of-sight path exists. The P paths arrive at the receiver in an angle uniformly distributed in [0, 2π]. Thus, the signal received at the mobile receiver becomes [30]

$$S_r = \sum_{i=1}^{P} A_i e^{j2\pi(f_0 - \frac{V}{\nu}\cos\theta_i)t} \tag{3.19}$$

where A_i is a complex random variable with zero mean and unit variance, f_0 is the central frequency, ν is the wavelength, V is the speed of the mobile station, and θ_i is the arrival angle of the i^{th} path and is uniformly distributed on [0,2π].

Each communication link is assumed to be operating in slotted time. The path loss is assumed to be staying constant during each time slot, but fluctuate randomly in different time slots. The transmitter power is assigned at the beginning of each time slot and kept stable during that time slot. The interference of the link is induced by all other transmitters sharing the channel and the background noise is Gaussian distributed.

Let us denote the link gain on the path from the base station in cell q to the h^{th} mobile station in cell j as g_{qkj}. Through some one to one corresponding relationship, we can map (h, j) to i, where $1 \le i \le M$. For simplicity, we let

$$i = \sum_{n=0}^{j-1} m_n + h, \quad 1 \le h \le m_j \tag{3.20}$$

where $m_0 = 0$, m_n ($n = 1, 2, \ldots N$) is the number of active mobile stations in cell n at the given moment. Using these notations, the downlink SIR γ_i for mobile station i in cell j can be given by

$$\gamma_i = \frac{p_i \cdot g_{ij}}{\sum_{\substack{k=1 \\ k \neq i}}^{M} p_k \cdot g_{il} + \eta_i} \tag{3.21}$$

where j and l denote the cells to which mobile stations i and k belong, respectively, and p_i is the downlink transmitter power for mobile station i by the base station in cell j, and η_i is the background noise received at mobile station i. Note that this model is general enough to represent DS-CDMA systems with matched filter receivers [31], [32] or TDMA/FDMA (Frequency Division Multiple Access) systems [21], by giving specific interpretations to the parameters.

Given the received the SIR γ_i at mobile station i, the probability of successful transmission is $Q(\gamma)$, where the function $Q(\gamma)$ depends on the details of the data transmission, including modulation, coding, interleaving, radio propagation, and receiver structure.

Each transmitter is equipped with a First In First Out (FIFO) queue. At the beginning of each time slot, a packet may arrive with probability λ, where $\lambda \in [0, 1]$. When a packet arrives at the time slot, it is first enqueued at the tail of the FIFO queue. Each arrival is assumed to be statistically independent from any other random events in the system (transmissions, interference, and other arrivals).

Packet transmission events are assumed to be statistically independent of each other and only one packet is transmitted in each time slot. If it was transmitted successfully, the packet is immediately removed from the queue and the transmitter attempts to transmit the next one in the queue in next time slot. In case of unsuccessful transmission, the transmitter continues to attempt transmission of the same packet in the following time slots, until this is successfully transmitted. We assume that the transmitter is immediately notified at the end of each slot, whether the packet transmitted in that slot was successfully received or not, through some highly reliable ACK/NACK process which takes negligible time.

3.3.3 Problem Formulation

We consider a radio channel which has a number of communication links in it, operating in slotted time. Each link can only observe the past and present states of its queue length, transmitted power and SIR received at the mobile stations. SIR measurements are fedback to its associated base station typically on a low-rate reliable reverse channel, which is perhaps a separate control channel. Based on this information, power control algorithm needs to decide what power to transmit in the next time slot. That is, each transmitter uses exclusively locally observed information to autonomously control its transmitted power.

Suppose at the beginning of one time slot, the queue length for mobile station i at its associated base station is l_i, the transmitted power for this mobile station from its associated base station during the time slot is p_i and the average SIR received by the mobile station is γ_i. Using these notations, we introduce two cost functions for power and buffering respectively:

1) The cost incurred by the power used in transmission during the time slot, denoted as $C_i^P(p_i)$. The specific cost function should reflect the expenses of power consumption to the user. There are at least two requirements for the cost function: $C_i^P(0)=0$, and that $C_i^P(p_i)$ is an increasing, positive function of p_i. In this thesis, we will use a linear cost function:

$$C_i^P(p_i) = \alpha p_i \tag{3.22}$$

where α is the price coefficient. Although α is a constant independent of p_i. It can be a function of environment factors such as user location and received interference. In fact, this kind of adaptive price setting can be very helpful in achieving fairness and robustness.

2) The cost incurred by buffering packets. It is a function of the queue length l_i, and is denoted as $C_i^B(l_i)$. The function must be an increasing, positive function of l_i.

Assume that at the beginning of the time slot, a packet for mobile i may arrive its associated base station with probability $\lambda_i \in [0,1]$ or no packet arrives with probability $1-\lambda_i$. The base station for mobile station i uses power p_i to transmit the first packet in its queue. The packet is received successfully at the receiver with probability $Q(\gamma_i)$. The total cost incurred, denoted by $C_i(p_i, l_i)$, is given by

$$\begin{aligned} C_i(p_i,l_i) = C_i^P(p_i) + \lambda_i[(1-Q(\gamma_i))C_i^B(l_i+1) + Q(\gamma_i)C_i^B(l_i)] + \\ (1-\lambda_i)[(1-Q(\gamma_i))C_i^B(l_i) + Q(\gamma_i)C_i^B(l_i-1)]. \end{aligned} \tag{3.23}$$

Rearranging the terms we get

$$C_i(p_i,l_i) = C_i^P(p_i) - Q(\gamma_i)X_i(l_i) + Y_i(l_i) \tag{3.24}$$

where

$$X_i(l_i) = \lambda_i[C_i^B(l_i+1) - C_i^B(l_i)] + (1-\lambda_i)[C_i^B(l_i) - C_i^B(l_i-1)] \tag{3.25}$$

$$Y_i(l_i) = \lambda_i C_i^B(l_i+1) + (1-\lambda_i)C_i^B(l_i)\,. \tag{3.26}$$

The power control problem can be formulated as

$$\min_{p_i \geq 0} C_i(p_i, l_i)\,. \tag{3.27}$$

3.3.4 Proposed Power Control Algorithm

For each *i*, the minimum cost occurs at a power level for which the partial derivative of $C_i(p_i, l_i)$ with respect to p_i is zero:

$$\frac{\partial(C_i(p_i, l_i))}{\partial p_i} = 0\,. \tag{3.28}$$

Referring to Equations (3.21), (3.22) and (3.24), we can express the derivative of cost with respect to power as

$$\frac{\partial(C_i(p_i, l_i))}{\partial p_i} = \alpha_i - \frac{dQ(\gamma_i)}{d\gamma_i} \cdot \frac{g_{ij}}{I_i} \cdot X_i(l_i) \tag{3.29}$$

where $I_i = \sum_{k \neq i} p_k \cdot g_{il} + \eta_i$ is the total received interference of mobile station *i*. Therefore, with $p_i \geq 0$, the necessary condition for mobile station *i* to minimize the cost is

$$Q'(\gamma_i) = \frac{\alpha_i I_i}{g_{ij} X_i(l_i)}\,. \tag{3.30}$$

In order to differentiate (3.30) with respect to γ_i, we need to know the function form of $Q(\gamma_i)$. A natural function form arising in some standard radio technologies [35] is given by:

$$Q(\gamma_i) = \frac{\gamma_i}{a \cdot \gamma_i + b} \tag{3.31}$$

with $a \geq 1$ and $b \geq 0$ (which guarantee that $Q(\gamma_i) \leq 1$). Using this equation, we can find the solution of the power control problem (3.30) to be:

$$p_i = \begin{cases} \frac{1}{a}\left(\sqrt{\frac{b}{\alpha_i}\cdot\frac{I_i}{g_{ij}}X_i(l_i)} - b\frac{I_i}{g_{ij}}\right), & \frac{I_i}{g_{ij}} < \frac{X_i(l_i)}{\alpha_i b} \\ 0, & \frac{I_i}{g_{ij}} \geq \frac{X_i(l_i)}{\alpha_i b} \end{cases}. \tag{3.32}$$

Clearly, p_i is a function of I_i / g_{ij} and the queue length l_i, for given power cost price coefficient α_i. From equation (3.32), we derive an iterative algorithm by putting in the time index. The power assignment for the current time step p^k is determined by the last channel status I^{k-1}/g_{ij} and the current queue length l^{k-1}. From equation (3.21), we can replace I_i / g_{ij} with p_i / γ_i in equation (3.32). The corresponding power assignment becomes:

$$p_i^{k+1} = \begin{cases} \frac{1}{a}\left(\sqrt{\frac{b}{\alpha_i}X_i(l_i^{k+1})\frac{p_i^k}{\gamma_i^k}} - b\frac{p_i^k}{\gamma_i^k}\right), \gamma_i^k > \frac{\alpha_i b p_i^k}{X_i(l_i^{k+1})} \\ 0, \qquad \gamma_i^k \leq \frac{\alpha_i b p_i^k}{X_i(l_i^{k+1})} \end{cases}. \tag{3.33}$$

Note that $p_i = 0$ is on the constraint boundary of (3.27), so the optimal power at step k is either p_i^k or 0, whichever results in a less total cost. This gives us a distributed power control algorithm for minimizing total cost. We call this algorithm *Power & Backlog Sensitive Power Control Algorithm (PBSPC).*

The mechanism of PBSPC is shown in Figure 3-2. We define the signal to interference ratio reconstructed by the base station at k^{th} time slot as $\hat{\gamma}_b^k$. On the mobile station side, $\hat{\gamma}_b^k$ is kept tracking by the mobile station, defined as $\hat{\gamma}_m^k$. Each mobile station measures the average received SIR denoted by γ^k , over the k^{th} time slot. The

