Soft Handoff in MC-CDMA Cellular Networks Supporting Multimedia Services

by

Jinfang Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada 2004

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

An adaptive resource reservation and handoff priority scheme, which jointly considers the characteristics from the physical, link and network layers, is proposed for a packet switching Multicode (MC)-CDMA cellular network supporting multimedia applications. A call admission region is derived for call admission control (CAC) and handoff management with the satisfaction of quality of service (QoS) requirements for all kinds of multimedia traffic, where the QoS parameters include the wireless transmission bit error rate (BER), the packet loss rate (PLR) and delay requirement. The BER requirement is guaranteed by properly arranging simultaneous packet transmissions, whereas the PLR and delay requirements are guaranteed by the proposed packet scheduling and partial packet integration scheme. To give service priority to handoff calls, a threshold-based adaptive resource reservation scheme is proposed on the basis of a practical user mobility model and a proper handoff request prediction scheme. The resource reservation scheme gives handoff calls a higher admission priority over new calls, and is designed to adjust the reservation-request time threshold adaptively according to the varying traffic load. The individual reservation requests form a common reservation pool, and handoff calls are served on a first-come-first-serve basis. By exploiting the transmission rate adaptability of video calls to the available radio resources, the resources freed from rate-adaptive high-quality video calls by service degradation can be further used to prioritize handoff calls. With the proposed resource reservation and handoff priority scheme, the dynamic properties of the system can be closely captured and a better grade of service (GoS) in terms of new call blocking and handoff call dropping probabilities(rates) can be achieved compared to other schemes in literature. Numerical results are presented to show the improvement of the GoS performance and the efficient utilization of the radio resources.

Acknowledgements

I wish to express my sincere gratitude to my supervisors, Professor Jon W. Mark and Professor Xuemin(Sherman) Shen for their guidance towards this thesis. Their constant encouragement, professional assistance made this thesis possible. I would especially like to thank them for challenging me to improve my writing and critical reviewing skills, which made my publications possible.

I would like to thank Dr. John Thistle, Dr. En-hui Yang, Dr. Raouf Boutaba for acting as my thesis committee members. I sincerely thank them for their valuable advice and helpful suggestions for my research work and for completing this thesis. I would also like to thank Dr. Ekram Hossain for serving as my external committee member and for providing valuable and helpful comments.

It has been my pleasure to work at the Centre for Wireless Communications (CWC) group, the place has provided me great opportunities to perform my research and to learn extensively from other team members. I sincerely acknowledge the support and assistance offered by all the members of the CWC. My Ph.D program was full of excitement and truly memorable thanks to the great place and the wonderful people.

I would like to acknowledge the funding agencies from Ontario Graduate Scholarship in Science and Technology (OGSST), Canadian Institute for Telecommunications Research (CITR) and Ontario Research & Development Challenge Fund Bell Scholarship (ORDCF

- Bell Scholarship) for supporting me to complete my Ph.D program. I would also like to thank University of Waterloo for its numerous scholarships.

My special gratitude is to my husband, Lei, who supported my graduate studies with deepest love. He took good care of me and tried to ease the pressure of the work. As a token of appreciation, I dedicate my thesis to him.

Finally, my deepest thanks go to my parents and sister in China, YiGen Zhang, Guifen Xu and Lifang Zhang, who always support my studies. Without their encouragement and love, little of this work would have been possible.

Contents

1 Introduction **1**

1.1 Wireless communications system 1

1.2 Handoff in mobile systems 5

1.2.1 Handoff categories 6

1.2.2 Challenging issues 8

1.3 Related research work on handoff 11

1.3.1 Cell capacity and call admission region derivation 11

1.3.2 Resource reservation 13

1.3.3 Transmission rate adaptation 15

1.3.4 Summary 17

1.4 Research motivation and objectives 17

1.5 Contributions of the thesis 20

1.6 Overview of the thesis 21

2 Proposed Research **23**

2.1 Proposed research 23

2.2 System structure 27

2.3 MAC layer structure 30

2.4 Propagation model 32

2.5 Traffic model 35

2.6 GoS parameters 37

2.7 Summary 38

3 Uplink Call Admission Region **39**

3.1 Bit error rate and packet loss rate 40

3.2 Uplink cell capacity estimation 41

3.2.1 Call admission region 49

3.3 Summary 54

4 An Adaptive Resource Reservation Strategy for Handoff **55**

4.1 Soft handoff signaling 56

4.2 Adaptive resource reservation scheme 59

4.2.1 User mobility adaptation 59

4.2.2 Traffic load adaptation 64

4.3 Summary 69

5 An Adaptive Handoff Priority Scheme Supporting Multimedia Applications 71

5.1 Adaptive handoff priority scheme 73

5.2 Summary 78

6 Numerical Results 79

6.1 Simulation model 79

6.2 Mobility model 82

6.3 Performance evaluation 86

6.4 Summary 108

7 Conclusions and Future Work 110

7.1 Summary of contributions 110

7.2 Future work 112

List of Tables

1.1 WCDMA vs. cdma2000 4

6.1 Simulation parameters 81

6.2 Call admission region with circuit switching transmission ($P^*_{WTE} = 10^{-4}$) . 86

6.3 Call admission region with packet switching transmission ($P^*_{WTE} = 10^{-4}$) . 87

6.4 Call admission region with packet switching transmission ($P^*_{WTE} = 10^{-5}$) . 88

6.5 GoS performance (new voice call arrival rate = 0.35 calls/s) 91

6.6 Mean and variation of R_b/R_d 96

6.7 GoS performance (the ratio of the number of pedestrian to vehicle users = 0.7, $P^*_{WTE} = 10^{-4}$) 104

List of Figures

2.1 System structure 28

2.2 Uplink and downlink frame structure 30

2.3 Two-state ON-OFF model 37

3.1 Cell structure 42

3.2 partial packet integration 53

4.1 Soft handoff signaling 56

4.2 RPSS difference vs. T_{rev} 62

4.3 New call blocking and handoff call dropping rates derivation 66

4.4 Reservation-request time threshold update curve 69

5.1 Voice handoff call admission control flow diagram 74

5.2 Video handoff call admission control flow diagram 76

5.3 Low-quality video call transmission rate adjustment 77

6.1 Simulation flowchat 82
6.2 Acceleration model 84
6.3 GoS performance vs. reservation-request time threshold with the ratio of the number of pedestrian to vehicle users being 0.8, 0.7 and 0.6 92
6.4 capacity utilization vs. reservation-request time threshold with the ratio of the number of pedestrian to vehicle users being 0.8, 0.7 and 0.6 93
6.5 GoS performance vs. simulation time with the ratio of the number of pedestrian to vehicle users being 0.7 and $c = 1.5$ 94
6.6 Reservation-request time threshold vs. simulation time with the ratio of the number of pedestrian to vehicle users being 0.7 and $c = 1.5$ 95
6.7 New call blocking probabilities with the packet switching and circuit switching transmissions 97
6.8 Handoff call dropping probabilities with the packet switching and circuit switching transmissions 98
6.9 Resource utilization with the packet switching and circuit switching transmissions 99
6.10 Ratios of low quality video calls over total video calls with the packet switching and circuit switching transmissions 100

6.11 New call blocking probability with the adaptive priority and non-adaptive handoff schemes 101

6.12 Handoff call dropping probability with the adaptive priority and non-adaptive handoff schemes 102

6.13 New call blocking probability with P^*_{WTE} being 10^{-4} and 10^{-5} 103

6.14 Handoff call dropping probability with P^*_{WTE} being 10^{-4} and 10^{-5} 104

6.15 Ratio of low quality video calls over total video calls with P^*_{WTE} being 10^{-4} and 10^{-5} 105

6.16 GoS performance vs. reservation-request time threshold with the new voice call arrival rate being 0.12, 0.125 and 0.13 calls/s 106

6.17 GoS performance vs. simulation time with the ratio of the number of pedestrian to vehicle users being 0.7 and $c = 2$ 107

List of Abbreviations

3G	third generation
3GPP	the third generation partnership project
3GPP2	the third generation partnership project 2
AMPS	advanced mobile phone system
AR	autoregressive
BER	bit error rate
BS	base station
CAC	call admission control
CDMA	code division multiple access
DS-CDMA	direct sequence-code division multiple access
ETACS	European total access cellular system
FCFS	first-come-first-serve

FDD	frequency division duplex
GoS	grade of service
GPRS	general packet radio services
GSM	global system for mobile
IMT	international mobile telecommunications
MAC	multiple access control
MAHO	mobile assisted handoff
MAI	multiple access interference
MC-CDMA	multicode-code division multiple access
MCHO	mobile controlled handoff
MS	mobile station
MSC	mobile switching center
MUP	most urgent packet
NCHO	network controlled handoff
PLR	packet loss rate
QoS	quality of service
RPSS	received pilot signal strength
SDU	selection/distribution unit

SIR	signal to interference ratio
UMTS	universal mobile telecommunications system
WCDMA	wideband code division multiple access

Glossary of Symbols

$a_{S,k}$	variation of the changes in RPSS
C	uplink cell capacity
d_0	the reference distance
d_i	the distance to the i-th BS
$\bar{d}_i$	shadowing fading correlation distance
E_b/I_0	bit energy-to-interference ratio
$G_{k,i}(r,\theta)$	channel gain of an MS located at (r,θ) at time t_k from the i-th BS
L	RPSS samples averaging window size
M	accumulated MUPs in packets/frame
N	averaging window size within which the number of handoff call arrivals
P_b	new call blocking probability
P_d	handoff call dropping probability

P^i_{BER}	bit error rate due to wireless transmission
P^i_e	overall bit error rate
$P_{k,i}$	the received RPSS
P^i_{PLR}	bit error rate due to packet loss
$p_{k,i}$	received pilot signal strength from the i-th BS at time t_k
P_{succ}	probability of a successful resource reservation
R_b	new call blocking rate
R_C	cell radius
R_d	handoff call dropping rate
R_I	inner cell radius
R_O	outer cell radius
$r_{loss_i\|m}$	the number of lost packets for user i given the MUP load in a frame is m
$\bar{r}_i$	average data generation rate for user i
S	required received power with perfect power control
S_e	imperfect received power
$S_{k,i}$	smoothed $p_{k,i}$
S_t	ADD threshold
T_{rev}	reservation-request time threshold

$v_{S,k}$	the changes in RPSS
α	relative handoff area
$\beta_{k,i}$	weighting factor to the previously estimated RPSS
γ	handoff call dropping probability weighting factor
η	path loss exponent
κ_I	inner cell user density
κ_S	soft handoff area user density
θ	transition probability
ρ	smoothing factor
$\varrho_{k,i}$	correlation coefficient of $\{\zeta^*_{k,i}\}$
σ	variance of the shadowing effect
σ_e	variance of the imperfect power control
ξ_e	Gaussian random variable
$\zeta^*_{k,i}$	lognormal shadowing effect
Δt	sampling interval
$\mathbf{r}_k$	correlated acceleration
$\mathbf{u}_k$	unexpected acceleration
$\mathbf{w}_k$	Gaussian vector

Chapter 1

Introduction

1.1 Wireless communications system

Wireless communications have experienced enormous growth over the last decade and the communication world is becoming more mobile than ever. Mobile cellular systems have evolved from the first generation analog system, such as *AMPS (Advanced Mobile Phone System)* and *ETACS (European Total Access Cellular System)*, to the second generation digital systems, such as *GSM (Global System for Mobile)* and *IS-95 cdmaOne.* The parallel, rapid growth in mobile and Internet-based services is the driving force for the development of the third generation mobile system (*3G*). The 3G systems can provide traditional voice services as an extension of the wireline telephone systems, and support mobile multimedia and mobile Internet-based services of the future. The data transmission rate is at least 144 kbps in vehicle speed, 384 kbps in pedestrian speed and up to 2 Mbps for indoor and picocell environment. 3G systems allow both circuit switching and packet switching wireless transmissions, which make it possible to provide advanced and flexible quality of service (QoS) support and efficient radio resource utilization. A goal of 3G systems is to provide universal coverage and to enable users to be capable of seamless roaming between multiple networks so as to provide mobile users with services anywhere anytime. The activities of the 3G standardization and development have focused on the radio interface proposals for IMT-2000/UMTS [1], including cdma2000 (code division multiple access) from North America [1], WCDMA (wideband CDMA) from Europe [2] and Japan [5], and Global CDMA from Korea [3][4]. Among these terrestrial radio interface proposals, WCDMA and cdma2000 have drawn much more attention than the others and likely emerged as the two dominant 3G radio transmission technologies.

WCDMA The WCDMA specification [29] was created in 3GPP (the Third Generation Partnership Project), which is the joint standardization project of the standardization bodies from Europe, Japan, Korea, the USA and China. Within 3GPP, the 3G system is called UMTS, which employs the WCDMA radio interface and is backward compatible with GSM/GPRS (Global System for Mobile/General Packet Radio

[1]International Mobile Telecommunications-2000/Universal Mobile Telecommunications System

Service) core networks.

cdma2000 cdma2000 is a multi-carrier CDMA radio transmission technique standardized by 3GPP2 (the Third Generation Partnership Project 2) [34][35][36][37]. cdma2000 is designed to be backward compatible with IS-95-A/B.

The main system characteristics and parameters of WCDMA and cdma2000 are listed and compared in Table. 1.1. In comparison, WCDMA and cdma2000 have some similarities, but differ in the chip rate and channel structure. In the best circumstances, some harmonization will occur between cdma2000 and WCDMA, facilitating deployment of hardware capable of supporting both systems.

A typical wireless cellular communication network consists of three components: mobile station, base station and mobile switching center with connections to link these components.

- *Mobile Station (MS).* An MS can be as simple as a cellular phone or as complex as a multimedia terminal with a wireless access card.

- *Base Station (BS).* The service area of a wireless network is divided into a number of cells. A BS ensures the radio coverage of a cell area and provides an entry point to the wireline networks.

- *Mobile Switching Center (MSC).* An MSC is responsible for switching functions nec-

Table 1.1: WCDMA vs. cdma2000

Characteristics	WCDMA	cdma2000
Channel bandwidth	5 MHz	1.25, 5 MHz
Chip rate	3.84 Mcps	3.68 Mcps
Forward link RF channel structure	Direct spreading	Multi-carrier or direct spreading
Synchronous	Asynchronous	Synchronous
Multirate	Variable spreading and multicode	Variable spreading and multicode
Frame duration	10 ms/ 20 ms (optional)	20 ms for data and control/5 ms for control information on the fundamental and dedicated control channel
Forward pilot	Common code-multiplexed, dedicated time-multiplexed	Common code-multiplexed
Spreading	Variable length orthogonal sequences for channel separation, long Gold codes for user separation	Variable length Walsh sequences for channel separation, long M-sequence for user separation
Power control	Open and fast closed-loop (1.6 kHz)	Open and fast closed-loop (800 Hz)

essary to interconnect MS's to the fixed network. It is a central coordinating element for a set of BS's.

- *Connection.* MS's are connected to networks with radio links, and BS and MSC's are connected with high speed wired links.

With the rapid growth of wireless communication and access requirements, enormous recent developments have been undertaken by both academia and industry to make efficient utilization of the limited radio resources. Cell capacity can be expanded by introducing advanced CDMA platform, such as multicarrier architecture [27], deploying adaptive antenna arrays [24][40], using orthogonal spreading sequence sets [64][68][61], cell sectorization [60][59], cell splitting, etc. Though cell splitting increases the system capacity, it introduces a handoff problem induced by user mobility. When a mobile user moves from one cell to another, the MS should be served by the new cell in order to maintain service continuity. Handoff, which is one of the most critical subjects for cellular wireless networks, is the focus of this thesis.

1.2 Handoff in mobile systems

Handoff refers to the process in which an MS switches its connection from one BS to another BS to maintain service continuity with quality of service satisfaction. Furthermore, in the wireline network, a path rerouting procedure may take place if the current BS and the new BS are associated with two different MSC's. A handoff process consists of three phases: initiation phase, execution phase and path rerouting phase.

1. Initiation. The handoff initiation phase includes monitoring of the current radio channel conditions and identifying the need to initiate a handoff process. Received pilot signal strength (RPSS), signal-to-interference ratio (SIR), bit error rate (BER)

and distance are usually used to evaluate radio link quality among which the most commonly used indicator is RPSS for its simplicity.