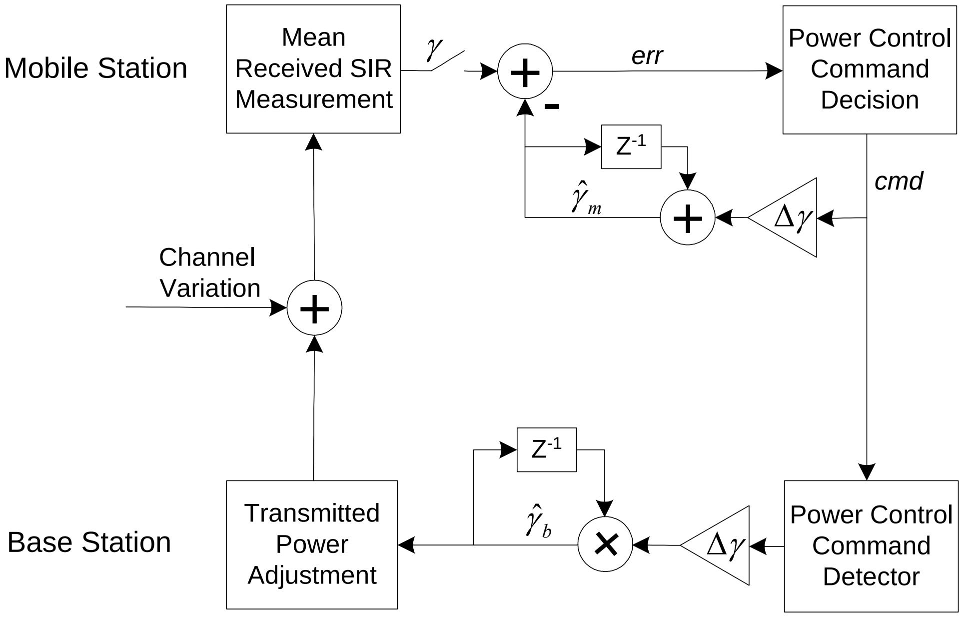


Figure 3-2 Block diagram of PBSPC

mobile station compares the received SIR γ^k with $\hat{\gamma}_m^{k-1}$. If $\gamma^k > \hat{\gamma}_m^{k-1}$, the mobile station will send a power control command to the base station to increase $\hat{\gamma}_b^k$. On the other hand, if $\gamma^k < \hat{\gamma}_m^{k-1}$, the mobile station will send a power control command to the base station to decrease $\hat{\gamma}_b^k$. Concurrently with the transmission of the power control command, the mobile station updates $\hat{\gamma}_m^{k-1}$ using the same command that has just been sent. The command, denoted by *cmd*, is a digital number, and is obtained by the rule described below:

$$cmd = \begin{cases} 1 & \gamma^k > \hat{\gamma}_m^{k-1} \\ 0 & \gamma^k < \hat{\gamma}_m^{k-1} \end{cases} \tag{3.34}$$

The base station will detect the power control command and updates $\hat{\gamma}_b^k$ by:

$$\hat{\gamma}_b^k = \begin{cases} \hat{\gamma}_b^{k-1} \cdot \Delta\gamma & cmd = 1 \\ \hat{\gamma}_b^{k-1} / \Delta\gamma & cmd = 0 \end{cases} \tag{3.35}$$

where $\Delta\gamma$ is the step size.

With the reconstructed SIR $\hat{\gamma}_b^k$ experiencing at the mobile station and the local information about queue length l^{k+1} and the transmitted power p^k during the last time slot, the transmitted power p^{k+1} can be achieved from (3.33). In summary, the iterative procedure of the PBSPC is as follows:

Algorithm PBSPC:

1. Each mobile station measures the average received SIR value γ^k over the k^{th} power control sampling period, and compares it with $\hat{\gamma}_m^{k-1}$. The difference is sent to the base station using one step feedback. The mobile station updates $\hat{\gamma}_m^k$ using (3.35).

2. The base station detects the power control command *cmd*, updates its $\hat{\gamma}_b^k$ using (3.35).

3. With the updated SIR value $\hat{\gamma}_b^k$, the value of last transmitted power used, p^k and the queue length of the base station l^{k+1}, the base stations decide the next transmit power p^{k+1} according to (3.33).

4. Let $k \leftarrow k+1$, and go to step 1.

3.3.5 Discussion and Extension of PBSPC

It is worthwhile to illustrate several desirable properties and a high degree of flexibility that can be achieved by properly tuning the parameters in our proposed algorithm.

The cost function represents the cost incurred by the transmitted power used and the cost incurred by buffering packets in the queue. By selecting different power cost function and buffering cost function, we can have a series of power control functions.

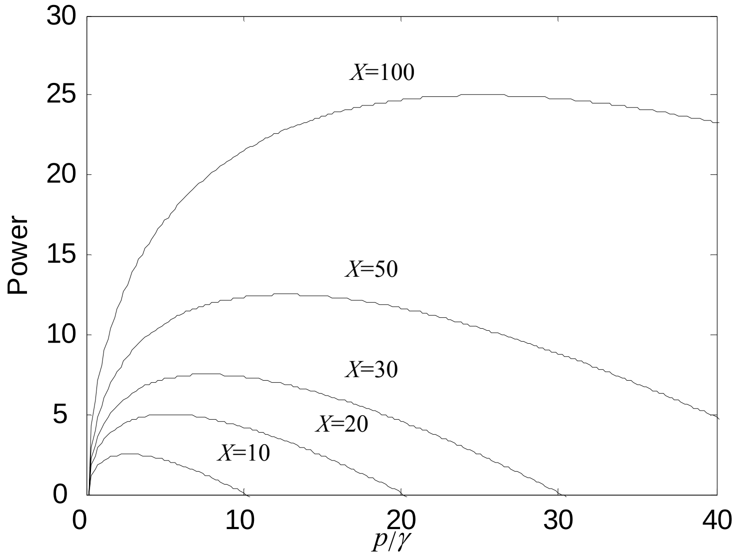


Figure 3-3 Plots of the power function of (3.33) for various X, assuming $\alpha = a = b = 1$.

In Figure 3-3, we show a series of power control functions with different buffering cost. From the figure we note that, for any given fixed queue length l, the algorithm generally have three operational phases that it goes through as p / γ increase. From equation (3.21) we know that p / γ equals to I / G which represents the quality of the transmission environment experienced by the link. It is clear that, at first when the quality of the transmission environment is favorable, the algorithm tries to overcome the interference and the channel loss by increasing the power quickly, thus achieving high SIR and more successful transmission. As the channel quality decreases below some points, it starts to decrease the power, beginning to buffer packets rather than transfer them at an arbitrarily high power. When the channel quality decreases further below another point, the power decreases to zero and the transmission is suspended.

The idea behind this is that, when the quality of the transmission environment is favorable, we make the best use of the channel by using relatively less power to achieve high SIR and more successful transmission while causing less interference to other users. When the quality of the transmission environment becomes worse, it doesn’t make sense to use much power to transmit because the success rate drops. Therefore, we’d better buffer the packets until the quality of the transmission environment becomes better before we start transmission again. As the packets accumulate in the buffer, the buffering cost increases. This pressure pushes the algorithm to increase the suspension threshold, thus makes it more aggressive to transmit at a relatively high power, which causes the packets to be transmitted successfully and in turn decreases queuing cost. That is to say, the mechanism always tries to tradeoff between the cost of power and the cost of buffering under certain quality of the transmission environment.

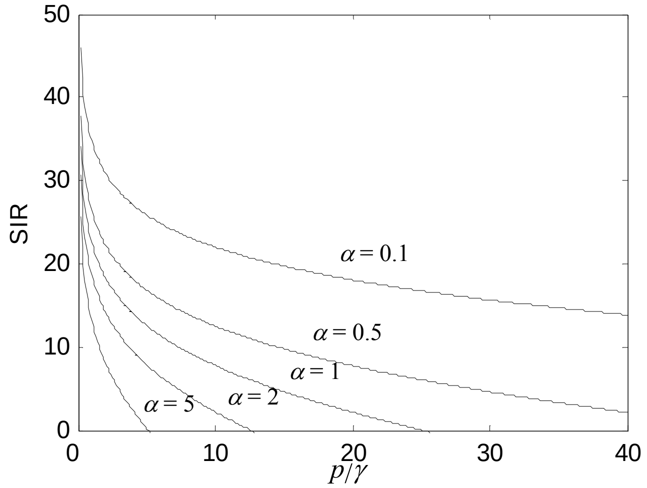


Figure 3-4 Plots of target SIR of (3.36) for various α, assuming $a = b = 1$ and $X(l) = 100$.

Now let’s turn to some properties of the power cost coefficient α. From the power control equation (3.33), we can deduce the target SIR that the algorithm tries to achieve:

$$\gamma_i^T = \begin{cases} \frac{1}{a}\left(\sqrt{\frac{b}{\alpha_i} X_i(l_i)\frac{\gamma_i}{p_i}} - b\right), & \gamma_i > \frac{\alpha_i b p_i}{X_i(l_i)} \\ 0, & \gamma_i \le \frac{\alpha_i b p_i}{X_i(l_i)} \end{cases}. \tag{3.36}$$

We illustrate the target SIR vs p/γ with different power cost coefficient α in Figure 3-4. We find that for the user with a smaller value of α, the target SIR is higher. Thus, a user having a smaller value α is more aggressive in transmitting its packets. In other words, transmission is discouraged by a larger value of α as well as by a hostile transmission environment (in terms of p/γ). It is natural if we adopt adaptive price setting mechanisms to help solve problems such as fairness, robustness, and admission control. We should also point out that price setting can be user dependent. For example, an automobile-mounted phone has a much larger supply of power than a handheld one, so its user may choose a lower price coefficient.

From the network point of view, the network can use the price coefficient as an effective way to manage resource to maximize revenue. When the transmission environment is desirable, we should set a low price, allowing the user to enjoy good quality of service (high SIR or high transmission rate). When the transmission environment is not so good, we could set high price, to impede the transmission which result in saving power and mitigating the interference it impedes to the network. Thus, it is desirable to have a price coefficient that is adaptive to the transmission environment. As it was shown in earlier section, a good measurement for the transmission environment experienced by user *i* is p/γ, thus a natural choice is to set the price coefficient as an increasing function of p/γ. For example, a simple adaptive setting can be

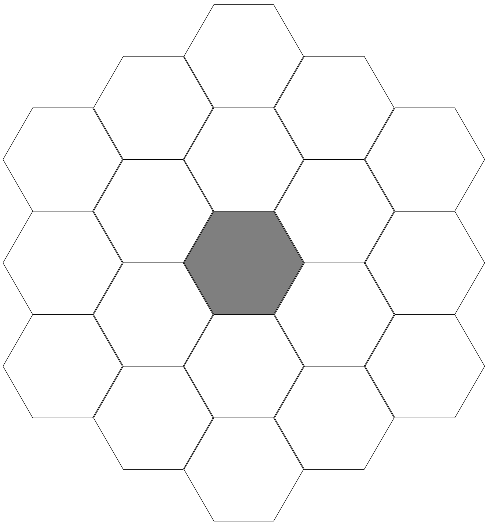


Figure 3-5 Cell layout.

$$\alpha_i = \alpha \frac{p_i}{\gamma_i} \tag{3.37}$$

where α is a constant. Then, we have a different power control algorithm,

$$p_i^{k+1} = \begin{cases} \frac{1}{a}\left(\sqrt{\frac{b}{\alpha} X_i(l_i^{k+1})} - b\frac{p_i^k}{\gamma_i^k}\right), & \gamma_i^k > p_i^k \sqrt{\frac{\alpha b}{X_i(l_i^{k+1})}} \\ 0, & \gamma_i^k \le p_i^k \sqrt{\frac{\alpha b}{X_i(l_i^{k+1})}} \end{cases} . \tag{3.38}$$

Finally, one should observe that by choosing different buffering cost function, we can make the algorithm more power sensitive or more buffering sensitive for a given load; by choosing different power cost coefficient, we can make the algorithm be favorable to some of the users while suppressing the others.

3.4 Performance Evaluation

Performance evaluation is carried out by computer simulation. For comparison, we simulated three different algorithms, Fixed SIR power control algorithm, PCMA-1 and PBSPC. The purpose of the simulations is to show that PBSPC achieves capacity improvement with less power consumption.

3.4.1 Simulation Parameters

To derive some numerical results, more specific system model and additional assumptions for simulation are described in the following.

1. The simulation model consists of 19 base stations in a hexagonal-grid configuration as shown in Figure 3-5. Base stations use omnidirectional antennas and are located at the center of the cell. Each cell is assumed to be identical with 6 mobile stations randomly distributed with uniform distribution in the cell. The propagation loss and long-term fading values are generated in the beginning of each simulation cycle and kept the same value during the same cycle. The air interface is DS-CDMA with processing gain $PG = 16$.

2. The radio signal is attenuated by (1) path loss proportional to d^4, d being the propagation distance, (2) log-normal shadowing with 0 dB expectation and 8 dB standard deviation, and (3) motion-induced multipath (Rayleigh) fading. The maximum Doppler frequency normalized by the power control sampling rate $f_D T_p$ for each user is a random variable uniformly distributed between 0.01 and 0.1. This corresponds to a speed range of 6 to 60 km/hr, assuming 900 MHz radio frequencies and a power control sampling rate $1/T_p$ on the order of 500Hz.

3. The probability of success used for all three power control algorithms is that of (3.31) with $a = b = 1$. The function $X(l) = 5l + 10$ is used for both PCMA-1 and PBSPC. The constant 10 is used to keep an acceptable throughput for very small queue length, so as to suppress excessive delays at very low arrival rates. The

coefficient 5 is used to make the algorithm more sensitive to the queue length in order to suppress the queue length when packets accumulate at the buffer.

4. The power control step size is 0.5 dB. The delay time due to control processing is ignored.

5. Maximum transmitter power limitation is an important parameter that determines the dynamic range of power control. As defined in [33], we use a maximum normalized power parameter Λ = 10 dB, where Λ is the ratio of mean signal (assuming maximum transmit power) to thermal noise power, for a shadow-fading value of 0dB, at the maximum distance within each coverage area.

6. The use of space diversity, soft-handoff, and antenna sectorization is not included in the simulation program.

7. The simulations are repeated for about 4,000 samples of random user location, log-normal shadowing, and multipath fading processes. Statistics are accumulated only for the center base station to avoid edge effects.

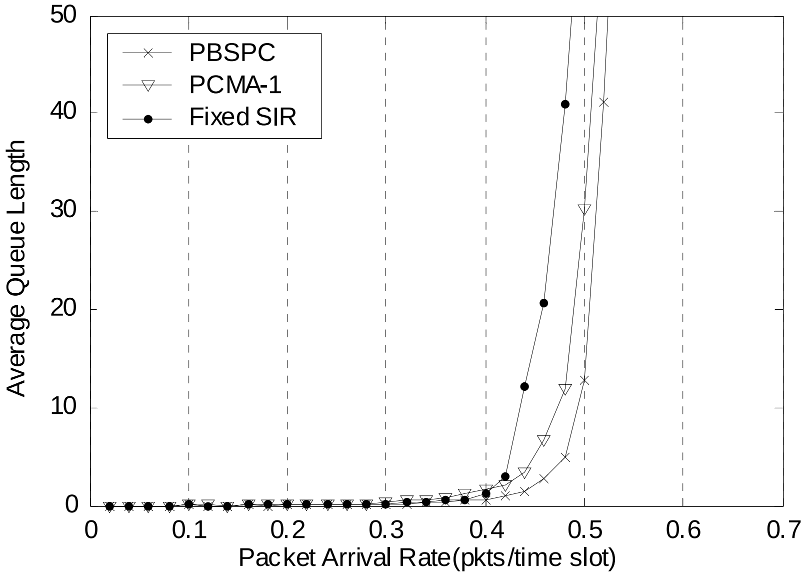


Figure 3-6 Average queue length vs packet arrival rate λ for three power control algorithms.

3.4.2 Average Queue Length

Figure 3-6 plots the queue length as a function of the packet arrival rate λ for three different algorithms: Fixed SIR power control algorithm, PCMA-1 and PBSPC. It can be found when the packet arrival rate below 0.4 (approximately), all three algorithm can keep their queue length stable. When the packet arrival rate is beyond 0.4, packets begin to accumulate in the queue of Fixed SIR scheme which means the arrival rate is already larger than the system throughput. At the same time, both PCMA-1 and PBSPC still can keep their queue length stable but at a larger value of l. For PCMA-1, packets begins to accumulate when the arrival is approximately 0.47 and for PBSPC it is approximately 0.5. Thus, PBSPC achieves 20 percent of throughput improvement over Fixed SIR power control scheme and 6 percent of improvement in throughput over PCMA-1. We also should note PBSPC always keep a shorter queue length compared to Fixed SIR scheme and PCMA-1 when arrival rate is less than system throughput, which means that PBSPC achieves less delay in transmission.

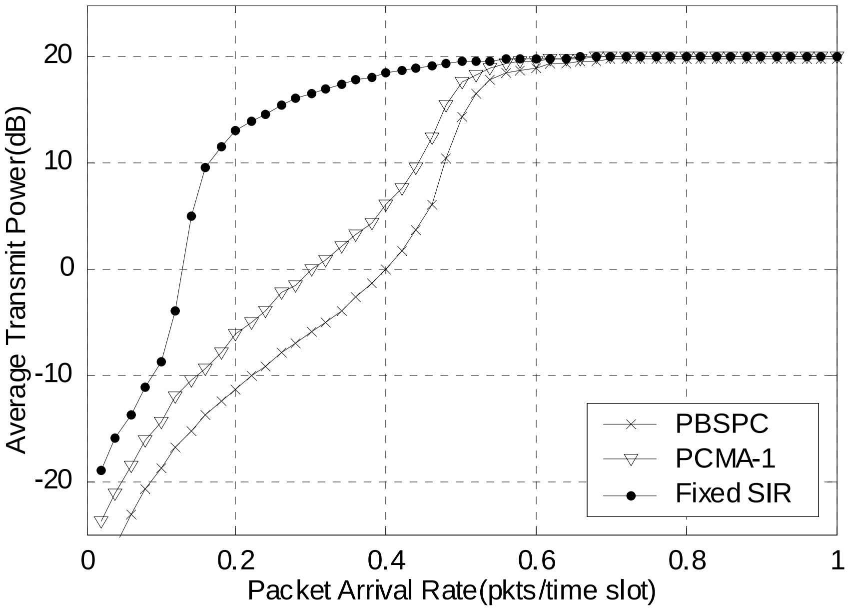


Figure 3-7 Average power vs packet arrival rate λ for three power control algorithms.

3.4.3 Average Transmit Power

To compare the power consumption of the three algorithms, we plot the average power vs packet arrival rate λ in Figure 3-7. Note that average power is defined as the total consumed power divided by the total power controlled time slots. This figure illustrates that as the packet arrival rate increases, Fixed SIR power control scheme quickly increases its power consumption at a very early stage and reache the maximum transmission power limitation when the packet arrival rate is approximately 0.4, which is equal to its throughput. While both PCMA-1 and PBSPC maintain lower average power level until the maximum power limitation is reached near. Compared to PCMA-1, PBSPC also achieves some power saving especially when the load is smaller than the system throughput.

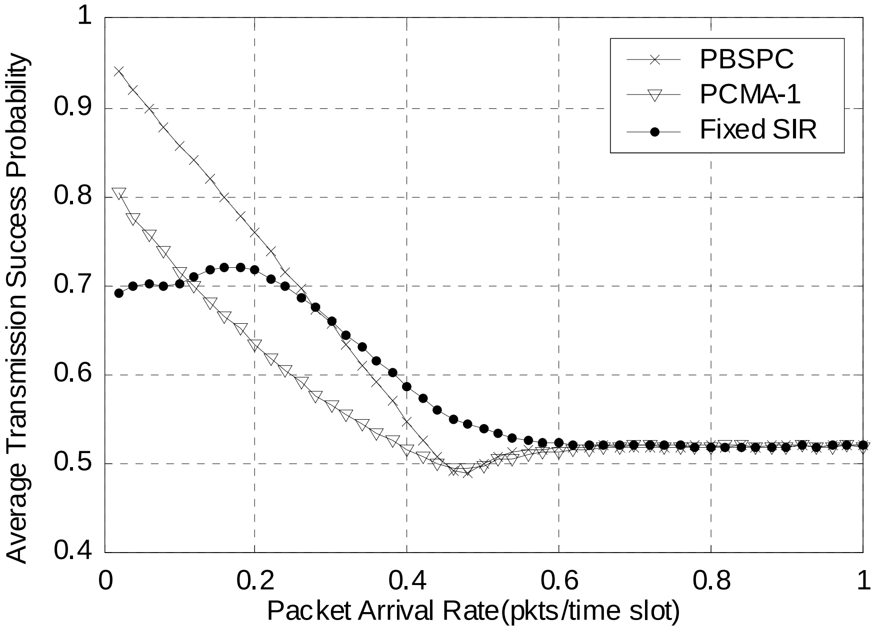


Figure 3-8 Average transmission success probability vs packet arrival rate for three power control algorithms.

3.4.4 Average Transmission Success Probability

Figure 3-8 shows the performance in terms of the average transmission success probability. We find that the curves of the three algorithms overlap when the packet arrival rate is greater than 0.6, which is larger than the throughput of all three algorithms. It is obvious from Figure 3-7 that when the system load is larger than the system throughput, all three algorithms always use the maximum power in transmission, resulting in same transmission success probability. When packet arrival rate is smaller than system throughput, PBSPC performs much better than PCMA-1. The performance difference decreases as the system load increases. It is because when the system load is light, less communication links are active, which results in less interuser interference. The channel variation is dominated by the multipath fading. Therefore PBSPC performs better than PCMA-1, because PBSPC can successfully combat the multipath as we have analyzed in earlier sections. As the system load increases, more and more transmitters try to transmit simultaneously. The interuser interference becomes more significant, which can be taken care of by either PBSPC or PCMA-1, thus the performance improvement decreases. Compared to Fixed SIR scheme, PBSPC achieves less performance improvement in terms of transmission success probability. However, one should note that Fixed SIR scheme achieves its performance at the expense of much higher power consumption.