2. Execution. The handoff execution phase includes the allocation of radio resources and the exchange of control signals.

3. Path rerouting. Since an MSC can only accommodate a limited number of BS's, when an MS moves to a new BS served by a different MSC, a path rerouting phase will take place to find an optimal connection based on network topology and traffic load distribution in each subnetwork.

1.2.1 Handoff categories

According to the BS or MSC involved in the handoff process, handoff can be classified into *intra-cell handoff* between sectors within a cell, *intra-switch handoff* between cells under the control of the same MSC and *inter-switch handoff* between cells belonging to two different MSC's. On the basis of the detection strategies employed, the handoff process can be divided into *mobile controlled handoff (MCHO)*, *network controlled handoff (NCHO)* and *mobile assisted handoff (MAHO)*. In MCHO, the MS is completely in control of the handoff process. The MS measures the signal strengths from all its neighboring BS's and initiates a handoff process when necessary. One critical defect of MCHO is that the MS has no information on the connection quality of other users and traffic load distribution in the system, and may trigger a handoff process in improper time epoch which will degrade cell capacity utilization. In NCHO, the network performs the measurements and makes the handoff decision. This type of handoff is not suitable for a rapidly varying communication environment and a high density user situation because of the introduced delay [12]. An MAHO distributes the measurement among MS's. The network makes decisions based on the feedback measurement reports. MAHO is widely used in second generation wireless systems such as GSM and IS-95 [22]. Depending on the way handoff is initiated, it may be soft or hard. Soft handoff is characterized by an MS having more than one radio connection during the handoff process. This scheme becomes possible in CDMA systems because of universal frequency reuse and the use of Rake receivers. In hard handoff, the old connection is terminated before a new one is activated, so there is only one radio connection at any point in the handoff process. Depending on whether an MS is having a call or not, there are ordinary handoffs and idle handoffs. Idle handoff takes place when an MS crosses the cell boundary while not in service. It is not accompanied by resource allocation and path rerouting. An ordinary handoff happens when an MS with an ongoing call crosses cell boundary and it will go through the three phases of initiation, execution and path rerouting.

1.2.2 Challenging issues

1. **User mobility**

   Wireless transmission relaxes the rigid fixed access point constraint and allows users to roam within the service area. However, random user mobility raises a challenging issue that needs to be dealt with in wireless networks. Location management and handoff are two main aspects of the user mobility issue [6][72][26][7]. Handoff is important for service continuity, especially for realtime services, when a user roams around the service area. As the cell size shrinks to enhance system capacity, an ongoing call may go through several handoffs during a call holding time which increases the possibility of a call being dropped before it hangs up. When a user randomly moves away from the serving BS, the required transmission power from the MS increases in order to achieve the predefined QoS. The received power should be closely monitored and a handoff process should be initiated at an appropriate time so that there is neither excessive handoff, nor excessive straying of the MS from its current BS. The time and manner in which a handoff is initiated are important factors in a successful handoff. Also, a well designed handoff initiation algorithm can reduce interference imposed on other ongoing users and potentially increase cell capacity. When a handoff occurs, signaling exchanges between the serving and the new BS's are involved. Effective signaling exchanges should introduce as little delay as as possible in order to reduce the possibility of data loss. The new cell where an MS is heading will execute resource allocation and try to provide the MS with the same QoS as agreed at the call setup time. With the random quality of user mobility and traffic loads in the system, the spare resource in the system is randomly varying which makes it difficult to maintain the service continuity and the QoS satisfaction. Therefore, intelligent resource allocation is essential for a successful handoff.

2. **Radio transmission**

   The available frequency bands are subject to technical limitations and the usage of the available radio frequency bands is regulated and coordinated by special organizations. Unlike wireline network resources, which can be increased simply by adding more devices, the available radio resource in wireless networks is very limited and precious. Radio transmission experiences path loss and is distance-limited. Besides large-scale attenuation due to path loss, wireless transmission also experiences slow fading due to shadowing and fast fading due to reflection of transmitted signals. Channel fading causes high transmission error rate. Power control is one way to smooth out transmission errors and deal with the hostile communication environment. Channel fading increases the system resources required to achieve the same QoS requirement and thereby reduces the system capacity. Since signals are transmitted in open air, transmissions from different users interfere with one another if they are using the same frequency band. Higher transmission power from one user increases interference with other users which causes lower SIR and high bit error rate. Therefore, wireless networks are also interference-limited, and the radio resource should be utilized efficiently. Efficient resource utilization should support as many users as possible with the satisfaction of QoS requirements. In fact, under the constraint of a low connection blocking rate, the resources can never be fully utilized.

3. **Multimedia service**

   Future wireless networks are expected to support both traditional voice service and Internet-based multimedia services with a wide range of QoS requirements in terms of transmission rate, delay, bit error rate, packet loss rate (PLR), etc. For example, voice and video are two types of realtime traffic that need immediate resource reallocation and cannot tolerate delay in the handoff process. In addition, both voice and video traffic are insensitive to tolerable data loss. Voice traffic has constant information generation rate during speech state, while a video source usually has variable information generation rate in an active period. On the other hand, non-realtime data traffic has strict packet loss rate requirements, but is insensitive to delay. Resources can be allocated to data transmission only when all other types of realtime traffic have been scheduled for transmission. In a system supporting all types of traffic, different amounts of resources should be allocated to different types of traffic in order to satisfy their respective QoS requirements. Services requiring higher QoS require more resources. Multimedia services coexisting in a system make the handoff process a comprehensive issue. In CDMA, the resource sharing property raises the opportunity to multiplex different types of traffic flexibly and to improve radio resource utilization. The coordination of different types of traffic to the satisfaction of their QoS requirements while capturing the CDMA transmission characteristics for efficient radio resource utilization is a challenging problem.

1.3 Related research work on handoff

An enormous amount of research work on handoff has been carried on since wireless communication came into being. As cell size is ever decreasing to enhance system capacity, handoff issues draw ever more attention in both academic and industry areas.

1.3.1 Cell capacity and call admission region derivation

Since handoff strategies go hand in hand with call admission control (CAC) schemes, cell capacity and call admission region must be determined on the basis either of online interference estimation, such as in [50][43], or of offline statistical calculation of the maximum acceptable number of users as in [42][65][28]. Practical capacity analysis for CDMA systems has been a challenging issue, especially in the presence of channel fading, imperfect power control, user mobility and multimedia services.

Power control is a basic system requirement for CDMA systems to combat the near-far problem so that the received powers at the BS from the same type of calls are the same to achieve the same QoS. In a CDMA system supporting heterogeneous traffic, power control can also be used to achieve a required power for each type of call in order to limit the multiple access interference (MAI) to other users. In 3G WCDMA and cdma2000 systems, the reverse link closed-loop power control keeps the signal-to-interference ratio of each connection at a target value. As the relatively fast variations associated with Rayleigh fading may at times be too rapid to be tracked by closed-loop power control, the target SIR cannot be guaranteed all the time. The effect of imperfect power control on cell capacity derivation has been discussed in [65][39].

The variability of the user spatial density due to mobility can impact significantly on offered traffic characteristics and hence on the CDMA radio interface capacity, especially in a microcellular environment. Impact of user mobility on cell capacity is investigated in [71][10].

Cell capacity of a system supporting multimedia services with different transmission rate and BER requirements has been discussed in detail in [49][18][67][65]. Multicode CDMA is a promising technology for the support of multimedia services, wherein each user employs a number of orthogonal codes for variable transmission rate satisfaction while semi-orthogonal pseudo-random codes are used for separation between users. The capacity of a multicode CDMA system supporting multimedia services is investigated in [42][9].

A media access control (MAC) protocol regulates how different users share and compete for the available system capacity. A packet scheduling scheme in the MAC protocol coordinates the packet transmission from different users so that: i) the required packet delay and packet loss requirements of different users can be guaranteed, and ii) high packet throughput can be achieved to accept more users. An efficient packet scheduling scheme can extend the call admission region and improve call level performance in terms of new call blocking and handoff call dropping probabilities while improving resource utilization efficiency [32].

1.3.2 Resource reservation

Since from a user's point of view, dropping a handoff call in progress is more annoying than blocking a new call at the time of call setup, handoff calls are given higher admission priority than new calls to reduce the probability of forced termination [19][55][69]. Resource reservation is an efficient way to prioritize handoff calls. In [46][15], a fixed number of channels are reserved exclusively for handoff calls. Fixed channel reservation has the risk of increasing new call blocking probability and under-utilizing the radio resources. To reduce the probability of handoff call dropping, it is better to have some knowledge of user mobility information so that the resources are reserved for potential handoff calls. User mobility information is determined by a user's location, velocity and acceleration. Different approaches have been proposed to predict future user mobility information from present and past mobility information. In [76], a mathematical formulation is developed for systematically tracking the random movement of an MS in a cellular environment. A hierarchical user mobility model, together with appropriate pattern matching and Kalman filtering techniques are used in [48] for user mobility estimation and prediction to improve the connection reliability and bandwidth efficiency. A fuzzy logic estimator for mobile positioning using received signal strengths is proposed in [63]. Dynamic channel reservation schemes based on user mobility are designed to decrease the handoff call dropping probability while keeping the new call blocking probability as low as possible [44][52][16]. Probabilistic channel reservation is another efficient way to improve resource utilization [17][66][75] wherein the probability of the potential number of calls that will be handed off from neighboring cells is estimated, and resources are reserved dynamically reflecting the current cell occupancy status and the user mobility. Zhuang et al [75] also propose a resource reservation scheme based on measurement of the received pilot signal strength. Each MS keeps listening to the pilot signal from neighboring BS's to detect the time epoch when handoff shall be triggered to a target cell with the strongest RPSS value. To improve resource utilization, reserved resources are usually used as a common reservation pool for all the coming handoff calls[13][14][30]. While dynamic resource reservation schemes achieve better performance measures in a non-stationary communications environment, they also introduce additional computation and signaling loads to the system.

1.3.3 Transmission rate adaptation

As future communication networks are expected to handle a variety of traffic types, handoff in mixed media, especially in the mixed voice-data cellular networks, have been widely studied. Performance measures such as call blocking and call dropping probabilities are applicable only to realtime traffic, but may not be suitable for data traffic. Packet loss rate and transmission delay are more suitable for data traffic evaluations. In [53][54][11], two-dimensional traffic models are proposed to support voice and data traffic simultaneously. As data is a time-insensitive service, handoff data packets that cannot be transmitted at arrival are queued. The queued handoff data packets can even be transferred from the buffer in the old BS to that in the target BS when the mobile user slips far into the target cell. Since data traffic does not ask for immediate resource allocation in handoff process, more attention has been put on realtime multimedia systems wherein resource management and handoff play an important role in service continuity and QoS assurance [51][57]. In [73], resource reservation is based on local information alone. A Wiener prediction scheme and a time series analytical method are applied to estimate potential resources for the coming multimedia handoff calls. A dynamic bandwidth sharing scheme is proposed in [31], where realtime traffic can borrow the necessary bandwidth from existing non-realtime traffic when the available bandwidth in the target cell is insufficient. However, resource reservation for broadband multimedia traffic has the risk of under-utilizing radio resources. A sub-rating channel assignment strategy for handoffs in a system with homogeneous service is proposed by Lin et al. [47] which reduces handoff call dropping probability by temporarily dividing a full-rate channel into two half-rate channels, one to serve the existing call and the other to serve the handoff call. For cellular multimedia wireless networks, this scheme exploits the ability of mobile users to tolerate some degradation in QoS. In fact, this is true for most multimedia services involving voice and video applications where multi-level or hierarchical source/channel coding schemes can be used. Reference [74] introduces two classes of traffic. The first class is assumed to be an adaptive QoS class that accepts channel sub-rating, while the other does not accept QoS degradation. However, a simple sub-rating scheme is not practical for a multimedia CDMA system due to its time-variant call admission region, neither is it suitable for traffic types with different transmission rates and performance requirements.

1.3.4 Summary

In summary, though research work on handoff has been extensively done. Most of the schemes and algorithms proposed in literature are specific on only one or two aspects of handoff. A more comprehensive research which can accommodate all the above listed issues in handoff is expected to be proposed and studied.

1.4 Research motivation and objectives

The evolution of wireline to wireless communications has freed users from fixed access points, and realized communications anywhere at anytime. Mobility is one of the essential issues in cellular wireless networks, and handoff is an integral part of mobility management. In the literature, user mobility characteristics are simplified with exponentially distributed cell residence time, which is also the inter-handoff time. However, the exponential inter-handoff time assumption underestimates the handoff rates, which leads to an actual dropping probability higher than that designed for cellular networks [20]. To deal with a dynamic system with highly random user movement, an appropriate mobility model which can mimic the behavior of human movements and represent the traffic flow among cells is required. It is quite difficult, if not impossible, to provide accurate and tight statistical models for randomly varying system parameters and even when a statistical model is available, only long run average performance measures can be provided. Consequently, measurement-based approaches have received much attention for their ability to capture the non-stationary characteristics of user mobility, propagation loss, traffic load, etc. Adaptive strategies based on the measured parameters can closely track the varying system parameters to guarantee the required QoS at all times. Considerable research work has been conducted in handoff and resource management, but it is a challenging task to apply measurement-based approaches to adaptively handle the time-variant handoff attempts with different mobility information and traffic characteristics, while keeping a low rate of new call blocking events. Handoff prioritization is a requirement from a user's point of view, while viewed from a service provider's perspective, revenue maximization with the promised QoS provisioning is the first concern. Scarce radio resources are expected to be efficiently utilized; namely, an efficient resource management scheme shall admit as many users as possible.

Different system characteristics at the physical, link and network layers may be correlated, and consideration of the individual layer in isolation can lead to inefficient resource utilization and bring up problems in QoS provisioning. Research at the physical layer studies the cell capacity in terms of simultaneously transmitted packets with different transmission rates and bit energy-to-interference density ratio requirements; there is no consideration of traffic characteristics, link layer packet loss and delay requirements and network layer connection blocking and resource utilization at physical layer studies. Packet scheduling at the link layer, CAC and handoff at the network layer are performed based on the physical layer analysis. Packet scheduling is necessary to coordinate the transmission rate, packet loss rate and packet delay of different traffic types to satisfy the packet level QoS requirements and to achieve high throughput for a given number of users. A well designed packet scheduling scheme shall expand the call admission region for better call accommodation. The number of admitted users is determined by CAC and handoff at the network layer. If a CAC scheme cannot admit enough users, it may be impossible for the packet scheduling scheme to achieve the best packet throughput and the call level connection blocking requirement may not be satisfied; while if a CAC admits an excessive number of users, it may cause network congestion, and the required QoS may not be ensured. Also, a CAC scheme should make an admission decision for a user such that once admitted, a call receives guaranteed QoS throughout its life-time.

The motivation of this research is to address handoff issues in integrated multimedia wireless cellular networks by taking the propagation environment, user mobility, traffic characteristics, and traffic load variations into consideration. Specifically, the objective of this research is to propose and to evaluate the performance of handoff management schemes, which can:

- integrate the physical and link layer analysis for an effective call admission region derivation;

- provide a measurement-based adaptive resource reservation strategy for handoff with different user mobility patterns and traffic load variations and subject to handoff prioritization and efficient radio resource utilization;

- provide an efficient CAC and handoff scheme to accommodate integrated narrowband voice and broadband multimedia services by taking advantage of the different traffic characteristics in wireless transmissions.

1.5 Contributions of the thesis

The main contributions of this thesis are listed as follows:

- uplink cell capacity is properly estimated for a cellular CDMA network;

- an efficient packet scheduling and partial packet integration scheme is proposed to guarantee the packet loss rate requirements of different traffics;

- a call admission region is effectively derived for call admission control and handoff management;

- the relation between the packet error rate due to wireless transmission and the packet loss rate due to link layer congestion is derived and discussed;

- a user mobility model mimicking practical user movement is applied to the resource reservation scheme for accurate handoff prediction;
- a reservation-request time threshold is introduced in an adaptive resource reservation scheme for better handoff performance and resource utilization.
- a pooling algorithm is proposed to improve multiplexing gain, to eliminate false resource reservation and to simplify the signaling process;
- an adaptive handoff priority scheme supporting multimedia services is proposed and its performance is evaluated.