3.5 Conclusion

In this chapter, a power & backlog sensitive power control algorithm has been proposed to control the power transmission for data traffic in a CDMA mobile radio system. The transmission performance of downlink channels using this method has been evaluated by computer simulations. Significant improvement over the existing algorithm has been demonstrated. Although the proposed algorithm is still non-cooperative in nature, some degree of cooperation emerges: a user will automatically decrease its transmission power (and may even turn off transmission) when the transmission environment is not desirable. Although our algorithm provides a promising framework for distributed power control of data traffic in cellular wireless system, there are still many open problems. For example, how to translate different QoS requirements into cost functions leading to a solvable power control problem, and how to manage the performance trade-offs between QoS, power, delay, and traffic load are still topics requiring further research.

CHAPTER FOUR

Scheduled Data Dissemination in Wireless Environment

In this chapter, the concept of data dissemination in wireless systems is first introduced. The performance metrics and design issues are discussed and the related work is briefly reviewed. We then consider a finite linear network with mobility limited mobile nodes and propose a new class of scheduling strategies for the data dissemination in mobile networks. Finally the algorithm is evaluated by simulations.

4.1 Wireless Data Dissemination

Traditionally, the mode of information dissemination is usually of request-response style. For example, users submit queries to servers that respond with the answers. However, this model is no longer adequate in a wireless computing environment. First, the wireless channel is unreliable and insecure. Second, the bandwidth available is very limited varying from 1.2Kbps for slow paging channels, through 19.2 Kbps (e.g. Cellular Digital Packet Data) to about 10 Mbps for the wireless LAN. Third, the environment is essentially asymmetric with a large number of mobile users accessing a small number of servers. Fourth, battery-operated portable devices can typically operate only for a short time because of the short battery lifespan (before recharging or changing of batteries becomes necessary). Thus, clients may frequently disconnect to conserve energy. To overcome these limitations, there has been a proliferation of research efforts on designing data delivery mechanisms to support wireless computing more effectively.

4.1.1 Fundamental Models for Wireless Data Dissemination

There are two fundamental delivery methods for wireless data dissemination [36]–[38]:

On-Demand Access: A client sends data requests to the server, and the server returns the results to the client individually.

Data Broadcast: The server periodically broadcasts information to the entire client population, and the clients monitor the broadcast channel to retrieve the data of their interest.

On-demand access provides fast response for a lightloaded system but the performance will deteriorate rapidly as the workload increases. On the contrary, data broadcast can scale up to a very large client population and fit well to an asymmetric communication environment.

4.1.2 Performance Metrics

In wireless computing environments, several criteria can be used to evaluate the performance of a data delivery method:

1. *Responsiveness*: The success of a data delivery scheme is determined by its ability to get the requested data to the clients quickly, In this regard, the first important metric is the average access time. The average access time is the amount of time spent on average from the instant the request is made to the time that the

requested object is received. The second metric is the worst-case access time that measures the maximum access time for any client request to be satisfied. The former provides a metric for the overall system performance, while the latter indicates individual performance.

2. *Robustness*: The effectiveness of a scheme is also determined by how well it adapts to workload or environmental changes. The scheme should be able to support increasingly larger population (i.e. scalable), or be robust to changing access patterns. A scheme that is rigid will be useful only in static environments where the workload is fairly stable, whereas a robust scheme can be employed for dynamically changing environments.

3. *Packet Efficiency*: In cellular telephony, one of the most important criteria is the cell capacity that measures the number of telephone calls which can be handled by the cell per unit time. In wireless computing, a query will be the analog of a telephone call. Thus, the criterion queries per Hz that measures the cell capacity in terms of the number of queries that the local mobile support station can handle per unit time is expected to provide a better indication of the efficiency.

4. *Power Efficiency*: Battery power, being a precious resource for portable devices, has to be minimized. As Such, the metric queries per watt has been proposed [4] to measure the power efficiency of a method. Queries per watt essentially indicates the number of queries that a client can process with 1 watt power of battery.

5. *Tuning Time*: Another metric that is commonly used as an indication of the energy consumption of an information dissemination strategy is the tuning time. The tuning time measures the amount of time that a mobile client listens to the channel.

Thus, it measures the time during which the client stays in the active mode and therefore determines the power consumed by the client to retrieve the relevant data.

Ideally, we hope to have a robust and energy efficient data delivery method that minimizes both the average and worst-case access time. However, some of these metrics may conflict. For example, building "indexed on air" for broadcast data can help to minimize energy consumption but the extra metadata lengthen the broadcast length and hence increase the access time. Thus, it is a challenge to design schemes that balance these factors.

4.1.3 Design Issues in Wireless Data Dissemination

There are many challenges that have to be addressed to realize the full potential of wireless computing [39]. These include networking issues (e.g. mobile IP, mobility management, congestion control in TCP), power management issues (e.g. operation systems level, I/O and file systems level), system issues (e.g. disconnection support, locations dependent queries, recovery) and information dissemination issues. We shall focus only on the information dissemination issues in this thesis.

Traditionally, users query data base using the request-response mode. However, because of the unique characteristics (or rather limitations) of wireless computing environments, the basic request-response approach is not the most desirable approach to be deployed in such environments. For example, the request-response approach does not scale well in terms of the number of queries/users that can be supported. Instead, a data delivery mechanism should have the following feature:

1. *Efficient wireless bandwidth utilization*: The wireless bandwidth can quickly

become congested as the client population increases. Moreover, as popular data objects may be continuously requested within a short span of time, multiple copies of the same data objects may be disseminated, leading to poor bandwidth utilization. Mechanisms are needed to effectively utilize the limited bandwidth. One promising approach is to periodically broadcast popular objects without explicit user requests, whereas those that are less frequently demanded can be accessed using the request-response approach.

2. *Efficient and effective scheduling strategies at the server*: The asymmetry between the large client population and the small number of servers can quickly generate huge spikes in the load at the servers. To handle demand-driven requests, the servers must employ efficient scheduling strategies that scale well and balance the average and worst case response time of queries. Such schemes can exploit the fact that an object that is broadcast will be received by all clients who have requested it to reduce multiple transmission.

3. *Energy-efficient data access for battery-powered portable devices*: The need to recharge or change batteries frequently will be a big obstacle to the wide acceptance of wireless technology. Designing energy-efficient solutions at the software level will be critical. Energy efficient solutions make it possible to use smaller and less powerful batteries to run the same set of applications for the same time. Small batteries are important from the portability point of view since devices like palmtop can be more compact and/or weigh less. Moreover, with the same batteries, the unit can run for a long time without the troublesome of changing the batteries so often. This can also result in substantial monetary savings and avoids recharging. In addition, we can minimize the disposal of batteries that can lead to environmental hazard. Software solutions should take

advantage of energy-management features that already exists at the hardware level. For example, the algorithms should minimize the time the CPU spends in the active mode and maximize the time the CPU spends in the doze mode, i.e. keep the receiver off most of the time or spin down the disk, etc.

4. *Support for disconnection*: Clients are expected to be frequently disconnected, be it voluntarily to reduce power consumption or bandwidth use, or forced when the portable device encounters a region where there is no network coverage. Thus, mechanisms are needed to allow clients to continue to operate even when disconnected. Caching of data is one approach that has been shown to be effective. However, disconnection makes it difficult to ascertain the validity of data objects that are cached when the mobile clients resume connection. Thus, novel cache invalidation strategies are needed.

5. *Support for secured and reliable transmission*: While existing techniques for secured transmission on wired environment can be applied to wireless context, the fact that the transmitted information can be picked up by everybody calls for more secure methods to be developed. It is also necessary to deal with the unreliable network in order to minimize retransmission and long waiting time. The basic approach of injecting redundancy into the broadcast has to be adapted and controlled since replication broadcast information implies a longer broadcast length.

From the discussion above, it is clear that we need novel data delivery mechanisms in wireless environments.

4.2 Related Works

Recently, there have been research efforts [40]-[43] in data sharing and dissemination in mobile (not necessarily ad hoc) environments. Such dissemination is peer to peer in nature: a mobile node may be a requester of a popular data file, such as an image or an audio clip; and it becomes a supplier of this data packet after it retrieves the packet from other supplier(s). Meanwhile, unlike wired or wireless cellular networks, every mobile node has a limited transmission range. Therefore, two mobile nodes can communicate with each other directly, only if they are in the transmission range of each other. Due to the free movement of mobile nodes, connections between peers can be disconnected very frequently.

In [40], nodes with all the data packets serve as suppliers for others. They broadcast their packet segments periodically. When a node receives all the packet segments, it becomes a supplier. In [42], it was shown that with a two hop relay model, the steady-state pernode throughput scales with the number of nodes. Unicast communications and content distribution using single hop multicast are considered in [41] and [42] respectively. In order to expedite data dissemination, a node also relays packets for other nodes if it has not done so for some time. In [43], a social contract was introduced to allow packet exchanges among mobile nodes while keeping the network essentially noncooperative. The above works either randomly choose the data packet to be disseminated or use a greedy method [43] which is based on an unrealistic assumption that each node has full knowledge of the circulation of each packet within the network.

We demonstrate in this chapter that some kinds of scheduling of the packet dissemination in a mobility limited system can greatly increase the system

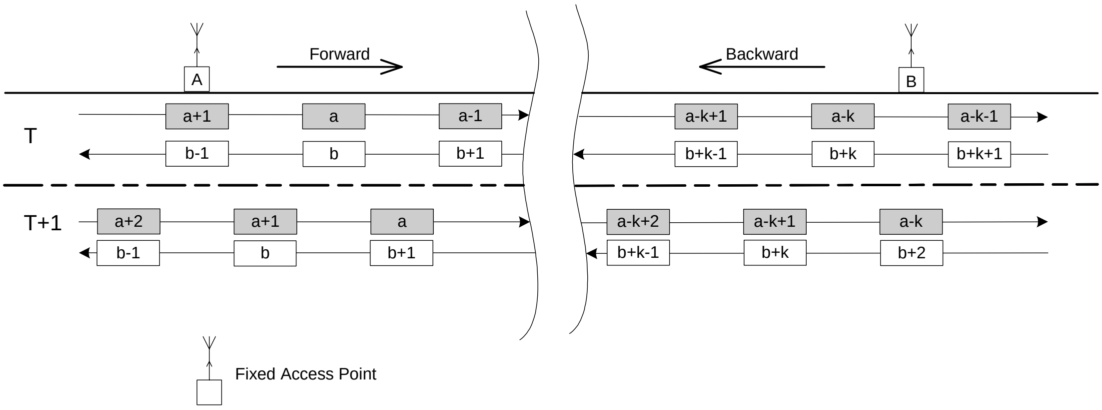


Figure 4-1 Illustration of the system model.

performance.