1.6 Overview of the thesis

The remainder of this thesis is organized as follows. Chapter 1 defines the system model, including the system structure, MAC layer structure, wireless propagation model, traffic model, and QoS and grade of service (GoS) parameters definition. Uplink cell capacity estimation and call admission region derivation are presented in Chapter 3. A measurement-based adaptive resource reservation strategy for handoff is proposed and its performance is evaluated in Chapter 4. Chapter 5 proposes an adaptive handoff priority scheme for wireless MC-CDMA cellular networks supporting multimedia applications. Numerical results are presented in Chapter 6 to evaluate the proposed resource reservation and handoff priority scheme. Concluding remarks and future research work are given in Chapter 7.

Chapter 2

Proposed Research

In this chapter, we first present the proposed research. Then the CDMA system structure and the scope of this research are defined.

2.1 Proposed research

In a packet switching wireless cellular system, the overall packet error rate for a connection includes on the one hand, a packet error rate in the physical layer wireless transmissions caused by interference due to other transmissions, channel fading and ambient noise, and on the other hand, a link layer packet loss rate caused by the failed scheduling of packets with delay exceeding the required delay bound. As the wireline network has much larger capacity than the wireless network, we ignore errors introduced in the wireline domain and focus on the wireless uplink transmission, namely the transmission from the mobile station to the base station. Uplink cell capacity is limited by the maximum number of packets that can be transmitted simultaneously with the satisfaction of the prescribed bit error rate requirement. A predefined bit error rate is applied for all traffic classes to guarantee the same performance in wireless transmission. Therefore, the uplink cell capacity derivation is simplified as it is traffic independent and accordingly depends solely on communication environments. For traffic types with different packet error rate requirements, the additional error rate is converted from the link layer packet loss rate due to failed scheduling of packets for transmission. On the basis of the uplink cell capacity, a packet scheduling scheme is presented, which achieves different packet loss rate requirements among different traffic types and satisfies a fair packet loss constraint among ongoing calls with the same traffic type. The traffic characteristics at the call level and the packet loss constraint at the packet level are applied to derive the call admission region, which is used by the serving mobile switching center to execute call admission control and handoff management.

To give call admission priority to handoff calls over new calls, two handoff priority schemes are proposed in this thesis. First, we propose a measurement-based adaptive resource reservation scheme. A reservation is made when a resource reservation request is triggered through a threshold mechanism by monitoring the RPSSs. The RPSS difference between the serving and the target BSs is used to determine whether a resource reservation request and handoff process should be initiated or not. The RPSS difference rather than the absolute RPSS value from the serving BS is used because CDMA is interference-limited. An MS should always be connected to the BS with better communication quality even though the RPSS from the serving BS is sufficient for satisfactory communication. To balance the fast-moving-user handoff failure due to lack of reserved resources and the resources wasted because of slow-moving users, a time threshold, which is the time remaining until a handoff is expected to happen, is introduced for administering resource reservation. This is based on the fact that the time during which a target BS prepares the requested resources depends solely on the traffic load and service characteristics, not on user mobility. A reservation request for a fast-moving user is triggered with much larger RPSS difference, while for a slow-moving user the RPSS difference is much smaller. Therefore, the RPSS difference adapts to different user mobility so that the amount of reserved resources can be appropriately set aside and efficiently utilized. In addition, the effect of traffic load variations on the time threshold in resource reservation is also investigated. When the system traffic load is heavy, it is desirable to initiate resource reservation requests early in order for the target cell to have the requested resources ready before a handoff occurs. On the other hand, in a light traffic load situation, the resource reservation requests can be sent to the target BS late to eliminate excessive new call blocking events. An adaptive time threshold can also capture the traffic burstiness. Thus, the reservation-request time threshold can adapt to the traffic load variations to achieve better performance. The reserved resources for individual requests form a common reservation pool which is used exclusively for handoff calls on a first-come-first-serve (FCFS) basis. This common reservation pool for handoff calls can take advantage of multiplexing use of the available reserved resources to decrease the handoff call dropping events.

In order to prioritize handoff calls further, an adaptive handoff priority scheme is proposed, which adjusts the transmission rate of ongoing rate-adaptive calls to gain capacity for handoff call prioritization. Whenever an incoming handoff call fails to be served with either reserved resources or spare resources, a transmission rate renegotiation process is triggered to degrade the service of any high-quality rate-adaptive call in progress in order to accept the handoff call. The service degradation process can continue till either the handoff call is admitted or there is no high-quality rate-adaptive call in the cell. The scheme aims to decrease the handoff call dropping probability without increasing the new call blocking probability.

The novelty of the proposed research includes (i) the transformation of different packet error rate requirements for different types of traffic to different packet loss rate requirements in the packet scheduling scheme, (ii) the joint consideration of the physical, link and network layer characteristics in the derivation of the call admission region, (iii) the integrated study of user mobility and traffic variations on handoff initiation and execution, and (iv) a handoff call priority admission scheme which exploits the transmission rate adaptability of rate-adaptive video calls. The proposed CAC and handoff management scheme takes advantage of multiplexing packet transmission property over wireless links to achieve better new call blocking and handoff call dropping probabilities and efficient cell capacity utilization and, at the same time, to provide satisfactory quality of service for each individual call.

2.2 System structure

In a Frequency Division Duplex/MC-CDMA (FDD/MC-CDMA) network, the uplink and downlink use different frequency bands to provide frequency isolation, so there is no interference between the uplink and downlink transmissions. MC-CDMA is adopted to support multi-rate transmission because its unified architecture can easily integrate traffic streams with significantly different transmission rates with all transmission channels having the same bandwidth and spread spectrum processing gain. In an MC-CDMA system, all the data signals over the radio channel are transmitted at a basic rate, B bits/sec. Any connection can only transmit at rates mB bits/sec, where m is a positive integer. When an MS needs to transmit at rate mB bits/sec, it converts its data stream, serial-to-parallel, into m basic-rate streams. These parallel streams are spread by a set of orthogonal codes generated by the so called "sub-code concatenation scheme". To simplify analysis, we do not distinguish interference introduced between code channels and subcode channels.

Fig. 2.1 shows a two-tier network structure which can effectively support soft handoff. The first tier is a mesh connection of MSCs which connect the wireless sub-networks with

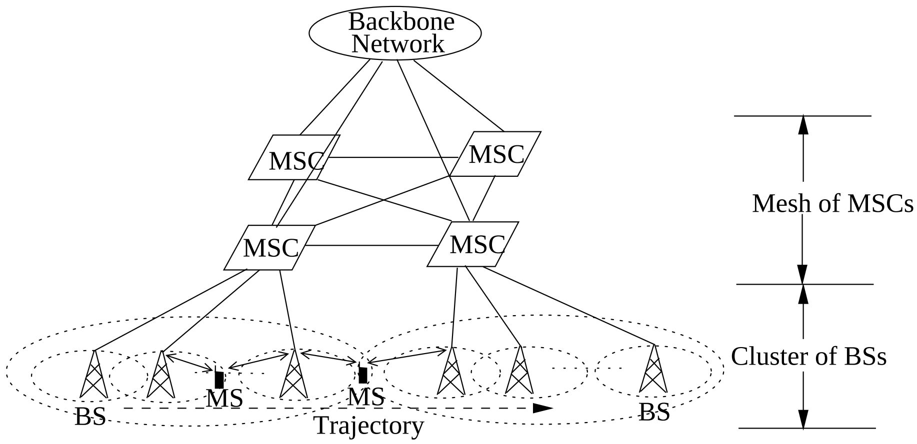

Figure 2.1: System structure

the backbone network. An MSC collects status information of all its serving BSs and performs most of the resource management functions of a CDMA wireless network, including call admission control, mobility management, radio resource management, and so on. In order to support soft handoffs effectively, a Selection/Distribution Unit (SDU) residing at the MSC schedules the packet transmission and reception for both downlink and uplink among the associated BSs. When an MS is involved in soft handoff, the SDU schedules the simultaneous transmission of uplink packets from the associated BSs. According to the bit energy-to-interference density ratio (E_b/I_0) of the transmissions, the SDU chooses the best one and discards the other copies of the same packet. At the downlink, the SDU distributes several copies of the same packet to the involved BSs for transmission, and the Rake receiver at the MS combines those copies together for better transmission quality. The second tier consists of a cluster of BSs, each connected to a serving MSC. A BS plays two roles. First, it takes part in the radio resource management under the control of the serving MSC. Second, it works as the interface between an MS and its serving MSC. An MS keeps an active connection with the serving BS by monitoring the RPSS. Each mobile user samples the RPSS values from its serving and the neighboring BSs every Δt s. Without loss of generality, we assume that only two BSs are involved in a given handoff. The handoff decision is based on three thresholds: a reservation-request time (T_{rev}) threshold, a target BS add (ADD) threshold and a previous BS drop (DROP) threshold. The T_{rev} threshold is selected on the basis of the remaining time until a soft handoff process takes place, while both ADD and DROP thresholds are based on the RPSS differences. The ADD and DROP thresholds together define the soft handoff area. We adopt the MAHO approach in the system, and whenever the remaining time reaches T_{rev}, or when the ADD or DROP threshold is reached, a message is formed by the MS and sent back either by piggyback or by a specific signaling channel to the serving MSC via the serving BS for decision making. If the target BS is served by a different MSC, the message will be further forwarded to the target MSC for a decision. This message includes the target cell identification, the potential handoff call traffic characteristics and all the other handoff-related parameters. The MSC directs the target BS to reserve resources and the MS to add or drop an active BS. With the deployment of a central data collection and mediation point for the radio access network, the knowledge of traffic load information makes it possible to determine the reservation-request time thresholds adaptively so as to maintain performance requirements.

2.3 MAC layer structure

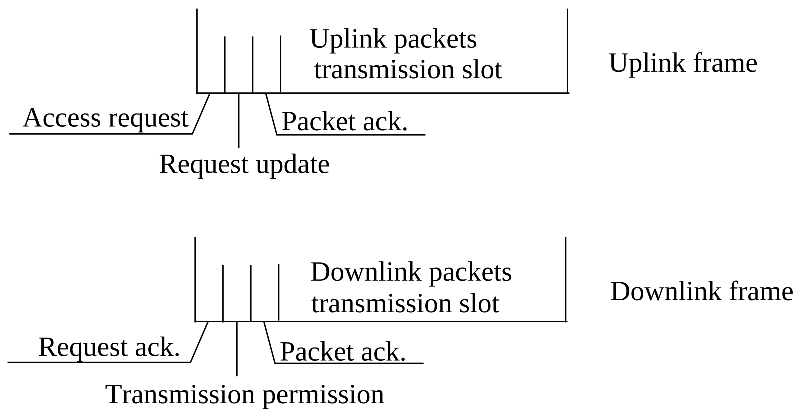

Figure 2.2: Uplink and downlink frame structure

The MAC protocol resolves contention resulting from radio resource sharing among a number of users. The frame architecture for uplink and downlink is shown in Fig. 2.2. In the uplink transmission, there are a access request mini-slot, a request update mini-slot and a packets acknowledge mini-slot for downlink transmission, followed by the uplink packets transmission slot. The request access mini-slot is used for both new and handoff calls to request call admission or for admitted calls to access transmission when there are packets generated. The request update mini-slot is used for dynamically adjusting the transmission rate of calls which can either adapt the transmission rate to the resources available or have variable transmission rate. The downlink packet acknowledgment mini-slot is followed by uplink packet transmission slot. The downlink transmission in each frame starts with the request acknowledgment which acknowledges the successfully received requests. The transmission permission mini-slot broadcasts the packet scheduling result for the uplink transmission in the next frame.

For the request access slot, a direct sequence (DS)-CDMA with slotted ALOHA random access protocol is used. Dedicated codes are used for the requests. The BS broadcasts these codes to MSs. When an MS is ready to send a request, it chooses randomly a code from the code pool and a request access slot for transmission. The user identity is included in the request. If a new or handoff call initiates a call admission request, the transmission parameters will be included in the request as well. If the request has been successfully received, the BS will broadcast the user identity in the request acknowledgment slot. Meanwhile the transmission parameters are sent to the serving MSC for call admission for a new arriving call or scheduling decision for an admitted call. If the MS does not receive its identity, the request process repeats in the next frame. When the MS has received its identity in the request acknowledgment slot, it will listen to the transmission permission slot to obtain the number of packets that can be transmitted in the next frame. Data packet transmissions start at the beginning of a packet transmission slot. Different calls can transmit simultaneously in one packet transmission slot. The satisfaction of QoS for each call in the system is guaranteed by limiting the number of calls admitted into the system and properly scheduling packet transmission from all the calls.

2.4 Propagation model

In this research, we assume mobile users are uniformly located in a cell. The received downlink pilot signal strength at each MS is affected by three components: path loss which depends on the user location, slow shadowing and fast fading. It is assumed that fast fading can be taken care of by the physical layer functions so that the channel is characterized by path loss and shadowing [23]. Given the location of an MS as (r, θ), the channel gain at time t_k from the i-th base station, $G_{k,i}(r, \theta)$, is modeled as

$$G_{k,i}(r, \theta) = d_{k,i}(r, \theta)^{-\eta} 10^{\zeta^*_{k,i}/10}, \qquad (2.1)$$

where $d_{k,i}(r,\theta) = d_i/d_0$, d_i is the distance between the i-th BS and the MS and d_0 is the reference distance from the BS; η is the path loss exponent; and $10^{\zeta^*_{k,i}/10}$ is the lognormal shadowing effect with $\zeta^*_{k,i}$ being a Gaussian distributed random variable with mean zero and standard deviation σ. By taking logarithms on both sides of (2.1), we can get the channel gain in dB as

$$G_{k,i}(r,\theta)_{[dB]} = -10\eta\mathrm{lg}[d_{k,i}(r,\theta)] + \zeta^*_{k,i}. \tag{2.2}$$

Because $-10\eta\mathrm{lg}[d_{k,i}(r,\theta)]$ is dependent on distance, $G_{k,i}(r,\theta)_{[dB]}$ is Gaussian distributed with mean $-10\eta\mathrm{lg}[d_{k,i}(r,\theta)]$ and standard deviation σ. After some manipulation, (2.1) is re-expressed as

$$\begin{aligned} &G_{k,i}(r,\theta) = 10^{\zeta_{k,i}(r,\theta)/10} \\ &\zeta_{k,i}(r,\theta) \sim N(-10\eta\mathrm{lg}[d_{k,i}(r,\theta)], \sigma^2). \end{aligned} \tag{2.3}$$

To capture the large scale autocorrelation property of shadow fading, a lognormal first order autoregressive (AR-1) model is applied [25]. Therefore, $\zeta^*_{k,i}$ can be written recursively as

$$\zeta^*_{0,i} = \sigma^2 W_{0,i},$$

$$\zeta^*_{k+1,i} = \varrho_{k,i}\zeta^*_{k,i} + \sigma\sqrt{1-\varrho^2_{k,i}}W_{k,i},$$

where $\varrho_{k,i}$ is the correlation coefficient of $\{\zeta^*_{k,i}\}$,

$$\varrho_{k,i} = \exp(-v_k\Delta t/\bar{d}_i).$$

$\{W_{k,i}\}$ are independent and identically distributed (i.i.d) Gaussian random variables with zero mean and unity variance; v_k is the mobile user's velocity and $\bar{d}_i$ is the shadow fading correlation distance. To incorporate the dependence of shadowing among multiple BSs, the model in [70] is adopted whereby the random component of the decibel gain for the i-th BS ($i = 0, 1, 2 \cdots$) is expressed as

$$\zeta^*_{k,i} = \phi y + \varphi y_i, \tag{2.4}$$

with

$$\begin{aligned}
&\frac{E[\zeta^*_{k,i}\zeta^*_{k,j}]}{\sigma^2} = \phi^2 = 1 - \varphi^2 \\
&\phi^2 + \varphi^2 = 1 \qquad \phi^2 = \varphi^2 = \frac{1}{2} \\
&E[y] = E[y_i] = 0 \qquad \text{for all i} \\
&Var[y] = Var[y_i] = \sigma^2 \qquad \text{for all i,}
\end{aligned} \tag{2.5}$$

where the first term of (2.4) is the component in the near field of the MS which is independent of the receiving BS; the second term of (2.4) is the component that is independent from one BS to another. Therefore, $p_{k,i}$, the RPSS signal received is formulated in decibel as

$$p_{k,i}(dB) = p_{0,i} + 10\ \log_{10} G_{k,i}(r, \theta), \tag{2.6}$$

where $p_{0,i}$ is a constant determined by the transmitted power, the wavelength, and the antenna gain of the i-th BS.