4.3 System Model

To demonstrate the characteristics of scheduling in popular data sharing and dissemination, a simple finite linear system as shown in Figure 4-1 is considered. Two fixed access points (AP) are deployed at both ends of the system. Nodes are injected into the system regularly with the same interval from both ends. All nodes move at the same speed towards the other end. Therefore, no nodes will overtake other nodes or being overtaken as it traverses the line. A node will only encounter nodes moving in the opposite direction. Assuming there are N nodes moving in one direction. The total number of nodes in the system is $2N$. We refer to this N as the size of the network.

The popular packet which is of interest for all nodes is divided into K packet segments of same size (for the rest of this thesis, we will refer the packet segments as *packets*). All packets are initially cached in the two access points. Each node has no packet in its cache when it enters the system. The node will download its initial packets from the access point. Since nodes move at a constant speed, the number of nodes in the system is fixed. Therefore, the size of the network can be defined by the number of nodes in it.

In our system we don't consider the channel conditions and all nodes and the access points use determinate transmission power, thus the access points and all nodes will have the same size of coverage area. When a node approaches the coverage area of the access point, it downloads packets from the access point. If two nodes enter each others coverage area so that they can exchange packets if possible, these two nodes are called a *node pair*. The interval between nodes along one direction is separated by the distance of at least two coverage areas wide. This is to avoid one node falling into two other nodes' coverage area simultaneously. Since nodes are distributed regularly along the line and they move with constant speed, we can divide time into *time units* such that during each unit of time, each node will only encounter one and only one node moving in the opposite direction. Given a particular system bandwidth, the size of packets can be chosen such that during each unit of time only two packets can be transmitted. That is, either a node can download two packets from an access point or a node pair can exchange one packet when they encounter.

There are two types of data transmission in this system. One is *packet exchange*. When two nodes encounter, *packet exchange* may happen. The exchange is done in two steps. First, these two nodes check whether the other node has any packets that it doesn't have. If both nodes find packets that can be downloaded from their counterpart, they will exchange packets. When there are more than one packet that can be chosen, a node must decide which packet to download. In our system, the random strategy is considered. A node randomly selects a packet it does not have from the other node. If one node has all the packets its neighbor has, then no packet exchange will happen. A node can not download packets from another node without providing packets for that node. This is different from the typical assumption in most wireless networking literature, which is often unrealistic. If a node provides services for others without getting any pay, the node will consumes its own bandwidth and exhaust it battery which is of crucial importance for a mobile user. A node therefore has no immediate incentive to forward packets for others [43]. Another type of data transmission is *packet download*. When a node is in the coverage area of the access point, two packets will be downloaded from the access point.

4.4 Proposed Scheduling Strategies

In order to improve the probability of packet exchange between two nodes in the subsequent packet exchange, we propose two scheduling schemes when nodes encounter and download packets from the access points. One is the ID-Schedule where packet download follows a determinate pattern. Assuming K is the total number of packets cached in the access points. In one of the two access points, packet download follows increasing order, i.e. packets #1 and #2 are downloaded for the first arriving node, packets #3 and #4 are downloaded for the second arriving node and etc. In another access point, packet download follows decreasing order, packets #K and #$K-1$ for the first arriving node, packets #$K-2$ and #$K-3$ for the second arriving node and etc. The other is II-Schedule which is similar with ID-Schedule. The difference is packet download follows increasing order in both access points for II-Schedule. To demonstrate the effectiveness of our schedule strategies, we also consider a random strategy as a comparison. For the random strategy, a node randomly selects two packets it does not have for downloading from the access point.

We evaluate the system performance in terms of throughput and transmission success probability to describe how quickly packets are disseminated and how effectively the packets are disseminated. We define C_T as the average number of packets disseminated per unit time in the system. It describes the rate of the system in disseminating data. C_p is defined as the average number of packets disseminated per encounter per unit time. Dissemination success probability, P_{Succ}, is defined in terms of the probability that a node successfully gets all packets to reconstruct the original packet when it leaves the system. If a node leaves the system without successfully acquiring all the packets, then all the former packets it acquires will be of no use and will be discarded. This is a great waste of the system resource and poor service is provided to the subscribers. Thus low dissemination success probability is not tolerable. A minimum dissemination success probability must be achieved from both the service provider and subscriber's points of view.

4.5 Performance Analysis

In this section, we try to do some mathematical analysis to the system described in Figure 4-1. We only consider the random download strategy here. Consider that all nodes move at a constant speed on the line from access point A to B. For forward traffic, we denote N as the number of nodes between access points A and B. Apparently there are N nodes traveling backwards at the same time. During the time one node traverses the distance between the access points, it will encounter $2N$ nodes moving in the opposite direction.

As we mentioned earlier, nodes are injected into the system for both access points. They will download two packets from the access points as their first two packets and then move toward the other end. When two nodes encounter, packets are to be exchanged with a probability which depends on the packet contents each node has. We define this probability as Pe. We make a key assumption here: Independent uniform content distribution. Given that node i has obtained m packets, all combinations of m

out of K packets are of equal probability, independent of the packets held by all other nodes. If two nodes, say node i and node j are to exchange packets, suppose node i has f_i packets and node j has f_j packets respectively in their caches. An exchange between the nodes will occur only when the collection of packets of one node is not the subset of the other's collection. Assuming, without loss of generality, that $f_i \le f_j$, an exchange failure occurs if the collection of packets of node i, f_i, is the subset of the node j's collection, f_j. The packet exchange probability Pe can be easily deduced:

$$P_e(f_i, f_j) = 1 - \frac{\binom{f_j}{f_i}}{\binom{K}{f_i}} \qquad 0 \le f_i \le f_j \le K. \tag{4.1}$$

Suppose at a time, node i encounters node j during its movement and node j is the m^{th} node that node i has encountered, where ($1 \le m \le 2N$). At this time, node i has f_i packets in its cache. Apparently, we can know that ($2 \le f_i \le m+1$). As to node j, node i is the $(2N - m + 1)^{\text{th}}$ node that node j encounters. Therefore, if node j has f_j packets in its cache, we know ($2 \le f_j \le 2N - m + 2$).

We denote $P^m(f)$ as the probability that one node has f packets in its cache at the beginning of its m^{th} encounter. The probability that packet exchange will occur during this encounter is given by

$$P_E^m(f) = \sum_{l=2}^{2N-m+2} P^{2N-m+1}(l) P_e(f, l), \qquad 1 \le m \le 2N. \tag{4.2}$$

Suppose after an encounter, a node has $(f + 1)$ packets in its cache, this may be due to this node had f packets in its cache at the beginning of that encounter and it exchanged a packet during that encounter, or due to that node had $(f + 1)$ packets in its cache at the beginning of the encounter and no file was exchanged during that encounter. Thus we acquire

$$\begin{aligned} P^{m+1}(f+1) = P^{m}(f+1)(1-\sum_{l=2}^{2N-m+2} P^{2N-m+1}(l)P_e(f+1,l)) \\ +P^{m}(f)\sum_{l=2}^{2N-m+2} P^{2N-m+1}(l)P_e(f,l). \end{aligned} \tag{4.3}$$

The overall packet exchange probability during an encounter is

$$P_E^m = \sum_{f=2}^{m+1} P^m(f) \sum_{l=2}^{2N-m+2} P^{2N-m+1}(l)P_e(f,l) \qquad 1 \le m \le 2N\ . \tag{4.4}$$

Solving equations (4.1)-(4.4), we can get the packet exchange probability during each encounter. As we mentioned earlier, if node j is the m^{th} node that node i encounters, then at the same time, node i is the $(2N - m + 1)^{th}$ node that node j encounters. Obviously for both nodes i and j, the packet exchange probability during that encounter is same. Therefore,

$$P_E^m = P_E^{2N-m+1}\ . \tag{4.5}$$

Then the system throughput C_T and C_p can be expressed by the packet exchange probability as follow

$$C_T = \frac{1}{2}\cdot\sum_{m=1}^{2N} P_E^m = \sum_{m=1}^{N} P_E^m \tag{4.6}$$

$$C_p = \frac{1}{N}C_T = \frac{1}{N}\sum_{m=1}^{N} P_E^m\ . \tag{4.7}$$

The dissemination success probability P_{Succ} is the probability that a node has K packets in its cache after it has encountered its $2N^{th}$ node. Therefore it can be given

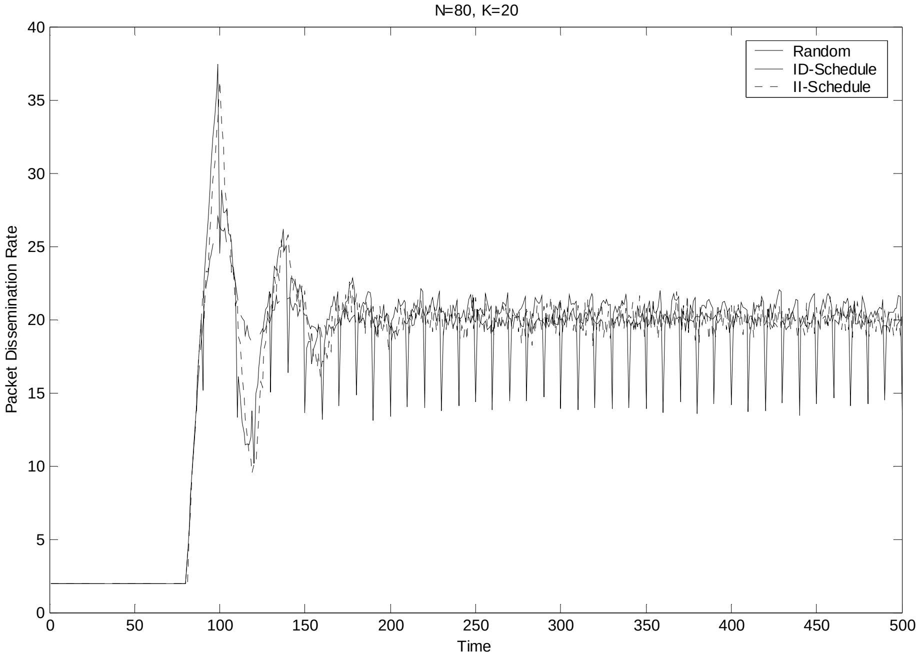


Figure 4-2 Average packet dissemination rate at each unit of time with $N/K = 4$.

from equation (4.3)

$$\begin{aligned} P_{Succ} &= P^{2N+1}(K) \\ &= P^{2N}(K) + P^{2N}(K-1)P^{1}(2)P_e(2,K-1) \\ &= P^{2N}(K) + (K-1)P^{2N}(K-1). \end{aligned} \tag{4.8}$$