2.5 Traffic model

Suppose there are K classes of traffic with different transmission rate, bit error rate, packet loss rate and delay requirements. Let B bits/sec be the basic transmission rate of a single code channel, and let a type i call have a transmission rate of c_iB, where c_i is a positive integer representing the number of subcode channels used for class i calls. For each traffic class, the wireless transmission bit error rate is defined by the upper bounded outage rate $(P_{out})_i$, which is the probability that the bit error rate exceeds the prescribed value. For a given transmission environment and coding and detection schemes, the BER requirement corresponds to the bit energy-to-interference-density ratio E_b/I_0 requirement. It is assumed that all calls of the same traffic class have the same E_b/I_0 requirement, denoted as $(E_b/I_0)_i$. It is noted that for a class i call requiring c_i subcode channels, each subcode channel will have the same $(E_b/I_0)_i$ requirement. For real-time service, a user requires its packets to be delivered to the destination within a delay bound; otherwise, the packets will be worthless and discarded. Here we only consider the access delay over wireless links, which is denoted as time-out value. For each session, a buffer is required at the MS side to accommodate bursty packet generation. Each arriving packet at the buffer is labeled with a time-out value (delay bound at the first arrival) in frames. This value is decreased by one for every frame interval till the packet is scheduled to transmit. The packet loss due to exceeding the delay bound is referred to as packet loss rate.

Voice and video are two types of real-time traffic requiring immediate resource allocation during handoff to maintain service continuity. Fig. 2.3 shows an ON-OFF model to represent a voice call with a the probability of transition from OFF (silent) state to ON (talkspurt) state and b from ON state to OFF state. The activity factor, β_1, which is defined as the fraction of average time that a call is in the ON state, is given by $a/(a+b)$. When in ON state, information is transmitted at a constant rate of R bits/sec to achieve small delay jitter requirement. The number of subcodes needed is set to $\lceil R/B \rceil$, which usually has the value of 1, where $\lceil \cdot \rceil$ is a ceiling function. Video calls generate packet rate according to their characteristics and the coding schemes used. Hierarchical source coding, which adapts the packet generation rate to the resources available, is often used for video coding [45]. Two types of video calls are defined according to the packet generation rate, namely high-quality coding video call and low-quality coding video call. Each video call can be approximately represented by a superposition of N ON-OFF minisources as in Fig. 2.3 with parameters a, b and R different from those of the voice call model. An approach to calculate the parameters a, b and R can be found in [62]. If the values of a and b, which can be used to decide the activity factor β_2 of a minisource, are given, different quality video calls can be represented by different values of R. It is noted that the number of subcodes needed for high- and low-quality video calls are time-varying.

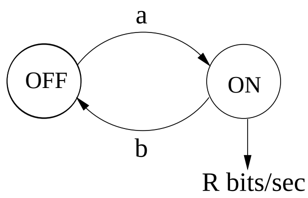

Figure 2.3: Two-state ON-OFF model

2.6 GoS parameters

The GoS parameters under consideration are new call blocking probability, P_b, and handoff call dropping probability, P_d, in a long run and new call blocking rate, R_b, and handoff call dropping rate, R_d, in a short term. The new call blocking and handoff call dropping rates are short-term performance measures of new call blocking and handoff call dropping probabilities. To achieve maximum revenue for the service provider and at the same time to meet user satisfaction, the GoS measures are defined as a combination of new call blocking and handoff call dropping probabilities(rates):

$$\text{GoS} = P_b(R_b) + \gamma P_d(R_d), \tag{2.7}$$

where γ is a weighting factor to put greater importance on handoff call dropping probability(rate). For example, γ equals 10 is used in [58]. It is noted that $P_b(R_b)$ and $P_d(R_d)$ are a pair of conflicting performance parameters. Insufficient resource reservation will cause high handoff call dropping probability(rate), while excessive resource reservation makes new call blocking probability(rate) high.

2.7 Summary

The proposed research and research scope have been presented in this chapter. MC-CDMA is used to support multimedia services in a cellular CDMA system. How to derive a proper call admission region and make efficient resource reservation for handoff call prioritization and at the same time to maintain new call admission performance is an important issue. Before this issue is addressed, uplink cell capacity and the call admission region are first derived in the next chapter.

Chapter 3

Uplink Call Admission Region

In this chapter, we first estimate the uplink cell capacity under a BER requirement. Given the derived cell capacity, a packet scheduling scheme is proposed which meets different PLR requirements for different types of traffic and satisfies a fair packet loss constraint among ongoing calls with the same traffic type. The traffic characteristics at the call level and the packet loss constraint at the packet level are applied to derive the call admission region, which is used by the MSC to make call admission decisions and to manage handoff.

3.1 Bit error rate and packet loss rate

In a packet switching multimedia wireless network, the overall packet error rate for user i, P_e^i, includes error rates due to wireless transmission, and packet loss.

$$P_e^i = P_{PLR}^i + (1 - P_{PLR}^i) \times P_{WTE}^i, \tag{3.1}$$

where P_{PLR}^i is the packet error rate due to packet loss and P_{WTE}^i is the packet error rate due to wireless transmission error. To simplify the uplink cell capacity estimation, all the uplink channels are specified to provide the same P_{WTE}^* value. It is noted that P_{WTE}^* should not be greater than the most stringent overall packet error rate requirement of the traffic in the system. In this way, different error rate requirements for different users can be achieved by providing different packet loss rates with a proper packet scheduling scheme. For user i, the required P_{PLR}^i to satisfy a predefined P_e^i with the specified P_{WTE}^* is expressed as

$$P_{PLR}^i = \frac{P_e^i - P_{WTE}^*}{1 - P_{WTE}^*}. \tag{3.2}$$

We denote the bit error rate caused by radio transmission by P_{BER}^o. If the physical layer error detection and correction are taken into consideration, the actual bit error rate seen by the link layer is P_{BER}^*. Therefore, the packet loss rate caused by the bit error rate at the link layer can be expressed as:

$$P_{WTE}^* = 1 - (1 - P_{BER}^*)^{L_{en}}, \tag{3.3}$$

where L_{en} is the packet length.

Let P_{PLR}^{1}, P_{PLR}^{2h} and P_{PLR}^{2l} denote the packet loss rate for voice, high- and low-quality video calls, respectively. A large P_{PLR} value implies a small P_{WTE} requirement and vice versa. It is expected that more efficient resource utilization can be achieved by choosing a less stringent P_{WTE} value because a more stringent P_{WTE} value indicates a small cell capacity and a degradation of multiplexing gain at the packet scheduling process. Also the noisy communication environment makes a stringent P_{WTE} requirement inappropriate for practical implementation. The effect of different P_{PLR} and P_{WTE} pairs on performance measures will be demonstrated through simulations in Chapter 6.

3.2 Uplink cell capacity estimation

A typical hexagonal cell structure with dashed lines representing inner and outer cells of radii R_I and R_O is shown in Fig. 3.1. The radius of the cell is R_C. The relative handoff area is defined as $\alpha = (R_O - R_I)/R_C$. The reverse link cell capacity of a multiple cell structure CDMA system is limited by interference from simultaneously transmitted packets from the other active calls connected and power controlled by the reference BS, the calls in soft handoff and power controlled by the neighboring BS, and the calls from outside the reference BS. With the assumption of uniformly located users in inner cell and soft handoff

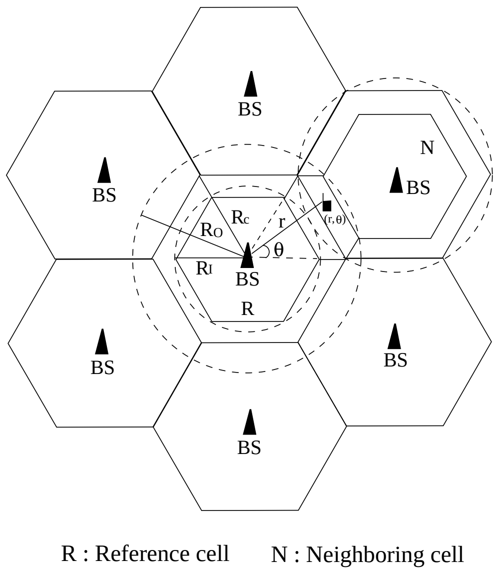

Figure 3.1: Cell structure

area with user density κ_I and κ_S, respectively, when the inner cell is loaded with N_I and the soft handoff area with N_S active MS's, κ_I and κ_S are expressed as

$$\begin{aligned} \kappa_I &= \frac{N_I}{\pi R_I^2} \frac{calls}{unit\ area} \\ \kappa_S &= \frac{N_S}{\pi(R_O^2 - R_I^2)} \frac{calls}{unit\ area}, \end{aligned} \tag{3.4}$$

respectively. A uniform distribution of calls in a cell is a special case of $\kappa_I = \kappa_S$.

We first derive the probability, P_R, that an MS in soft handoff is power controlled by the reference BS. For demonstration simplicity, we ignore the time epoch subscript k and use R and N to denote the reference and a neighboring BS, respectively. The probability of an MS at location (r, θ) in soft handoff and power controlled by the reference BS, $P_R(r, \theta)$, is the probability that the channel gain to the reference BS is greater than the channel gain to the neighboring BS, namely

$$\begin{aligned} P_R(r,\theta) &= P\{G_R(r,\theta) \geq G_N(r,\theta)\} \\ &= P\{10^{\zeta_R/10} \geq 10^{\zeta_N/10}\} \\ &= P\{\xi_k(r,\theta) \geq 0\} \qquad R_I \leq r \leq R_O, \end{aligned} \tag{3.5}$$

where $\xi(r,\theta) = \zeta_R - \zeta_N \sim N(-10\eta \lg[d_R(r,\theta)] + 10\eta \lg[d_N(r,\theta)], \varphi^2 2\sigma^2) = N(d_{RN}, \sigma^2)$.

Normalizing $P_R(r,\theta)$ over the handoff area, P_R can be obtained as

$$\begin{aligned} P_R &= \frac{6}{\pi(R_O^2 - R_I^2)} \int_0^{\frac{\pi}{3}} \int_{R_I}^{R_O} P_R(r,\theta) r \, dr d\theta \\ &= \frac{6}{\pi(R_O^2 - R_I^2)} \int_0^{\frac{\pi}{3}} d\theta \int_{R_I}^{R_O} r \, dr \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp - \frac{(\xi - d_{RN})^2}{2\sigma^2} \, d\xi. \end{aligned} \tag{3.6}$$

The probability that an MS in soft handoff area is power controlled by the neighboring cell is then $P_N = 1 - P_R$.

It has been experimentally verified that imperfect received power can be modeled as a random variable S_e with lognormal distribution [8]:

$$S_e = S 10^{\xi_e/10}, \tag{3.7}$$

where S is the required received power if perfect power control can be achieved; ξ_e is a Gaussian random variable with mean zero and standard deviation σ_e. For user l, the serving base station will receive a composite signal including the desired signal power $S_e(l)$, interference from other users connected and power controlled by the same BS I_I , interference from users connected with the same BS but power controlled by a neighboring BS I_S, interference from outer cell users I_O and background noise n_0.

Define C to be the uplink cell capacity, and $C = \lceil N_I + P_R N_S \rceil$ represent the number of simultaneously transmitted packets from calls connected and power controlled by the reference BS; the interference to signal power ratio imposed from these interference packets is expressed as

$$\frac{I_I}{S_e(l)} = \sum_{i=1,i\neq l}^{C} \frac{S10^{\xi_e(i)/10}}{S10^{\xi_e(l)/10}}. \tag{3.8}$$

Since each item in (3.8) is an independently distributed lognormal random variable, by using Central Limit Theorem, $I_I/S_e(l)$ can be approximated by a Gaussian random variable with mean $E[I_I/S_e]$ and second moment $E[(I_I/S_e)^2]$ given as follows

$$\begin{aligned} E\left[\frac{I_I}{S_e}\right] &= E\left[\frac{\sum_{i=1,i\neq l}^{C} S10^{\xi_e(i)/10}}{S10^{\xi_e(l)/10}}\right] \\ &= E\left[\sum_{i=1,i\neq l}^{C} 10^{\frac{\xi'_e(i)}{10}}\right] \\ &= (C-1)e^{\beta^2\sigma_e^2} \qquad \beta = \frac{ln10}{10}, \end{aligned} \tag{3.9}$$

$$\begin{aligned} E\left[\left(\frac{I_I}{S_e}\right)^2\right] &= E\left[\left(\frac{\sum_{i=1,i\neq l}^{C} S10^{\xi_e(i)/10}}{S10^{\xi_e(l)/10}}\right)^2\right] \\ &= E\left[\sum_{i=1,i\neq l}^{C} 10^{\xi'_e(i)/5} + \sum_{j=1,j\neq l}^{C} \sum_{k=1,k\neq j,k\neq l}^{C} 10^{(\xi'_e(j)+\xi'_e(k))/10}\right] \\ &= (C-1)e^{4\beta^2\sigma_e^2} + (C-1)(C-2)e^{3\beta^2\sigma_e^2} \end{aligned} \tag{3.10}$$

where $\xi'_e \sim N(0, 2\sigma_e^2)$. So $Var[I_I/S_e] = E[(I_I/S_e)^2] - E^2[I_I/S_e]$.

Given an MS m with an active call located at (r, θ) in the soft handoff area and power controlled by a neighboring BS N, the received power level at the reference BS R is

$$\begin{aligned} S_I(r,\theta) &= \frac{d_R(r,\theta)^{-\eta} 10^{\frac{\varsigma_R^*}{10}}}{d_N(r,\theta)^{-\eta} 10^{\frac{\varsigma_N^*}{10}}} S 10^{\xi_e(m)/10} \\ &= S e^{\beta[\xi(r,\theta)+\xi_e(m)]}, \end{aligned} \tag{3.11}$$

with

$$d_R(r,\theta)^{-\eta} 10^{\frac{\varsigma_R^*}{10}} < d_N(r,\theta)^{-\eta} 10^{\frac{\varsigma_N^*}{10}} \qquad R_I \le r \le R_O. \tag{3.12}$$

Because users in handoff are near to the cell boundaries and power controlled by the first-tier neighboring BS's, the variation of $S_I(r,\theta)$ is small compared to the received power from outer cell users. Therefore we consider the interference from calls in soft handoff and power controlled by neighboring BS's and that from calls in outer cells separately. The sum of independent lognormal random variables can be approximated by another Gaussian random variable I_S. Accordingly the total interference-to-signal power ratio from this kind of calls, I_S/S_e, is Gaussian distributed with mean $E[I_S/S_e]$ and second moment $E[(I_S/S_e)^2]$ given as follows

$$\begin{aligned} E\left[\frac{I_S(r,\theta)}{S_e}\right] &= E[e^{\beta[\xi(r,\theta)+\xi_e(m)-\xi_e(l)]}, \xi(r,\theta) < 0] \\ &= E[e^{\beta\xi(r,\theta)}, \xi(r,\theta) < 0] E[e^{\beta\xi_e'}] \\ &= e^{\beta^2\sigma_e^2} \int_{-\infty}^{0} e^{\beta\xi} f(\xi)\, d\xi \\ &= e^{\beta^2\sigma_e^2} \int_{-\infty}^{0} e^{\beta\xi} \frac{1}{\sqrt{2\pi\sigma^2}} exp - \frac{(\xi - d_{RN})^2}{2\sigma^2}\, d\xi, \end{aligned}$$

$$
\begin{aligned}
E\left[\left(\frac{I_S(r,\theta)}{S_e}\right)^2\right] &= E[(e^{\beta[\xi(r,\theta)+\xi_e(m)-\xi_e(l)]})^2, \xi(r,\theta) < 0] \\
&= E[e^{2\beta\xi(r,\theta)}, \xi(r,\theta) < 0]E[e^{2\beta\xi_e'}] \\
&= e^{4\beta^2\sigma_e^2}\int_{-\infty}^{0} e^{2\beta\xi} f(\xi)\, d\xi \\
&= e^{4\beta^2\sigma_e^2}\int_{-\infty}^{0} e^{2\beta\xi}\frac{1}{\sqrt{2\pi\sigma^2}} exp - \frac{(\xi - d_{RN})^2}{2\sigma^2}\, d\xi.
\end{aligned} \tag{3.13}
$$

So $Var[I_S/S_e] = E[(I_S/S_e)^2] - E^2[I_S/S_e]$.