4.6 Simulation Results

Figure 4-2 and Figure 4-3 plot the evolution of the system. Average packet dissemination rate vs time is plotted for different N and K. From the figure we can see there are three phases. At first nodes entered the system and download two packets from the access points. There is no node encounter during this phase. The packet dissemination rate is exactly two. Then after the first node reaches the middle of the system, it encounters a node. Packet exchanging begins. After that more and more pairs of node encounter, the packet dissemination rate keeps increasing until it is near the upper threshold $2N+2$. Because there are at most N encounters in the system, at

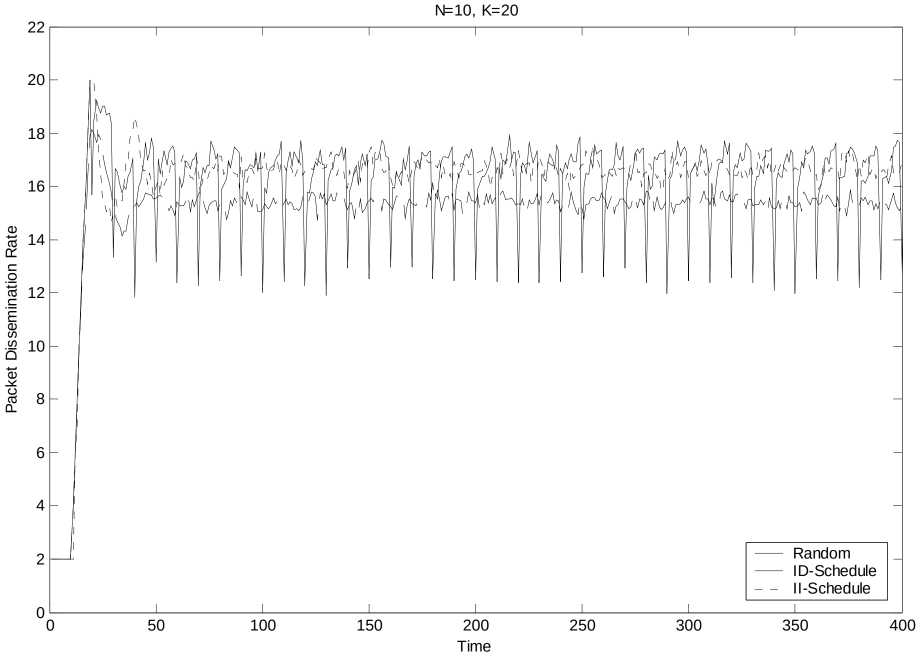
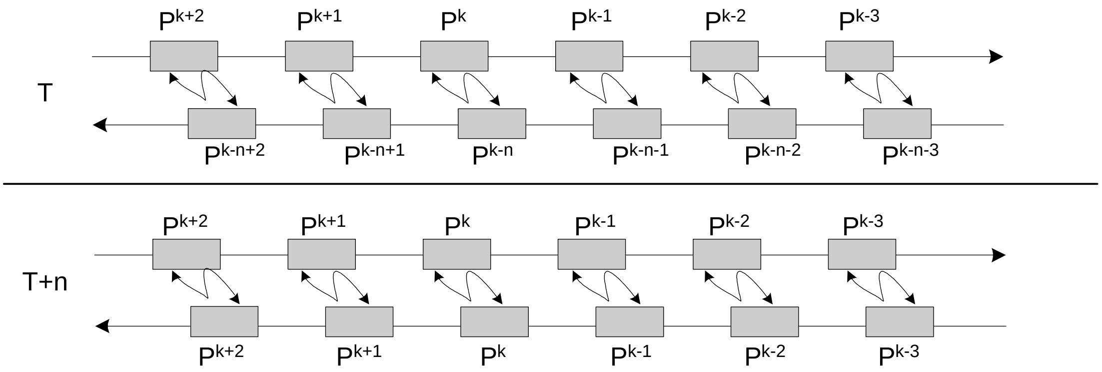


Figure 4-3 Average packet dissemination rate at each unit of time with $N/K = 0.5$.

most $2N$ packets can be exchanged. The packet dissemination rate oscillates for some time before it reaches a stable state. The larger the N, the slower the system becomes stable.

It is interesting to note that the packet dissemination rate of the ID-Schedule strategy has a negative pulse periodically. Since packet downloading follows either the increasing or the decreasing direction, two packets can be grouped as a pair. Given the total packet number K, there are $K/2$ pairs of packets. Suppose nodes move from access point A to B and download their initial packets from access point A following the increasing order. Nodes moving from access point B to A will download their initial packets from access point B following the decreasing order. This is shown in Figure 4-4. At time T, the node pairs encountered have different initial two packets downloaded form the access points. If the node pair haven't acquired the initial two packets of the other, they will definitely exchange packets. If not, the exchange probability will depend on the remaining packets they cached. However, due to the



P^k: The initial pair of files that the node downloads from the access point.
K means it is the Kth pair.

Figure 4-4 Explanation of the periodicity of ID-Schedule.

scheduling, the possibility that the two nodes have acquired the initial two packets of the other is low. Thus it has a high probability of packet exchange. At time *T+n*, the encountered node pairs have same initial two packets. In such condition, the packet exchange probability only depends on the packets except the initial two packets, which they cached. Obviously, this probability is lower than the previous one. Over all *N* node pairs, the packet dissemination rate is significantly lower when those encountered nodes have the same initial packets compared to those have different initial packets. Since there are *K/2* pairs of packets, the encountered node pairs with initial packets will happen once every *K/2* time units. Thus the period of the waveform of the evolution of the packet dissemination rate is *K/2* for the ID-Scheduling strategy.

This conclusion is consistent with the observation from Figure 4-2 and Figure 4-3. For the II-Scheduling strategy, one and only one encountered node pair will have same initial packets downloaded from an access point at any time. The packet dissemination rate will not have any periodicity when the system evolves into its stable state. From the waveform we can easily verify this.

Figure 4-5 plots the average packet dissemination rate against the ratio of *N*/*K* for

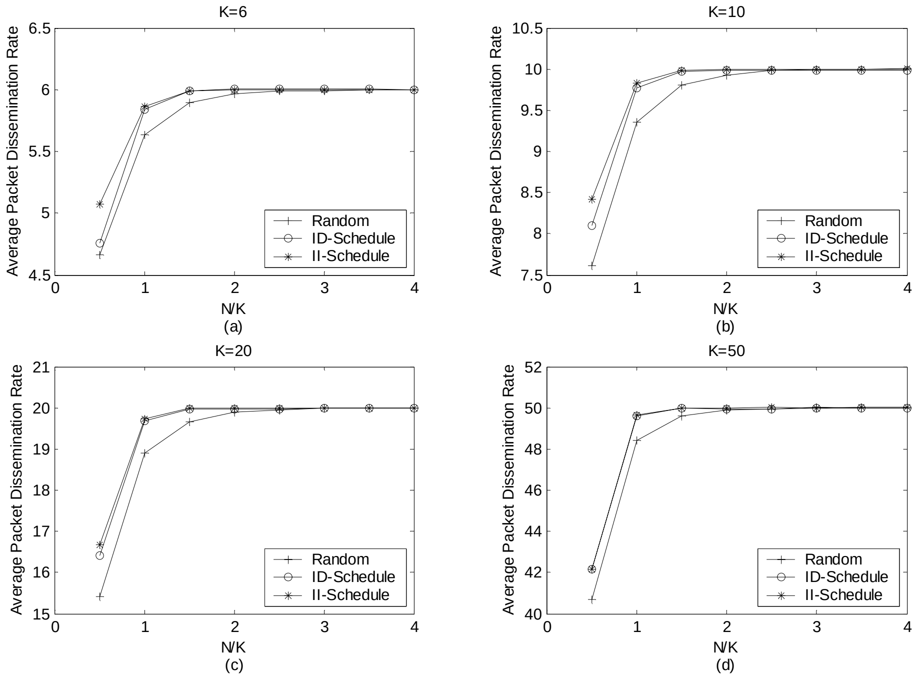


Figure 4-5 Average packet dissemination rate vs the ratio *N*/*K* for various *K*.

different number of packets (*K* = 6, 10, 20, 50). In all cases, the scheduling strategy has significant improvement over the random strategy, especially when *N/K* is small. With the increase of *N/K*, the difference decreases. When *N/K* is larger than 2, the difference between the three strategies is very small. All of the strategies reach the capacity of the system, which is determined by the number of packets. Because during each unit of time, only one node enters the system and one node leaves the system. If a node successfully acquires all the packets before it leaves, there will be *K* packets transferred during each unit of time in the system given every packet transmission isn’t wasted (no packet was transferred to a node twice or more). We also note that both ID-Schedule and II-Schedule strategies have similar performance. The difference decreases with the increase of *K* and *N/K*.

In Figure 4-6, the average packet exchange probability is plotted against *N*/*K* for different *K*. We notice that the II-Schedule strategy has best performance and the random strategy performs worst. However, the difference between the three strategies

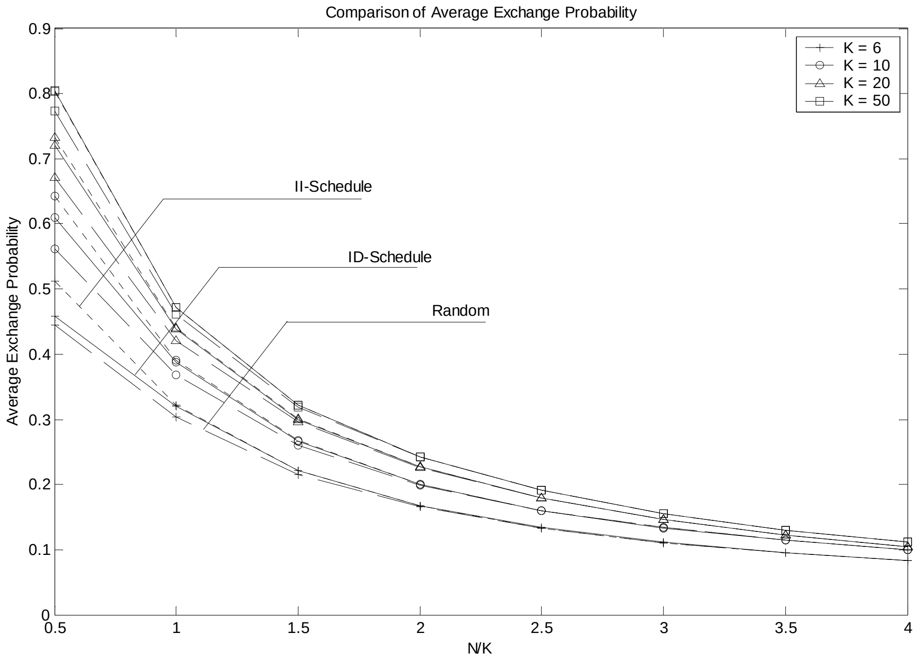
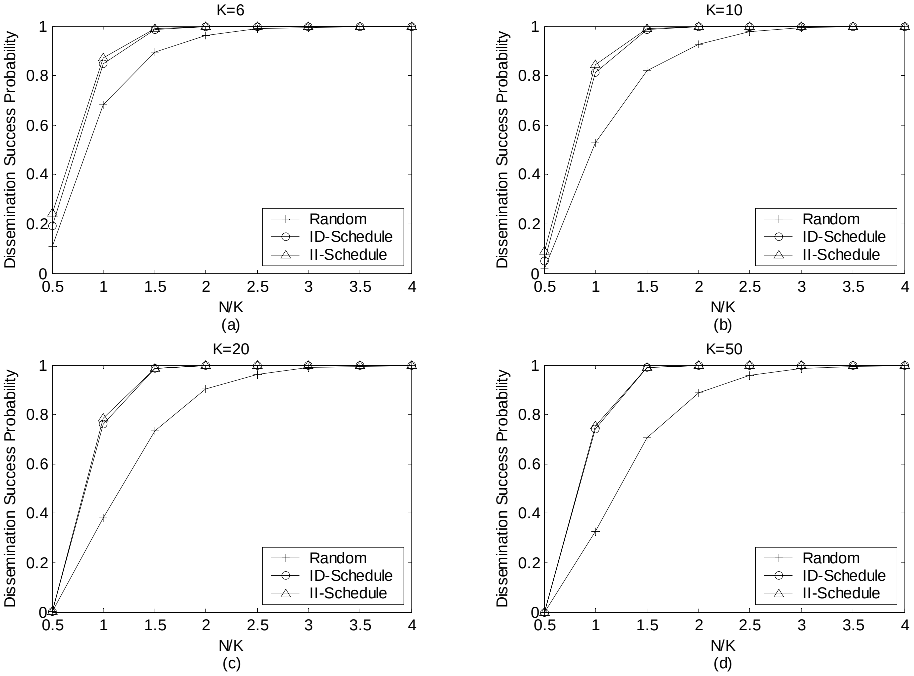


Figure 4-6 Average packet exchange probability vs the ratio of *N*/*K* for various *K*.

decreases as *N/K* increase. As to any strategy, the packet exchange probability increases as the number of packets. This can be easily verified from (4.1). The larger the number of packets *K*, the greater the exchange probability P_e for a determined pair of cached packets number (f_i, f_j).

Figure 4-7 compares the probability of a node obtaining all packets when it leaves the system between the three strategies for various *K*. It is shown that two proposed schedule strategies have significant improvement on the dissemination success probability than the random strategy. It is also shown, whatever the *K* is, the probability will be close to 1 when *N/K* is over 2 for all schedule strategies. The curve for the proposed scheduling strategies is steeper than the curve for the random strategy. At first, all of the three strategies have similar performance because there are no enough nodes in the system, even if the proposed schedule strategies have higher packet exchange probability as shown in Figure 4-6. Most of the nodes still fail to acquire all the packets when they leave the system. As *N/K* increases, a node has more

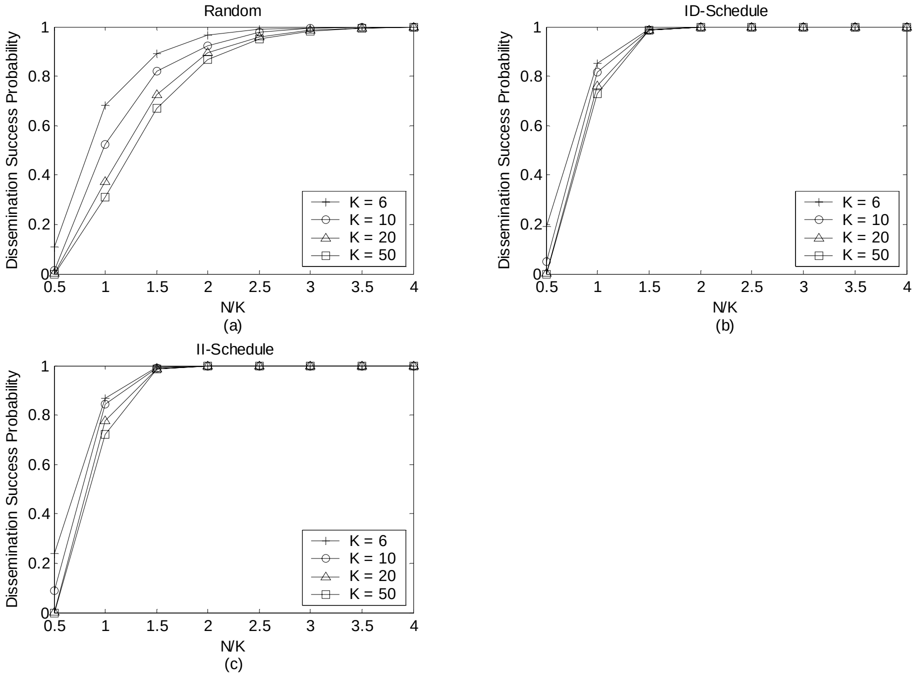


(a) (b) (c) (d)

Figure 4-7 Dissemination success probability vs *N*/*K* for various *K*.

chance to encounter another node when it moves from one end to the other, thus the probability it acquires all packets increases. Since the proposed scheduled strategies successfully coordinate the packets each node has in order to increase the packet exchange probability when two nodes encounter, the success probability of the proposed scheduled strategies outperform the random strategy by over 100% in Figure 4-7 (d) when *N/K* is 1. Even when *N/K* is 1.5 where the proposed schedule strategies nearly provide 100% packet dissemination success probability, the improvement is still near 30%. However, when *N/K* is large enough, all three strategies can guarantee the nodes acquiring all packets, the corresponding curves of theirs monotonically increase to 1 at last.

This experiment demonstrates the efficiency of the proposed scheduled downloading at the access points in this packet dissemination environment. Furthermore, we also conclude that *N/K* is a key parameter affecting the system performance. For a given packet number *K*, the number of nodes *N* should be at least



(a)

(b)

(c)

Figure 4-8 Difference between dissemination success probabilities.

two times of *K* for scheduled strategies and 3 times for random strategy to guarantee packet successful dissemination.

Finally, Figure 4-8 shows the success probability difference for different packet numbers *K*. In all three figures, smaller *K* has higher success probability. However the difference is quite small for the proposed two scheduled strategies. It is quite natural that the chance for a node to successfully get more packets is lower than to get fewer packets in a finite time duration.

4.7 Conclusion

In this chapter, we have examined the issue of data dissemination among mobile devices. Our results show that system capacity and performance can be improved significantly by properly scheduling the packet download at the access points, because scheduling can provide some degree of order in the packet distribution among the mobile devices, which increases the packet exchange probability. The advantage of the proposed schedule strategies is that they are simple and effective. They do not require any additional information about the system. No additional communications will be induced into the system to facilitate the schedule. Another merit of these schedule strategies is that they are implemented in the access points and no additional requirement is imposed on the mobile nodes.

However we would like to emphasize that this result is obtained under several idealistic assumptions. In particular, we assume the identical node mobility model in the network and the size of the network is restricted within one access point section. As such, the result should be viewed as a theoretical one. In any case, it does suggest that for data dissemination among mobile devices, there is ample opportunity to increase the capacity by exploiting mobility and dissemination strategy. The result of this chapter can be considered as an extreme point in the scheduling, with strict constraint on the mobility model. With a random mobility model and/or a complex system model, the capacity improvement with scheduling may decrease.

CHAPTER FIVE

Conclusions

This chapter highlights the contributions we have made in this thesis and presents the conclusions drawn from the evaluations conducted in previous chapters. Possible future work is also addressed here.

5.1 Summary of Thesis

Data traffic will comprise a major part of the future wireless communication system. There are many challenges that have to be addressed to realize the full potential of wireless computing. These include networking issues (e.g. mobile IP, mobility management, congestion control in TCP), power management issues (e.g. operation systems level, I/O and file systems level), system issues (e.g. disconnection support, locations dependent queries, recovery) and information dissemination issues. We focus only on the power management issues and the information dissemination issues in this thesis. In this thesis, we have developed a power control algorithm for data traffic in a wireless system and proposed two scheduling strategies to disseminate popular data to mobile users.

In Chapter 2, we have introduced some background of mobile computing. The reference architecture of wireless computing environment and the characteristics of wireless computing environments are discussed. The QoS concept is also introduced.

In Chapter 3, we have briefed the power control schemes used for current cellular mobile systems. A famous power control algorithm used for voice-oriented traffic is studied and the motivation of designing power control algorithms for data traffic is examined by analysis and a simple example. We have then developed a power & backlog sensitive power control algorithm for data traffic in a CDMA wireless communication system. Our solution reallocates transmit power once every time unit. The primary objective of the reallocation at time is to tradeoff the desire for increasing transmit power against the associated cost, achieving throughput improvement and power saving. The performance of downlink channels using this method has been evaluated by computer simulations. The results show that, our scheme improves the system performance in terms of both system throughput and power consumption over the existing algorithms. It is also shown that, although our proposed scheme is non-cooperative in nature, some degree of cooperation emerges: a user will automatically decrease its transmit power (and may even turn of transmission) when the transmission environment is not desirable.

In Chapter 4, the concept of data dissemination in wireless system is first introduced followed by the performance metrics and design issues for dissemination. We then propose a new class of scheduling strategies for the data dissemination in mobile networks. A finite linear network with mobility limited mobile nodes is considered. We show by analysis and simulations that the network performance depends on both the number of nodes in the network and the number of packets to be disseminated. Our results show that system capacity and performance can be improved significantly by properly scheduling the packet download at the access points. This is because scheduling successfully coordinates the packet distribution among mobile devices, which increases the packet exchange probability.

The advantages of the proposed scheduling strategies are the simplicity and effectiveness. They do not require any additional information about the system. No additional communications will be induced into the system to facilitate the scheduling. Another merit of the scheduling strategies is that they are implemented in the access points and no additional requirement is imposed on the mobile nodes. The packet exchange among mobile nodes is still kept in an essentially noncooperative manner since no node forwards packets for others. Furthermore, we point out that such scheduling is very simple, distributed and induces no additional communication overhead to the mobile networks.