Accordingly,

$$
\begin{aligned}
E\left[\frac{I_S}{S_e}\right] &= 6P_N\kappa_S\int_0^{\frac{\pi}{3}} d\theta \int_{R_I}^{R_O} E\left[\frac{I_S(r,\theta)}{S_e}\right] r\, dr \\
Var\left[\frac{I_S}{S_e}\right] &= 6P_N\kappa_S\int_0^{\frac{\pi}{3}} d\theta \int_{R_I}^{R_O} Var\left[\frac{I_S(r,\theta)}{S_e}\right] r\, dr,
\end{aligned} \tag{3.14}
$$

which are functions of the number of simultaneously transmitted packets from calls in soft handoff area N_S.

The interference from outer cell users may vary widely because of their random locations in the outer cell. Given an MS m with an active call located at (r, θ) in the outer cell, the received power level at the reference BS R is

$$
\begin{aligned}
S_I(r,\theta) &= \frac{d_R(r,\theta)^{-\eta}10^{\frac{\varsigma_R^*}{10}}}{d_N(r,\theta)^{-\eta}10^{\frac{\varsigma_N^*}{10}}} S10^{\frac{\xi_e(m)}{10}} \\
&= Se^{\beta[\xi(r,\theta)+\xi_e(m)]} \qquad r \geq R_O,
\end{aligned} \tag{3.15}
$$

where

$$
d_N(r,\theta)^{-\eta}10^{\frac{\varsigma_N^*}{10}} = \min(d_i(r,\theta)^{-\eta}10^{\frac{\varsigma_i^*}{10}}) \qquad i \neq R. \tag{3.16}
$$

By using the similar approach as in the previous interference estimation, the outer cell interference to signal power ratio can be approximated by another Gaussian random variable, I_O/S_e, with mean and variance expressed as

$$\begin{aligned} E\left[\frac{I_O}{S_e}\right] &= \sum_{outercell} E\left[e^{\beta[\xi(r,\theta)+\xi_e(m)-\xi_e(l)]}\right] \\ &= \int\int_{outercell} E\left[e^{\beta[\xi(r,\theta)+\xi'_e]}\right]\kappa\, dA, \\ Var\left[\frac{I_O}{S_e}\right] &= \sum_{outercell} \{E[e^{2\beta[\xi(r,\theta)+\xi_e(m)-\xi_e(l)]}] - E^2[e^{\beta[\xi(r,\theta)+\xi_e(m)-\xi_e(l)]}]\} \\ &= \int\int_{outercell} \{E[e^{2\beta[\xi(r,\theta)+\xi'_e]}] - E^2[e^{\beta[\xi(r,\theta)+\xi'_e]}]\}\kappa\, dA, \end{aligned} \tag{3.17}$$

where $\kappa = (N_I + N_S)/(\pi R_O^2)$, and A is the outer cell area. For simplicity, we use the smallest distance rather than the smallest attenuation to determine the moment statistics of the random variable I_O/S_e.

Based on the above analysis of three kinds of interference, during a snapshot, the received bit energy-to-interference density ratio E_b/I_0 is given by

$$E_b/I_0 = \frac{W/B}{I_I/S_e + I_S/S_e + I_O/S_e + n_0/S_e}, \tag{3.18}$$

where W is the spread spectrum bandwidth and n_0 is the background noise. The uplink cell capacity C is the number of simultaneously transmitted packets that the system can support while satisfying the specified wireless transmission error requirement P^o_{BER}.

$$
\begin{aligned}
&P_r\{P_{BER} \geq P^o_{BER}\} \\
&= P_r\left\{I_I/S_e + I_S/S_e + I_O/S_e + n_0/S_e \geq \frac{W/B}{(E_b/I_0)^o}\right\} \\
&= \int_0^\infty \int_0^\infty Q\left(\frac{\frac{W/B}{(E_b/I_0)^o} - n_0/S_e - y - z - E[I_I/S_e]}{\sqrt{Var(I_I/S_e)}}\right) f_{\frac{I_S}{S_e}}(y) f_{\frac{I_O}{S_e}}(z)\, dydz \\
&\leq P^*_{out},
\end{aligned}
\tag{3.19}
$$

where P^*_{out} is the outage requirement and $(E_b/I_0)^o$ is associated with P^o_{BER}; $f_{I_S/S_e}(y)$ and $f_{I_O/S_e}(z)$ are the probability density functions of I_S/S_e and I_O/S_e, respectively, and both of them have Gaussian distributions with different means and variances; $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$.

In this uplink cell capacity derivation for wireless transmission, the uplink cell capacity is derived on the basis of persistent transmissions. User mobility is a decided influence on how much of the capacity can be utilized. And traffic characteristics, such as transmission rate, delay requirement and packet loss rate, are used to derive a call admission region which determines how many users with different types of traffic can be accommodated in the system.

3.2.1 Call admission region

For a practical packet switching network, resource allocation can be performed on a frame basis. The frame duration is determined so that each voice source generates one packet in each frame during a talkspurt. A video call has a variable packet arrival rate in each frame duration. Buffering at the MS smoothes out the burstiness of the traffic in each connection. Waiting packets with time-out value equal to one are referred to as the *most urgent packets* (MUPs), waiting packets with time-out value equal to two are referred to as the second MUPs, and so on. With the assumption of infinite buffer size, packet loss occurs only when the accumulated number of MUP exceeds the cell capacity. Let N_1, N_{2h}, and N_{2l} denote the number of ongoing voice, high- and low-quality video calls in a cell. The number of MUPs from each call depends on its data generation rate, the delay bound and the wireless transmission capacity. Let the integer random variable $M, M \in [0, M_{max}]$, denote the number of accumulated MUPs in packets/frame, and let M_{max} be the maximum value of M determined by the admitted number of calls in the cell and the packet generation rate of each traffic type. Let $\bar{r}_i$ denote the average number of generated data packets by user i over the course of a particular call. Conditioned on the MUP load in a frame, $M = m$, the number of lost packets for user i is denoted by $r_{loss_i|m}$. Thus, the actual packet loss rate for user i, P^i_{PLR}, is the average number of lost MUPs divided by the average number of generated packets and is given by

$$P^i_{PLR} = \frac{\sum_{m=C+1}^{M_{max}} P_r\{M = m\} r_{loss_i|m}}{\bar{r}_i}, \tag{3.20}$$

where $P\{M = m\}$ is the probability distribution function of M. Whatever M is, in each packet scheduling process, if $r_{loss_i|m}$ is chosen in such a way that the following relation is satisfied,

$$\frac{r_{loss_i|m}}{r_{loss_j|m}} = \frac{\bar{r}_i P^i_{PLR}}{\bar{r}_j P^j_{PLR}} \quad \forall i, \forall j \in \{1, \cdots, N_1 + N_{2h} + N_{2l}\}, \tag{3.21}$$

then when the required P^i_{PLR} is achieved for user i, the packet loss rate for all the other users can be satisfied. This approach may lead to the choice of a noninteger value for $r_{loss_i|m}$, which is inappropriate in wireless transmissions. Appropriate integer values may be chosen by assigning priority to traffic with stringent packet loss rate requirements. For instance, suppose that during a frame only two connections i and j have information to transmit and i has a more stringent packet loss rate requirement than traffic j. The total MUP load from both connections is m. The actual packet loss for i and j can be chosen to be $\tilde{r}_{loss_i|m} = \lfloor r_{loss_i|m} \rfloor$ and $\tilde{r}_{loss_j|m} = m - C - \tilde{r}_{loss_i|m}$, where $\lfloor \cdot \rfloor$ is a floor function. The packet loss rate requirement for j is satisfied by limiting the admitted number of calls as shown in the following call admission region derivation. Given that the total number of MUPs from all the users is m, the packet loss in a frame is constrained by

$$\sum_{i=1}^{N_1+N_{2h}+N_{2l}} \tilde{r}_{loss_i|m} = m - C \qquad m \geq C. \tag{3.22}$$

Combining (3.21) and (3.22), the packet losses for each user in a frame can be obtained.

A scheduling procedure, which aims at providing fair packet transmissions among voice, high-quality and low-quality video calls with the strictly bounded packet loss rate P_{PLR} for each individual call, can proceed as follows. At the beginning of each frame, all the waiting packets in the buffer have their time-out values decreased by one. Transmission requests from each connection including the number of waiting packets and the tagged time-out values are sent to the MSC via the serving BS. For instance, a call with a delay bound of D in frames can submit a transmission request as $(n_1, 1), (n_2, 2), \cdots, (n_D, D)$ representing the number of MUPs, the second MUPs, $\cdots$, and the least urgent packets. According to (3.21) and (3.22), the serving MSC determines the transmission permissions of MUPs. If there are still resources available after all the MUPs are scheduled for transmission, the remaining cell capacity is used to accommodate the less urgent packets. The same criterion is applied to the less urgent packets until all the waiting packets in the buffer are scheduled or there is no cell capacity left. The transmission permissions are passed via the serving BS's to each user for uplink packet transmissions.

A burst of data arrival in a frame can be divided into packets with fixed length. Because of the randomly varying data generation rate, the last packet may not be in full length for a video connection. This kind of packet is referred to as a partial packet. If a partial packet is filled with redundant bits and occupies a full channel for transmission, resources are wasted. To make efficient use of scare radio resources, a partial packet integration scheme is proposed to combine partial packets with different time-out values for transmission. Because of the identification header in each packet for right order and correct reassembly [21], it is possible to implement the packet integration at the sender and the disintegration at the receiver. When there is a partial MUP scheduled for transmission, the MSC will check if there is a second MUP transmission request so that it can be combined with the MUP for a possible transmission in the following frame. Fig. 3.2 demonstrates three cases of partial packet integration where 1 and 2 represent the time-out values of the consecutive waiting packets. The second MUPs are disassembled and reassembled with the partial MUP to form a new packet. In case (a), the partial MUP and the second MUP are combined into an exact full packet and occupy a full channel. Case (b) shows that after combining there is still a partial second MUP packet left, while in case (c), the combined packet is still a partial packet. In both cases (b) and (c), the same partial packet integration scheme can be applied to the third MUP and even less urgent packets to utilize limited radio resources efficiently. With the estimated uplink cell capacity and the proposed scheduling and partial packet integration scheme, a call admission region can be derived as a vector

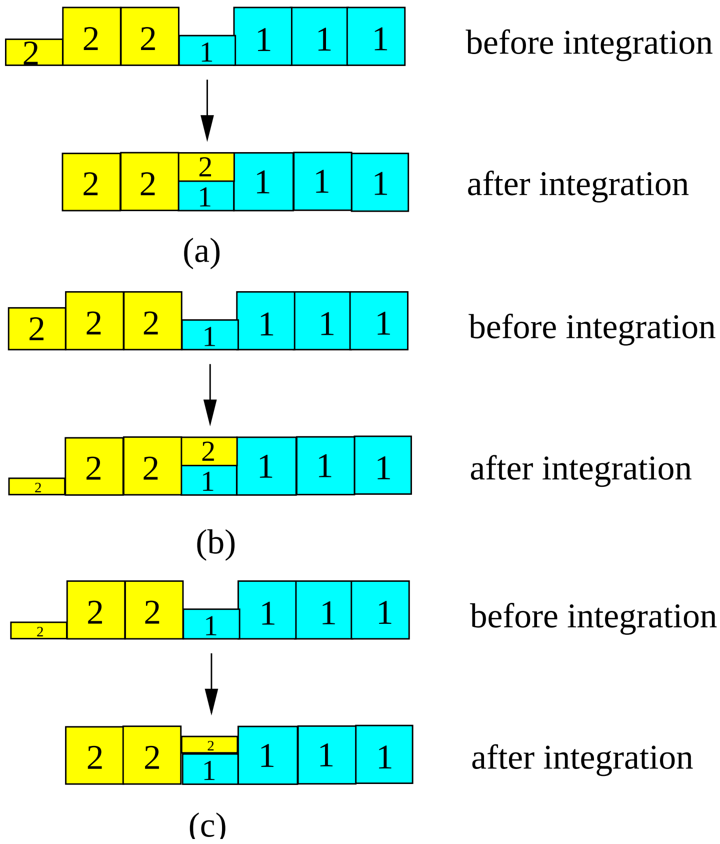

Figure 3.2: partial packet integration

(N_1, N_{2h}, N_{2l}) subject to the packet loss rate constraints for all users:

$$\begin{cases} Min(P^i_{PLR} - P^j_{PLR}) \quad \& \quad P^i_{PLR}, P^j_{PLR} \leq P^1_{PLR} \quad 1 \leq i, j \leq N_1 \\ Min(P^i_{PLR} - P^j_{PLR}) \quad \& \quad P^i_{PLR}, P^j_{PLR} \leq P^{2h}_{PLR} \quad 1 \leq i, j \leq N_{2h} \\ Min(P^i_{PLR} - P^j_{PLR}) \quad \& \quad P^i_{PLR}, P^j_{PLR} \leq P^{2l}_{PLR} \quad 1 \leq i, j \leq N_{2l} \end{cases} \tag{3.23}$$

The first term tries to minimize the difference of packet loss among users with the same traffic class, and the second term guarantees the strict packet loss rate requirement for each traffic type. A call admission region regulates the admitted number of voice, high-quality and low-quality video calls under the constraint that a call admission takes place only within the call admission region.

3.3 Summary

In this chapter, we first discuss the relation between the overall packet error rate, the bit error rate and the packet loss rate. The bit error rate requirement is satisfied by arranging simultaneously transmitted packets, which determines the uplink cell capacity. The packet loss rate is satisfied by properly scheduling packet transmissions from different traffic sources. With the proposed packet scheduling and partial packet integration scheme, a call admission region is derived which is used by the MSC to make resource reservation, call admission control and so on.

Chapter 4

An Adaptive Resource Reservation Strategy for Handoff

Soft handoff takes place when an MS is in the intersection of the coverage area of two or more BS's. In order to ensure the predefined handoff call dropping probability, some reservation resources are required before handoff. User mobility and traffic variations cause handoff attempts to change, which makes a fixed amount of resource reservation inappropriate. A reservation scheme which adapts to variations of handoff attempts is required. In the following, we first give a brief overview of the soft handoff signaling process and its related extra signaling overhead; after that we present in detail the proposed adaptive resource reservation scheme.

4.1 Soft handoff signaling

An MSC is solely responsible for the execution of soft handoff. Intra-MSC handoff is handled by the serving MSC locally without intervention of other system elements. Here we focus on the inter-MSC soft handoff signaling process.

Two phases comprise the inter-MSC soft handoff: add and drop. Fig. 4.1(b) and (c) shows the message flows in the add and drop phases. The process for add and drop is shown as pseudocode 1.