5.2 Future Work

As we have mentioned in earlier chapters, previous work has been focused on a system in which all nodes join the system follows i.i.d distribution. The nodes in the system are distributed evenly and they move with common constant speed. We showed the effectiveness of our proposed scheduling strategies. However, with a random mobility model, the capacity improvement with such scheduling may decrease. It would be interesting to investigate the performance with a more realistic mobility model which captures the mobility pattern of mobile nodes with reasonable accuracy. Some recent work uses either a random walk model [44] or a group clustering model [45]. The former model is not able to characterize the regularity of user mobility, commonly found in a civilian environment. On the other hand, the latter model makes a too strong assumption about the uniformity of mobility among mobile nodes. Recently a Street-and-Building model for mobility modeling has been proposed in [40]. This model is designed especially for civilian environments. However it introduces a population density factor which is based on the assumption that increased population density will decrease the walking speed of people. We argue that this will hold only when the population density is high enough which rarely happens. To overcome these problems, we are thinking of proposing a mobility model in the future based on the fact that all walking users on a given street have similar velocity, with certain adjustable variation. Walking users may change their velocity from time to time with very low probability. Furthermore, there is a restriction on the maximum moving speed of mobile nodes.

We have limited the packet transfer between mobile node to packet exchange and the packets are randomly picked. However these random picked packets might lead to some disorder in the packet content that each node has. Obviously this result is inconsistent with our goal of scheduling the packet download in the access points, to speed-up the data dissemination by creating some order in the packet contents of those mobile nodes. Thus in the future, we will investigate how to coordinate the packet exchange between mobile nodes. A simple approach will be dividing the packet exchange into two processes. At first, one mobile node should try to pick up the packet for exchange from the two packets initially downloaded from the access point. If this packet can not be exchanged, then the packet for exchange is randomly chosen from the packet contents that the node has. Another approach will be to loose the packet exchange restriction between two mobile nodes. If two nodes can not exchange packets, then one node can download packets from another. However, this will cause another problem that one node may continuously provide data for others to download. Its battery will soon be exhausted. One way to solve this problem is to let a node be active if it wants to download or exchange packets. If it has got all the packets, then it will be inactive which means it will not share its data anymore. The system throughput, the probability of nodes getting all packets, and the time delay should be studied further.

In previous work we have only considered the two access point model. In future, we may extend it to a more realistic and complicated model which contains several access points in a one dimensional or two dimensional space. The nodes move around through some determined routes just like in the streets. The system throughput, the probability of nodes getting all packets, and the average time for a node to get all packets are of interest. How to coordinate the packet downloading to increase the system performance should be investigated as well.

Bibliography

[1] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. Int. Conf. Communications*, vol. 1, pp. 331–335, Seattle, WA, June 1995.

[2] D. Tse, "Optimal power allocation over parallel Gaussian channels," in *Proc. Int. Symp. Information Theory*, Ulm, Germany, pp. 27, 1997.

[3] D. Barbara and T. Imielinski, "Sleepers and workaholics: Caching in mobile distributed environments," In *Proceedings of the 1994 ACM-SIGMOD International Conference on Management of Data*, pp. 1-12, June 1994.

[4] T. Imielinski and B.R. Badrinath, "Mobile wireless computing: Challenges in data management," In *Communications of the ACM*, 37(10), October 1994.

[5] D. Duchamp and N.F. Reynolds, "Measured performance of a wireless LAN," *In Proc. 17th Conf. on Local Computer Networks*, pp. 494-499, Sep. 1992.

[6] T. H. Meng et al., "Portable video-on-demand in wireless communication," In *Proceedings of the IEEE*, Vol. 83, No. 4, pp 659-680, April, 1995.

[7] L.C. Yun and D. G. Messerschmitt, "Power control for variable QoS on a CDMA channel," In *Proceedings of the IEEE MILCOM*, pp. 178-182, October, 1994.

[8] M. Khasari et al., "Low bit-error video transmission over fading channels for wireless microcellular systems," In *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 1, pp. 1-11, February, 1996.

[9] R. Yuan et al., "Mobility support in a wireless ATM network," In *Proc. WINLAB Workshop on 3rd Generation Wireless Systems*, pp. 335-345, 1995.

[10] Clint Smith and Daniel Collins, 3G Wireless Networks, McGraw-Hill, 2001.

[11] P. Haskell, "Some research issues in a heterogeneous terminal and transport environment for multimedia services," In *Proc. Cost #229 Workshop on Adaptive Systems, Intelligent approaches, Massively Parallel Computing and Emergent Technologies in Signal Processing and Communications*, October, 1994.

[12] A.S. Acampora and M. Naghshineh, "Control and quality-of-service provisioning in high-speed microcellular networks," In *IEEE Personal Communication Magazine*, Vol. 1, No. 2, Second Qtr. 1994.

[13] H. Koraitim et al., "Quality of service aspects of transport technologies for the UMTS radio access network," In *Personal Wireless Communications*, Kluwer Academic Publishers, 2000.

[14] J S. Choi and K. Shin, "An uplink CDMA system architecture with diverse QoS

guarantees for heterogeneous traffic," In *IEEE/ACM Trans. Net.*, Vol 7, pp. 616-628, Oct, 1999.

[15] K.S. Gilhousen et al., "On the capacity of a cellular CDMA system," In *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, pp. 303-312, May 1991.

[16] W. C. Y. Lee, "Power control in CDMA," In *Proceedings of IEEE Vehicular Technology Conference (VTC)*, St. Louis, MO, pp. 77-80, May 1991.

[17] Z. Rosberg and J. Zander, "Toward a framework for power control in cellular systems," In *ACM/Baltzer Wireless Networks Journal*, Vol. 4, pp. 215-222, 1998.

[18] N. Bambos, "Toward power-sensitive network architectures in wireless communications: Concept, issues, and design aspects," In *IEEE Personal Communications*, pp. 50-59, June 1998.

[19] R. Yates, "A framework for uplink power control in cellular radio systems," In *IEEE JSAC*, vol. 13, no.7, pp. 1341–1347, 1995.

[20] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," In *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 641-646, Nov. 1993.

[21] S. C. Chen, N. Bambos, and G. J. Pottie, "Admission control schemes for wireless communication networks with adjustable transmitter power," In *Proceedings of IEEE INFOCOM*, Toronto, Canada, pp. 21-28, June 1994.

[22] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio system," In *Proceedings 4th WINLAB Workshop Third Generation Wireless Information Network*, Rutgers University, pp. 249-257, Oct. 1993.

[23] N. Bambos, S. C. Chen, and G. J. Pottie, "Radio link admission algorithm for wireless networks with power control and active link quality protection," In *Proceedings of IEEE INFOCOM*, Boston, MA, pp. 97-104, April 1995.

[24] R. Gejji, "Forward-link-power control in CDMA cellular system," In *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 532-536, Nov. 1992.

[25] S. Ariyavisitakul, "SIR-based power control in a CDMA system," In GLOBECOM'92, pp. 868-873, 1992.

[26] N. Bambos and S. Kandukuri, "Power controlled multiple access (PCMA) in wireless communication networks," In *IEEE INFOCOM*, pp. 386-395, 2000.

[27] Q. Wu, "Performance of optimum transmitter power control in CDMA cellular mobile systems," In *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 571-575, March 1999.

[28] M. Xiao, N. B. Shroff and E. K. P. Chong, "Utility-Based power control in cellular wireless systems," In *IEEE INFOCOM*, pp. 412-421, 2001.

[29] T. S. Rappaport, Wireless Communications Principles & Practice, Prentice-Hall,

Inc. 1996.

[30] W. C.Y. Lee, Mobile Communication Design Fundamentals, 2nd edition, John Wiley & Sons, New York, 1993.

[31] C.Y. Huang, Radio Resource Management in Power Controlled CDMA Systems, Ph.D. thesis, Rutgers University, Piscataway, NJ, 1996.

[32] D. Kim, "Rate-regulated power control for supporting flexible transmission in future CDMA mobile networks," In *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 968-977, May 1999.

[33] S. Ariyavisitakul, "Signal and interference statistics of a CDMA system with feedback power control – Part II," In *IEEE Transactions on Communications*, vol. 42, pp. 597-605, Feb. 1994.

[34] K. Pahlavan and A.H. Levesque. Wireless Information Networks, John Wiley & Sons, 1995.

[35] J. Rulnick and N. Bambos, "Mobile power management for wireless communication networks," In *Wireless Networks*, Vol.3, pp. 3-14, 1997.

[36] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik, "Broadcast disks: Data management for asymmetric communications environments," In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 199-210, San Jose, CA, USA, May 1995.

[37] T. Imielinski and S. Viswanathan, "Adaptive wireless information systems," In *Proceedings of the Special Interest Group on DataBase Systems (SIGDBS) Conference*, pages 19-41, Tokyo, Japan, October 1994.

[38] W. C. Lee, Q. L. Hu and D. L. Lee, "A study of channel allocation methods for data dissemination in mobile computing environments," In *ACM/Baltzer Journal of Mobile Networking and Applications (MONET)*, 4(2), pp. 117-129, 1999.

[39] E. Pitoura and G. Samaras. Data Management for Mobile Computing. Kluwer Academic Publisher, 1998.

[40] S.K. Goel, M. Singh, D. Y. Xu and B. C. Li, "Efficient peer-to-peer data dissemination in mobile ad-hoc networks," In *Parallel Processing Workshops, 2002. Proceedings. International Conference*, pp. 152–158, Aug. 2002.

[41] M. Papadopouli and H. Schulzrinne, "Effects of power conservation, wireless coverage and cooperation on data dissemination among mobile devices," In *Proc. IEEE mobiHoc '01,* 2001.

[42] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," In *IEEE/ACM Transactions on Networking*, Vol. 10, Issue 4, pp. 477-486, Aug. 2002.

[43] W. H. Yuen, R.D. Yates, and S. C. Mau, "Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks," In *INFOCOM 2003.*

Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE, Vol. 3 , pp. 2218 -2228, April 2003.

[44] M. Papadopouli and H. Schulzrinne, “Design and implementation of a peer-to-peer data dissemination and prefetching tool for mobile users,” In *Proceedings of the 2nd ACM International Workshop on Modeling and Simulation of Wireless and Mobile Systems*, 1999.

[45] K. Wang and B. Li, “Efficient and guaranteed service coverage in partitionable mobile ad-hoc networks,” In *Proceedings of IEEE INFOCOM 2002*, June 2002.

Appendix: Author's Publication

[1] Qi Wang, Francois Chin and Zhongding Lei, "Power & backlog sensitive power control for wireless data in CDMA mobile systems under multipath environments", In *IEEE Vehicular Technology Conference 2004-Fall*, LA, USA, 2004.

[2] Qi Wang, Zhongding Lei and Francois Chin, "Scheduled data dissemination among mobile devices", In *IEEE Vehicular Technology Conference 2004-Fall*, LA, USA, 2004.

[3] Qi Wang, Francois Chin and Zhongding Lei, "A new power control algorithm in CDMA systems", *IEEE International Symposium on Spread Spectrum Techniques and Applications*, Sydney, 2004.