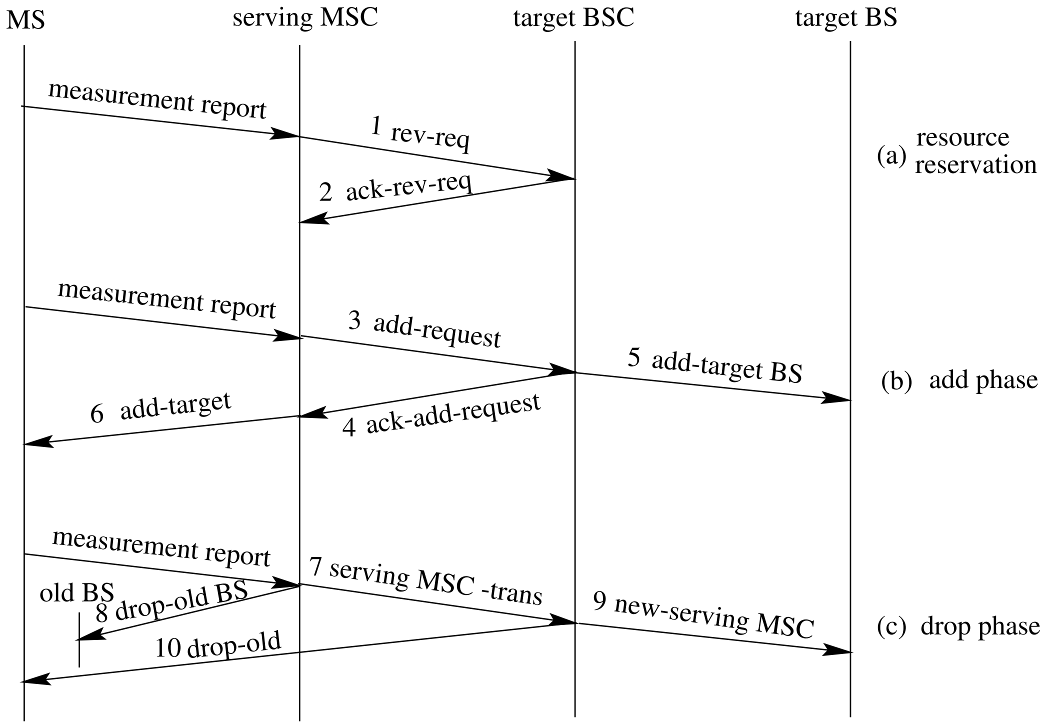

Figure 4.1: Soft handoff signaling

Pseudocode 1: Add and drop process in soft handoff

1 Each MS searches and averages the RPSS measurements;

2 RPSS measurements are reported;

3 If (a pilot signal is detected not associated with any forward traffic channel) && (RPSS measurement > ADD threshold) /* add process */

- **3.1** The serving MSC sends *add-request* message to the new target MSC;
- **3.2** The new MSC performs resource allocation;
- **3.3** If (successful resource allocation)
  - **3.3.1** Acknowledge the serving MSC with *ack-add-request*;
  - **3.3.2** The new MSC sends an *add-target BS* message to the new BS;
  - **3.3.3** The serving MSC directs the MS to add the new BS into its active set with *add-target*;

4 If (a pilot signal is detected with RPSS > DROP threshold) /* drop process */

- **4.1** The serving MSC sends a *serving MSC-trans* to the new MSC;
- **4.2** The serving MSC asks for dropping of the old BS with message *drop-old BS*;
- **4.3** The new MSC informs the new BS that it will act as a serving MSC with message *new-serving MSC*;

4.4 The new MSC sends a *drop-old* message to direct the MS to drop the old BS from its active set.

By introducing an adaptive resource reservation scheme to the system, some extra signaling overhead will be incurred. Resources must be reserved before a handoff takes place. The extra signaling is shown in Fig. 4.1 (a), and the process is shown as pseudocode 2.

Pseudocode 2: Resource reservation

1 Each MS searches and averages the RPSS measurements;

2 RPSS measurements report;

3 If (the estimated time interval during which a handoff will occur $< T_{rev}$) /* resource reservation */

3.1 The serving MSC sends a *rev-req* message to the new target MSC to request resource reservation;

3.2 The target MSC performs resource reservation;

3.3 The target MSC acknowledges the serving MSC with *ack-rev-req*, indicating reservation success or failure.

A successful reservation can avoid excessive reservation requests from the same MS, while a failed one indicates that reservation request can still be sent to the target MSC in the next time interval. According to the signaling process in Fig. 4.1, the extra signaling load for resource reservation is relatively light.

4.2 Adaptive resource reservation scheme

In this section, we first study user mobility adaptation, and then traffic load adaptation for resource reservation.

4.2.1 User mobility adaptation

Each MS samples the RPSS from both the serving and the neighboring BS's every Δt s. Since an instantaneous sampled RPSS is affected by many factors, such as fading and measurement noise, it cannot exactly reflect the point at which resource reservation should be triggered. Exponential averaging [56] is applied to smooth out the uncertainty by its virtue of arithmetic simplicity and relaxed memory requirements. The exponential averaging process is carried out upon each successive measurement, so the smoothed function $S_{k,i}$ of the measurements is

$$S_{k,i} = \beta_{k,i} S_{k-1,i} + (1 - \beta_{k,i}) p_{k,i}, \tag{4.1}$$

where $\beta_{k,i}$ is a smoothing factor representing the weighting value given to the previously estimated RPSS, which is an important parameter in estimating the present RPSS. With high mobile velocity, the path loss varies significantly, which indicates that the previous RPSS should have less effect on the current RPSS estimation. On the other hand, the path loss variations may be negligible with low mobile velocity, which indicates a strong correlation between the previous RPSS and the present RPSS values. The shadow fading correlation coefficient $\varrho_{k,i}$ also plays an important role on the smoothing factor $\beta_{k,i}$. With a strong correlation effect, the value for $\beta_{k,i}$ should be close to 1, while a weak correlation effect of shadow fading implies a small $\beta_{k,i}$ value. Therefore, $\beta_{k,i}$ reflects the correlation between the previous and the present measurements. A consecutive sequence of L recent RPSSs $p_{k-j,i}, j = 0, 1, \ldots, L-1$, are used to estimate the smoothing factor $\beta_{k,i}$. Let $\bar{p}_{k,i}$ be the mean RPSS over L samples, and let $\bar{\sigma}^2_{k,i}$ and $Cov_{k,i}$ be the sample variance and covariance, respectively. Then the smoothing factor $\beta_{k,i}$ is estimated as $\beta_{k,i} = Cov_{k,i}/\bar{\sigma}^2_{k,i}$, where

$$
\begin{aligned}
Cov_{k,i} &= \frac{1}{L-1}\sum_{j=0}^{L-2}(p_{k-j,i} - \bar{p}_{k,i})(p_{k-j-1,i} - \bar{p}_{k,i}), \\
\bar{\sigma}^2_{k,i} &= \frac{1}{L}\sum_{j=0}^{L-1}(p_{k-j,i} - \bar{p}_{k,i})^2, \\
\bar{p}_{k,i} &= \frac{1}{L}\sum_{j=0}^{L-1} p_{k-j,i}.
\end{aligned}
$$

Except for its serving BS, the BS with the most significant $S_{k,i}$ is expected to be the target BS the mobile user is approaching. To make efficient use of radio resources, resources are only reserved when a potential handoff call triggers a resource reservation request in the target BS. Rather than a specific reservation in which a reserved channel can only be used by the user who reserves it, the reserved resources form a common reservation pool for all the incoming handoff calls. A handoff call dropping event occurs when an incoming handoff call cannot find a spare channel. There are two scenarios for handoff dropping; first, all the reserved resources may be in use by some other handoff calls and the target cell may be unable provide the incoming handoff call with the necessary resources; secondly, the traveling mobile may have already entered the target cell, but the target cell may not have spare resources to admit the handoff call – at the same time, the mobile can no longer communicate with the serving BS. On the other hand, when a user moves slowly and the traffic load is light, there is more chance that the requested resources will be available when a handoff takes place. So a time threshold, which reflects dynamic user mobility rather than a RPSS threshold, plays an important role in resource reservation and handoff prioritization. Fig. 4.2 illustrates the proposed reservation-request time threshold with different user mobility. S_{rev} is the difference in RPSS between the smoothed RPSS from the serving BS, $S_{k,i}$, and that from the potential target BS, $S_{k,j}$, where i and j denote the serving and the target BS's when a reservation request is initiated. In the same communication environment, given the reservation-request time threshold T_{rev}, the value

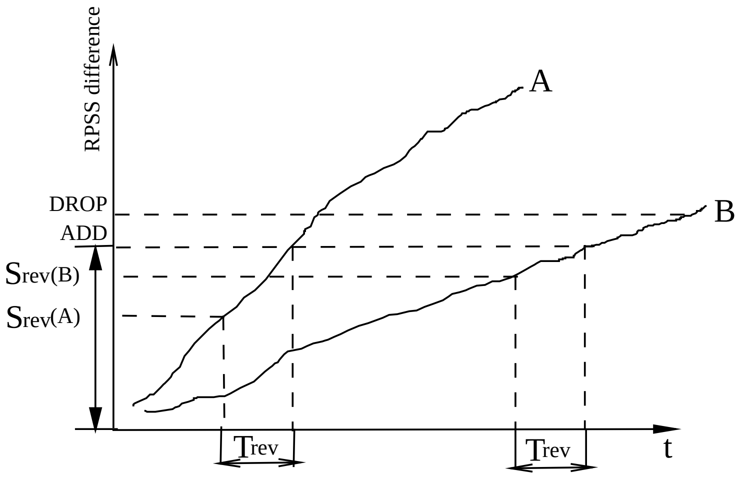

Figure 4.2: RPSS difference vs. T_{rev}

of S_{rev} varies according to user movement. A user moving quickly away from its serving BS, such as user A in Fig. 4.2, will initiate resource reservation request with a much lower S_{rev} value compared to the slowly moving user (B in Fig. 4.2) to make the reserved resources available when a handoff is actually requested. Because of the randomness in user mobility and shadow fading effects, the RPSS difference varies nonlinearly with time. Since making a reservation too early will waste resources and potentially increase new call blocking, while late reservation will cause handoff call dropping, close prediction of the reservation-request time is desired. As a compromise between prediction accuracy and computational complexity, we propose to use the following second order prediction,

$$S_t = S_k + v_{S,k}t + \frac{1}{2}a_{S,k}t^2, \tag{4.2}$$

where S_t is the ADD threshold beyond which a soft handoff occurs and the MS takes advantage of the better link quality of two simultaneously communicating BS's to maintain minimum interference; S_k is the RPSS difference at time t_k; $v_{S,k} = (S_k - S_{k-1})/\Delta t$ reflects the changes in the RPSS and $a_{S,k} = (S_k - 2S_{k-1} + S_{k-2})/\Delta t$ captures the variation of the changes in RPSS; t is the expected time interval during which a handoff will occur. Whenever the predicted time interval $t \leq T_{rev}$, a reservation request is triggered. The reserved resources are tagged as a common reservation pool which is exclusively used by handoff calls.

Compared to the fixed channel reservation-request threshold based on RPSS values in [38], the second order formulation of RPSS difference in the prediction of reservation-request time captures the mobility variations. A common reservation pool reduces the risk of false reservation since the available reserved channels can be used for fast-moving users that fail to reserve resources before handoff. Also, it can accommodate mobility and fading variations in which the early reserved resources can be used for the actual early handoffs rather than for the mobile user who reserves the resources but hands off later due to speed decrease or moving direction changes. Thus, the pooling mechanism can achieve resource sharing which provides multiplexing gain to the resource utilization.

4.2.2 Traffic load adaptation

In the previous discussion, T_{rev} is assumed to be known for each MSC to make comparisons for reservation decisions. In the remainder of this subsection, we investigate how T_{rev} is adaptively determined for the varying traffic load. T_{rev} is the allowed time duration for the target BS to get the requested resources ready before a handoff occurs. It determines the number of reserved channels which in turn affect the predefined GoS. A successful channel reservation is achieved either when there is a spare channel at the first time a reservation request is triggered or when there is a channel release before the handoff occurs. The probability of a successful channel reservation, P_{succ}, is expressed as:

$$\begin{aligned} P_{succ} = &\, P_r\{\text{a spare channel at request}\} \\ &+ P_r\{\text{no spare channel at request}\} \times P_r\{\text{ channel holding time} < T_{rev}\}. \end{aligned} \tag{4.3}$$

Intuitively, with the same user mobility, the higher the traffic load in the system, the higher the channel occupancy, which lowers the probability of a successful reservation at request time. To compensate for this, a larger T_{rev} is needed to increase the probability of a channel release before a handoff (the second term in (4.3)). Therefore, it is expected that heavy traffic load will prompt a larger reservation-request time threshold T_{rev}, while light traffic load will lead to a smaller T_{rev}. In addition to gradual longer-term traffic load variations, for example, traffic load and traffic flow during morning rush hour and during daytime are different, and new call and handoff call arrival events are also bursty and dynamic in a specified period. The reserved resources, and accordingly the new call blocking and handoff call dropping rates are functions of T_{rev}. To use a fixed T_{rev} value is not appropriate for such a dynamic system to achieve minimum GoS performance at any time epoch. Therefore, the resource reservation-request time threshold T_{rev} should be designed to be adaptive both to the traffic variations and to user mobility.

To closely track these two varying system parameters, short-term GoS measures in terms of new call blocking rate R_b and handoff call dropping rate R_d are used. Because the traffic load varies with time, derivation of the new call blocking and handoff call dropping rates using a fixed time duration T is not appropriate. For instance, in a low user-mobility and low traffic-load communications environment, T can take a large value before a call event happens, which includes a new call arrival, a handoff call arrival, a new call blocking or a handoff call dropping; while at high user mobility or heavy traffic load situation, T may take a rather smaller value to capture the variations of these four call events. Here we use a time-varying value T for the averaging window size within which a prescribed number of N handoff call arrivals occur. The reason to use the handoff call arrival event is that the number of handoff call arrivals indicates not only the traffic variations, but the user mobility variations. Four buffers are introduced for each BS to temporarily store the numbers of call events of four types. Each buffer is updated after every time interval during which a number of N handoff call arrival events have happened. At the k-th time interval

t_k, the new call blocking rate $\tilde{R}_b(k)$ and handoff call dropping rate $\tilde{R}_d(k)$ are obtained from the newly accumulated call event counts:

$$\begin{cases} \tilde{R}_b(k) &= \dfrac{\text{number of new call blocking events}}{\text{number of new call arrival events}}, \\ \tilde{R}_d(k) &= \dfrac{\text{number of handoff call dropping events}}{\text{number of handoff call arrival events (N)}}. \end{cases} \tag{4.4}$$

We introduce a sliding window averaging to smooth out the variation in the newly obtained instantaneous new call blocking and handoff call dropping rates as shown in Fig. 4.3. Let x be the averaging window size for the historical effect and y be the averaging window size

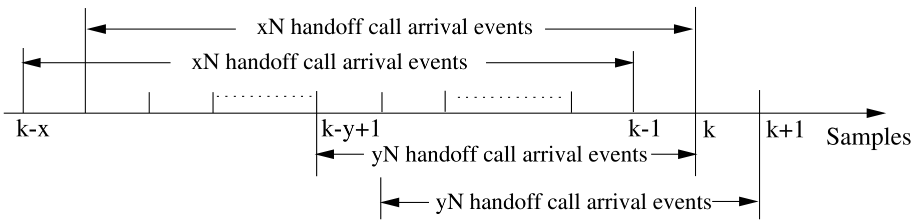

Figure 4.3: New call blocking and handoff call dropping rates derivation

for the recent effect where x and y satisfy the relation $x \geq y$. So, the smoothed functions $R_b(k)$ and $R_d(k)$ take the form of

$$\begin{cases} R_b(k) = (1-\rho)\dfrac{\sum_{i=k-x}^{k-1} \tilde{R}_b(i)}{x} + \rho\dfrac{\sum_{j=k-y+1}^{k} \tilde{R}_b(j)}{y}, \\ R_d(k) = (1-\rho)\dfrac{\sum_{i=k-x}^{k-1} \tilde{R}_d(i)}{x} + \rho\dfrac{\sum_{j=k-y+1}^{k} \tilde{R}_d(j)}{y}, \end{cases} \tag{4.5}$$

where ρ is a smoothing factor. By giving a greater weighting value to the recently obtained y new call blocking and handoff call dropping rates, the new call blocking and handoff call dropping events can be closely followed for T_{rev} adaptation. As the new call blocking and handoff call dropping rates are functions of the reservation-request time threshold T_{rev}, the GoS measure can be expressed as:

$$GoS = \frac{R_b(T_{rev})}{R_d(T_{rev})} = \gamma. \tag{4.6}$$

In the simulation, the ratio of the new call blocking rate and handoff call dropping rate is kept at γ to achieve the desired GoS performance.

As the ratio of the new call blocking rate R_b and the handoff call dropping rate R_d is the driving force that steers the adaptive adjustment of the T_{rev} value, the following three conditions are used to deduce a new reservation-request time threshold denoted as $T_{rev}(k+1)$ based on the ratio of the new call blocking and handoff call dropping rates $R_b(k)/R_d(k)$, and the old $T_{rev}(k)$ value.

1. When $R_b(k) \approx R_d(\text{k})$, the old $T_{rev}(k)$ value is implied to be too small and the handoff calls are not properly prioritized so that the handoff call dropping rate reaches as high as the new call blocking rate. In this case, a large $T_{rev}(k+1)$ value is required to reserve more resources for handoff calls. We increase the $T_{rev}(k)$ value by c times, so $T_{rev}(k+1) = cT_{rev}(k)$.

2. When $R_b(k)/R_d(k) \to \infty$, the old $T_{rev}(k)$ value is implied to be too large and excessive resource reservation has made the new call blocking rate and the handoff call dropping

rate far away from the objective ratio γ. In this case, we force $T_{rev}(k+1) = 0$ to immediately reduce the coming resource reservation requests to make efficient utilization of radio resources.

3. When $R_b(k)/R_d(k) = \gamma$, the old $T_{rev}(k)$ value is implied to be appropriately set and the minimum GoS performance measure is achieved. In this case, we set $T_{rev}(k+1) = T_{rev}(k)$.

To satisfy these three boundary conditions, T_{rev} can be periodically updated by

$$\begin{aligned} & T_{rev}(k+1) = cT_{rev}(k)exp(-\frac{1-x}{1-\gamma}\ln c), \\ \text{where} \qquad & x = \frac{R_b(k)}{R_d(k)}, \qquad 0 \leq T_{rev} \leq T_{max}. \end{aligned} \tag{4.7}$$

T_{max} is limited by the cell size and user mobility information.

The effect of different c values on the reservation-request time threshold update curve is shown in Fig.4.4. A larger c leads to fast convergence of R_b/R_d to the objective requirement γ, but a small deviation of R_b/R_d from γ will cause frequent vibration of R_b/R_d. On the other hand, a small c results in a more smooth and steady R_b/R_d value though the convergence speed to γ is relatively slow. Since the traffic load changes slowly with time, and call events burstiness are alleviated by the proposed sliding window average scheme, it is expected that a small c will result in better GoS performance measurements.

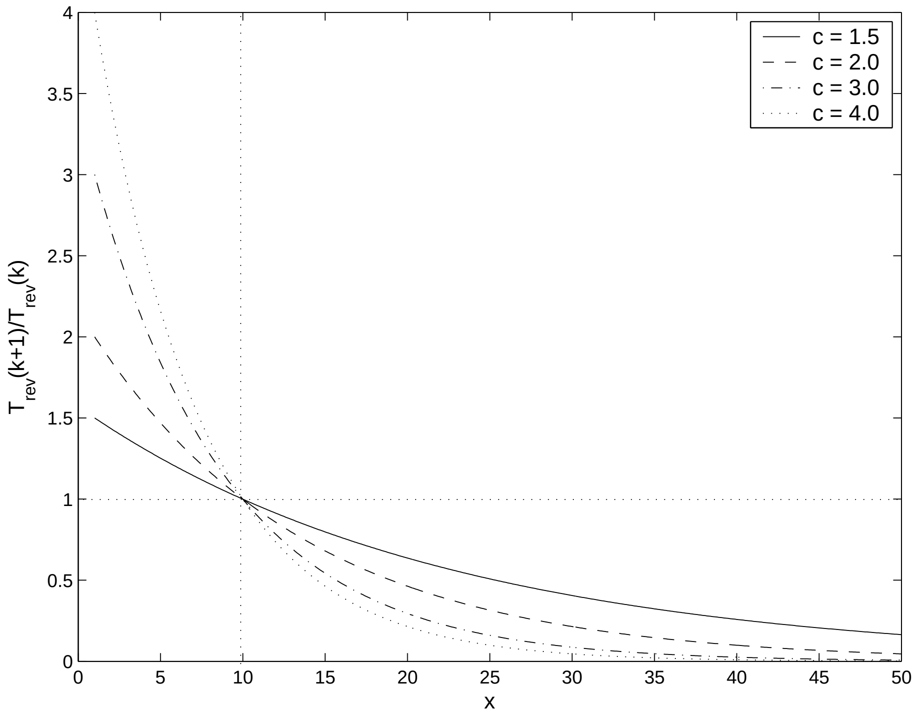

Figure 4.4: Reservation-request time threshold update curve

4.3 Summary

In this chapter, we have presented an adaptive resource reservation strategy for handoff priority. In summary, the measurement-based adaptive resource reservation scheme applies a user assisted handoff algorithm. The resource reservation decisions are exclusively based on the RPSS information fed back from each mobile user. There is no fixed number of reserved resources for handoff calls and each call makes reservation requests before handoff happens. The reserved resources are kept in a pool for all the incoming handoff calls rather than for a specific user in order to improve resource utilization. This can also reduce signaling cost because it makes cancellation of false reservations unnecessary. By adjusting the reservation-request time threshold with traffic variation, the required minimal GoS can be achieved. Based on the above analysis, by introducing the proposed signaling overhead, the remaining time prediction, and the reservation-request time threshold update in the resource reservation phase, the proposed adaptive resource reservation scheme can be well fitted into the current wireless CDMA networks without introducing excessive computational complexity.

Chapter 5

An Adaptive Handoff Priority Scheme Supporting Multimedia Applications

The role of CAC at the call level is to make an admission decision for a new or handoff call upon detecting an admission request from the request access mini-slot in an uplink frame. The serving BS forwards the admission request to the MSC, which determines whether to admit or reject the call request. The admission decision is then sent back from the MSC to the MS via the serving BS in the request acknowledgment mini-slot in the following downlink frame. The BS also notifies the MS of the allocated transmission rate via the transmission permission mini-slot if the MS has a call with variable rate traffic in progress. A dedicated coding channel and the corresponding receiver are allocated to each admitted call after the admission and released when the call is completed or handed off to a new cell. The call admission region limits the admitted number of calls of all classes in the system. The proposed packet scheduling scheme and call admission region guarantee that once a call is admitted, it can be provided with satisfactory delay and overall packet error rate performance throughout the lifetime of the connection. The admission decision should also ensure low handoff call dropping probability. In Chapter 4, we have proposed an adaptive resource reservation scheme where the reserved resources are used exclusively for handoff calls for handoff prioritization. If a handoff call can neither find the required reserved resources nor spare resources at arrival, while at the same time the connection quality to the old BS is no longer acceptable, the handoff call will have to be dropped. In order to further prioritize handoff calls for user satisfaction, in this chapter we propose another adaptive handoff priority scheme for wireless MC-CDMA cellular networks which can effectively support multimedia applications. The scheme utilizes the characteristics of multimedia traffic to temporarily degrade service quality when available resources are insufficient for handoff call admission, and gain back the lost quality whenever there is resource release in the system. Therefore, by adaptively transforming multimedia calls between high- and low-quality modes at heavy traffic load environments, handoff call admission priority and handoff call dropping probability can be further improved.

5.1 Adaptive handoff priority scheme

The proposed adaptive handoff priority scheme prioritizes handoff calls by reducing the transmission rate of ongoing high-quality video calls and admitting handoff calls with the gained resources. When a call admission request, which is either a voice or a video call which is served with either high- or low-quality, arrives, a call admission decision is made on the basis of the derived call admission region. For new and handoff calls, different call admission controls are applied. New voice and video calls are accepted only when the transmission rate and the overall packet error rate can be satisfied. On the other hand, if a voice handoff call arrives, the MSC will first check to see if the spare resource can be used to accept the handoff voice call. When the spare resource cannot support a satisfactory voice handoff call or there is no spare resource available, and if there is a rate-adaptive high-quality video call in progress, rate renegotiation will be triggered. After both correspondents agree to reduce the transmission rate, the MSC will perform call admission control again. Service degradation of more than one rate-adaptive high-quality video call may be necessary to admit a handoff call according to the traffic characteristics. A voice handoff call can be accepted if the overall packet error rate requirements of the handoff call and of all other ongoing calls in the system can be achieved; otherwise, it is dropped. Fig. 5.1 shows the voice handoff call admission control flow diagram.

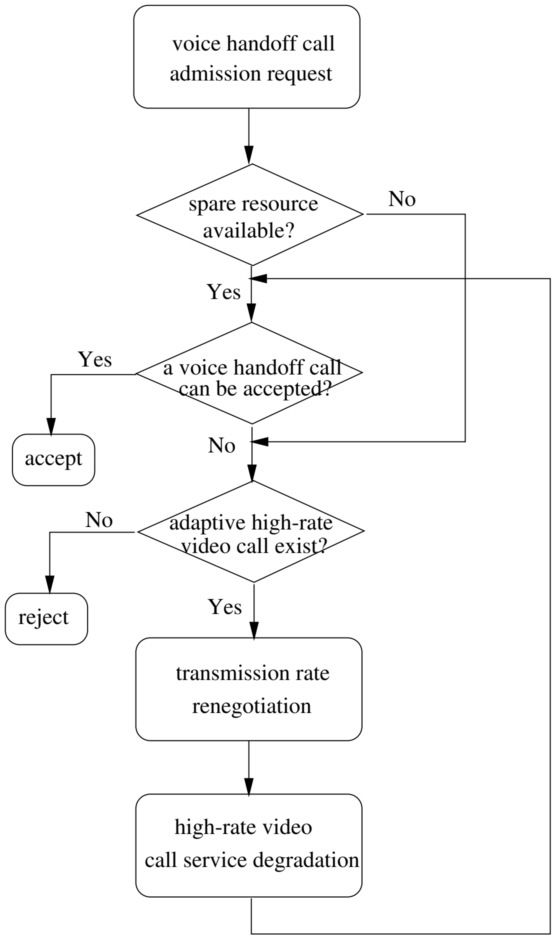

Figure 5.1: Voice handoff call admission control flow diagram

Because there are two types of video calls, high-quality and low-quality, in the system, handoff and call admission control schemes for video calls differ from that for voice calls. Fig. 5.2 shows the video handoff call admission control flowchat. No matter whether the incoming handoff video call is served with high or low quality, the MSC will first try to provide the handoff video call with high quality service with the spare resource. If the available resource is not enough, then low quality service will be provided. When the spare resource can neither provide high-quality nor low-quality service for the video handoff call, or when traffic is heavy and no spare resource is available, the same rate adaptation process is executed. In this way, the admitted video handoff call is served with a low transmission rate. It is seen in Fig. 5.2 that the resource gained from rate adaptation is only used to support a low-quality video handoff call. There are four possible service qualities for a video call before and after handoff. A high-quality video call can be provided with either high or low quality service after a successful handoff, and the same holds for a low-quality video handoff attempt depending on the traffic load and resource utilization in the new target cell.

In summary, let the number of ongoing calls of all classes in the cell be $(N_1, N_{2h}; N_{2l})$, a voice or video handoff is successfully executed if and only if

$$(N_1 + 1, N_{2h} - n_2, N_{2l} + n_2), \qquad N_{2h} \geq n_2 \tag{5.1}$$

for a voice handoff call or

$$\begin{cases} (N_1, N_{2h} + 1, N_{2l}), & or \\ (N_1, N_{2h} - n_2, N_{2l} + n_2 + 1), & N_{2h} \geq n_2 \end{cases} \tag{5.2}$$

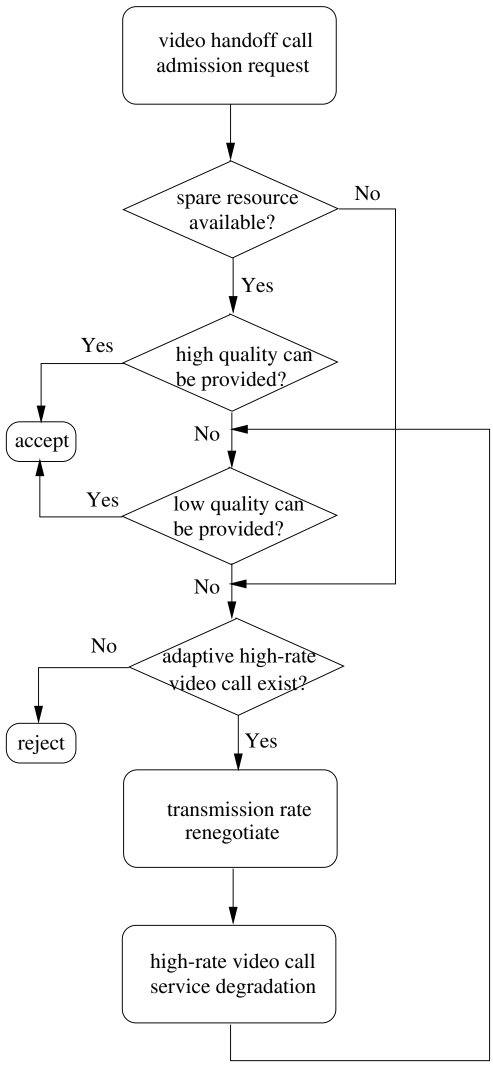

Figure 5.2: Video handoff call admission control flow diagram

for a video handoff call is still within the derived call admission region. Here $n_2 \geq 0$ is the total number of rate-adaptive high-quality video calls that have to suffer quality degradation. $n_2 = 0$ indicates that there is no service degradation for high-quality video calls for handoff call priority. Equation (5.2) also indicates that either high or low quality of service can be provided to a video handoff call, and only low quality of service can be provided if service degradation is needed.

From both the user's and the service provider's perspective, the ratio of the low-quality video calls over the total number of video calls in the system should be kept as low as possible for better resource utilization and user satisfaction. Therefore, whenever there is resource release due to call termination or call handoff to a neighboring cell, a low-quality video call will try to improve its transmission rate. Fig. 5.3 shows the flowchat of transmission rate adjustment of low-quality video calls at resource release. If the released

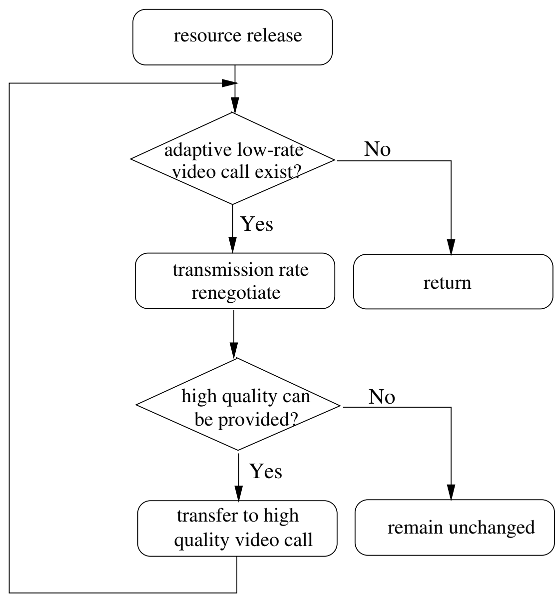

Figure 5.3: Low-quality video call transmission rate adjustment

resource is not sufficient for a low-quality video call to be provided with high quality service, the video call remains with low quality service till next time resource release happens.

5.2 Summary

In this chapter, we propose an adaptive handoff priority scheme supporting multimedia applications. To give handoff calls higher call admission priority, if there is no resource available on handoff call arrival, the rate-adaptive high-quality video calls can be temporarily converted to low-quality video calls to gain some resource to admit the incoming handoff call. The low-quality video call will be provided with high quality service whenever the released resource is sufficient to satisfy user satisfaction.

Chapter 6

Numerical Results

6.1 Simulation model

Consider a conventional hexagonal cellular network with a cluster of 7 cells. The spread spectrum bandwidth is 5 MHz. The cell radius is 300 m and the relative soft handoff area is chosen to be $\alpha = 0.4$; i.e., the soft handoff area corresponds to 36% of a normal cell area [41]. The basic transmission rate B is 16 kbps for one single code channel. The path loss exponent η is set to 4 and the standard deviation for the shadowing effect is $\sigma = 8$ dB. The shadow fading correlation distance $\bar{d}_i$ is set to 20 m for all the cells. The standard deviation due to imperfect power control is set to $\sigma_e = 2$ dB. Because CDMA is an interference dominant system, we ignore background noise for simplicity. Without consideration of bit error rate imposed from the wireline domain, the overall packet error rate requirements for voice and low-rate video (for applications such as video conferencing) communications are 10^{-3} and 10^{-4}, respectively. The upper bound of the outage probability P^*_{out} is set to 10^{-2}. Each frame is 10 ms in length which is equivalent to $L_{en} = 160$ bits. Call holding time is assumed exponentially distributed with a mean of 120 s.

The voice traffic is simulated by the ON-OFF model. During the ON state, one packet is generated in each frame. On average, an ON state lasts 0.35 s and an OFF state lasts 0.65 s. The video traffic has a variable transmission rate which is simulated by 20 minisources with the average ON and OFF durations of 0.2 s and 0.8 s. During the ON state of a minisource, the transmission rates are 8 kbps and 4 kbps for high- and low-quality video calls, respectively. The delay bound for voice and video packets are 2 and 3 frames, respectively. The simulation parameters are summarized in Table. 6.1. The simulation model is shown in Fig. 6.1. The mobility of each ongoing call in a cell is updated every Δt s according to the mobility model described in the next section. New call arrivals follow a Poisson process. Each MS periodically reports its RPSS measurements. When the measured RPSS difference from the serving and the potential target BSs is within the ADD and DROP thresholds, the call is in the soft handoff region. A resource allocation and call admission control process is triggered if a voice or video call just enters the handoff region as discussed in Chapter 5. When the incoming handoff call cannot be provided

Table 6.1: Simulation parameters

Parameters		Value
Spread spectrum bandwidth W		5 MHz
Basic transmission rate B		16 kbps
Path loss exponent η		4
Lognormal shadowing ζ^*		$\zeta^* - N(0, 8 \text{ dB})$
Imperfect power control ξ_e		$\xi_e - N(0, 2 \text{ dB})$
Frame length		10 ms (160 bits)
Outage probability P^*_{out}		10^{-2}
Voice traffic:	Average talk spurt duration	0.35 s
	Average silent period	0.65 s
	Tolerable time-out value in frames	2
	Packet error rate requirement	10^{-3}
Video traffic:	Number of minisources N	20
	ON duration	0.2 s
	OFF period	0.8 s
	Transmission rates during ON state	8 kbps, 4 kbps
	Tolerable time-out value in frames	3
	Packet error rate requirement	10^{-4}

with the required resources, a handoff call dropping event occurs. If the call is not in soft handoff, the remaining time to handoff is predicted and compared to the reservation-request time threshold to determine whether this call needs resource reservation in its potential target cell. The reservation-request time threshold is periodically adjusted on the basis of the estimated new call blocking and handoff call dropping rate and the old threshold value. Whenever a resource reservation request is formed, the potential target cell performs the resource reservation and updates its reservation pool if the process is successfully implemented. All the modules in Fig. 6.1 have been studied except the user mobility

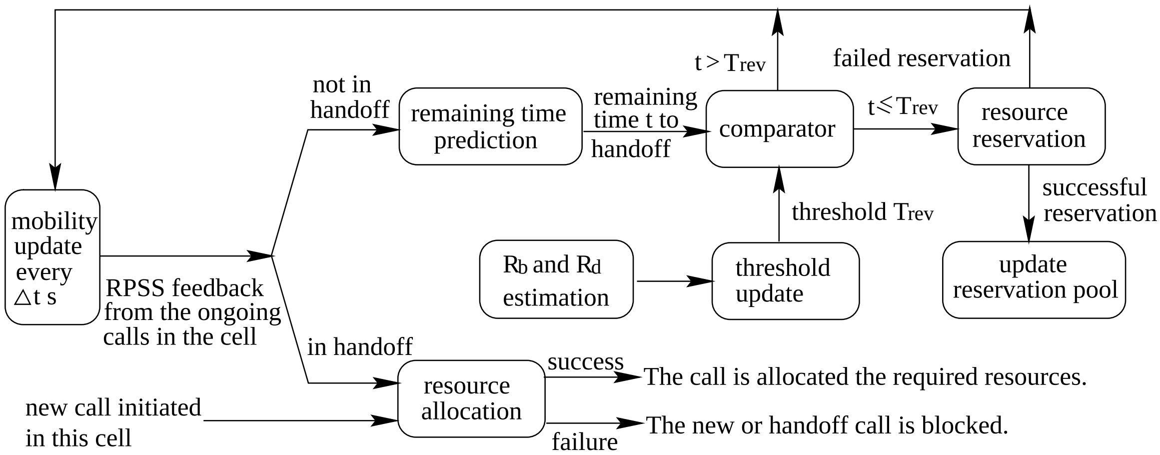

Figure 6.1: Simulation flowchat

update process. In the following section, a detailed mobility model which characterizes a practical moving pattern is presented.

6.2 Mobility model

There are two kinds of mobile users in the system: (1) pedestrians who are assumed to have a fixed velocity of 1 m/s with uniformly distributed direction in $[0, 2\pi)$, which indicates a T_{max} value of 240 s; (2) mobile users in moving vehicles characterized by the following mobility model.

A mobility model should mimic human movement behavior. We model user mobility using three parameters: location, velocity and acceleration, on a two dimensional plane. For velocity and acceleration, a two dimensional vector determines both the magnitude and direction. The mobile speed varies because of acceleration or deceleration. For description simplicity, we use the term acceleration with both positive and negative values to represent acceleration and deceleration. Therefore, the following model focuses on the mobile acceleration establishment. Assume a user's mobility is updated at discrete time instances $t_k = t_0 + k\Delta t$, where t_0 is the initial time epoch, Δt is the sampling interval and k is an integer. In reality, a mobile user may have unexpected changes in acceleration caused by traffic lights or road conditions; on the other hand, the acceleration is highly correlated, i.e., if a moving user is accelerating at time t_k, it is likely to continue accelerating at time t_{k+1} till it reaches the speed limit. In order to take these two effects into consideration, following [48], we model a user's mobility as a dynamic system driven by an unexpected acceleration $\mathbf{u}_k = [u_{x,k}, u_{y,k}]^T$ and a correlated acceleration $\mathbf{r}_k = [r_{x,k}, r_{y,k}]^T$ at time t_k, as shown in Fig. 6.2. $\mathbf{u}_k$ is modelled as a Markov process with a finite number of "states", $Q_1, Q_2, \ldots, Q_m$, as possible discrete levels of acceleration. The transition probability $\theta_{i,j} = P\{\mathbf{u}_k = Q_j | \mathbf{u}_{k-1} = Q_i\}$ can be approximated by a value ϵ near unity for $i = j$, and $(1-\epsilon)/(m-1)$ for $i \neq j$ [48]. The correlated acceleration can be modelled as a zero mean Gaussian random variable with a variance that is chosen to cover the "gap" between adjacent acceleration states. To represent the correlation feature of $\mathbf{r_k}$, a first order AR model is used, $\mathbf{r}_{k+1} = \alpha\mathbf{r}_k + \mathbf{w}_k$, where $\mathbf{w}_k$ is a zero mean white Gaussian vector with variance $\sigma_{\mathbf{w}}^2$ in one dimension.

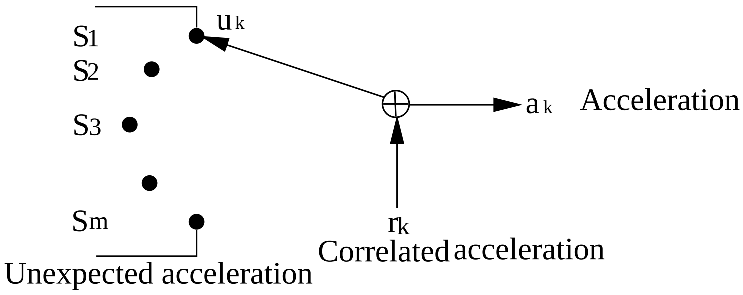

Figure 6.2: Acceleration model

Denote x_k and y_k as the horizontal and vertical coordinates of a mobile user's random location at time t_k, and $v_{x,k}$ and $v_{y,k}$ as the corresponding velocities. The user mobility vector denoting user location, velocity and acceleration can be expressed as

$$\mathbf{\Psi}_k = [x_k, v_{x,k}, r_{x,k}, y_k, v_{y,k}, r_{y,k}]^T. \tag{6.1}$$

Combining the driving model in Fig. 6.2 with the second order motion model, the discrete time dynamic equations in the X dimension can be expressed as:

$$\begin{cases} x_{k+1} = x_k + v_{x,k}\Delta t + (r_{x,k} + u_{x,k})\frac{\Delta t^2}{2} \\ v_{x,k+1} = v_{x,k} + (r_{x,k} + u_{x,k})\Delta t \\ r_{x,k+1} = \alpha r_{x,k} + w_{x,k} \end{cases} \tag{6.2}$$

The discrete time dynamic equations in the Y dimension have the same expressions as (6.2) with x replaced by y. Rewriting these equations in a more compact matrix form, we have

$$\mathbf{\Psi}_{k+1} = A\mathbf{\Psi}_k + B\mathbf{u}_k + C\mathbf{w}_k, \tag{6.3}$$

where

$$A = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} & 0 & 0 & 0 \\ 0 & 1 & \Delta t & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & \alpha \end{bmatrix}, B = \begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ \Delta t & 0 \\ 0 & 0 \\ 0 & \frac{\Delta t^2}{2} \\ 0 & \Delta t \\ 0 & 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Five levels, $(-4, -2, 0, 2, 4)$ m/s^2, are selected as unexpected accelerations. The probability of having the same unexpected acceleration in the next time interval, ϵ, is 0.9, so that the probability to change to any one of the other levels is 0.025. The correlation coefficient of the correlated acceleration α is chosen to be 0.6 and the variance of $\mathbf{w}$ is set to 0.72. The speed limit is set to 17 m/s which is reasonable in an urban area. Considering computational complexity, practical service continuity requirements and the fading property, we choose $\Delta t = 1$ s for the mobility update and RPSS averaging [19].

6.3 Performance evaluation

In this section, we provide some numerical simulation results on the proposed adaptive resource reservation scheme and the proposed handoff priority scheme supporting multimedia applications. A discrete-event-driven simulation is performed for performance evaluations. Comparisons are also given to demonstrate the performance improvement of the proposed schemes over other schemes in literature.

Table 6.2: Call admission region with circuit switching transmission ($P^*_{WTE} = 10^{-4}$)

	N_{2l}									
N_{2h}	0	1	2	3	4	5	6	7	8	9
0	38	35	31	28	25	21	17	14	11	7
1	32	29	25	21	18	14	11	8	3	
2	26	23	19	15	12	8				
3	20	16	12	9	2					
4	12	9	5							
5	6									

Table 6.3: Call admission region with packet switching transmission ($P^*_{WTE} = 10^{-4}$)

	N_{2l}												
N_{2h}	0	1	2	3	4	5	6	7	8	9	10	11	12
0	40	35	32	28	25	22	19	16	12	8	5	1	
1	33	31	27	24	21	18	15	12	8	5	3		
2	28	25	22	19	16	13	10	7	3				
3	23	20	17	13	10	7	3	1					
4	17	14	11	7	4	2							
5	11	7	5	1									
6	6	1											

Tables 6.2 and 6.3 give the call admission regions (N_1, N_{2h}, N_{2l}) respectively estimated with the traditional circuit switching transmission and the proposed packet switching transmission, respectively. The row numbers represent the number of low-quality video calls and the column numbers represent the number of high-quality video calls that are in the system. The table entries represent the maximum number of voice calls that can be admitted under the constraints of uplink cell capacity and packet loss. The call admission region is derived with $P^*_{WTE} = 10^{-4}$. It can be seen from these two tables that the proposed packet switching transmission enlarges the call admission region significantly compared to that obtained from circuit switching transmission for the same packet error rate requirement and cell capacity. The enlarged call admission region can accommodate more calls, which

can improve the new call blocking probability, the handoff call dropping probability and the radio resource utilization. With different P^*_{WTE} settings, different multiplexing gain and different call level performance measures can be achieved. For demonstration purposes, we set P^*_{WTE} to 10^{-4} and 10^{-5}, respectively, for uplink cell capacity derivation. When P^*_{WTE} equals to 10^{-4}, additional error rate can be introduced in packet scheduling for voice traffic, while no more packet loss can be endured for video traffic. On the other hand, when P^*_{WTE} is set to 10^{-5}, both voice and video traffic can introduce packet loss according to (3.2). Table. 6.4 gives the shrunken call admission region derived with $P^*_{WTE} = 10^{-5}$ due to the degraded multiplexing gain in wireless transmission compared to that in Table. 6.3.

Table 6.4: Call admission region with packet switching transmission ($P^*_{WTE} = 10^{-5}$)

	N_{2l}										
N_{2h}	0	1	2	3	4	5	6	7	8	9	10
0	35	32	27	24	21	17	14	11	8	4	
1	29	26	23	20	16	14	10	7	4	1	
2	24	21	17	15	11	9	5	2			
3	18	16	12	9	6	3					
4	13	9	6	3							
5	7	4	1								
6	1										

To evaluate the performance of the proposed adaptive resource reservation scheme, we consider a homogeneous voice traffic system with each voice call occupying one channel. Comparisons are made with two other schemes in the literature: (a) specific channel reservation based on RPSS values [38] and (b) specific channel reservation based on the remaining time to handoff [66]. In specific channel reservation, the reserved resources are only for the user who reserves them. Therefore, reservation cancellation is required to release falsely reserved channels because the potential handoff call either terminated before handoff or changed direction. In (a), if the RPSS drops below the reservation threshold for a predefined period, the MS requires the associated BS to release the reserved capacity. In (b), whenever the expected remaining time to handoff elapses, the reserved resources are released to avoid false reservation. On the other hand, there is no reservation cancellation for the proposed scheme. In addition, there is no reservation checkup at the time a user moves out of the handoff region as in [38] because in a reasonable mobility pattern, it is assumed that there are no abrupt acceleration changes. Therefore, the proposed reservation scheme based on the remaining time to handoff performs better than that based on RPSS values since the probability of false reservation is reduced. We denote schemes (a), (b) and the proposed schemes as spec_RPSS, spec_time, comm_time, respectively. Table. 6.5 tabulates the achieved minimum GoS performance with the ratio of the number of pedestrian to vehicle users (denoted as mobility in the table) being 0.6, 0.7, and 0.8, respectively. As both schemes (a) and (b) are based on a fixed threshold value and none of them takes traffic variations into account, we set the new call arrival rate at a fixed value of 0.35 calls/sec, and ignore the call event burstiness with a fixed threshold to achieve minimum GoS for comparison purposes. Therefore, the GoS performance is defined as a combination of the new call blocking probability and the weighted handoff call dropping probability GoS $= P_b + \gamma P_d$. It can be seen from Table. 6.5 that the proposed scheme outperforms the other two schemes in that the GoS performance is improved significantly in the same communication environments with the same traffic load. As the ratio of the number of pedestrian to vehicle users decreases, the performance of GoS improves gradually for all the three schemes because the high channel occupancy and release rates result in efficient resource utilization. The proposed adaptive resource reservation scheme is also an efficient approach for prioritizing handoff calls while at the same time keeping the new call blocking probability low.

Fig. 6.3 shows the effect of user mobility on GoS performance with different reservation-request time thresholds. It can be seen that, with small T_{rev}, the GoS degrades because of the high handoff call dropping probability caused by insufficient resource reservation. As T_{rev} increases, the GoS degrades as well owing to the high new call blocking probability caused by excessive resource reservation. As the ratio of the number of pedestrian to vehicle users increases, the smallest reservation-request time threshold that achieves minimum GoS performance increases from 6 to 9 s. This is because that high user mobility increases

Table 6.5: GoS performance (new voice call arrival rate = 0.35 calls/s)

mobility	0.8	0.7	0.6
spec_RPSS	0.1240	0.1143	0.0815
spec_time	0.1113	0.0981	0.0782
comm_time	0.0785	0.0742	0.0692
improvement over spec_RPSS	36.7%	35.1%	15.1%
improvement over spec_time	29.5%	24.4%	11.5%

multiplexing gain in resource utilization, which can decrease the T_{rev} value. The increasing reservation-request time threshold causes inefficient utilization of the reserved resources, especially in low mobility environments. Capacity utilization vs. reservation-request time threshold is shown in Fig. 6.4. When the reservation-request time threshold is small, the capacity is efficiently utilized in a low mobility environment owing to a low handoff rate and, accordingly, a low proportion of reserved resources. As the time threshold increases, more resources are reserved for potential handoff calls which makes capacity utilization inefficient. A low mobility environment suffers even worse capacity utilization degradation than a high mobility environment because a call with a slow moving user has more chance to terminate before handoff, which makes an excessive resource reservation even though those reserved channels can be used by other handoff calls.

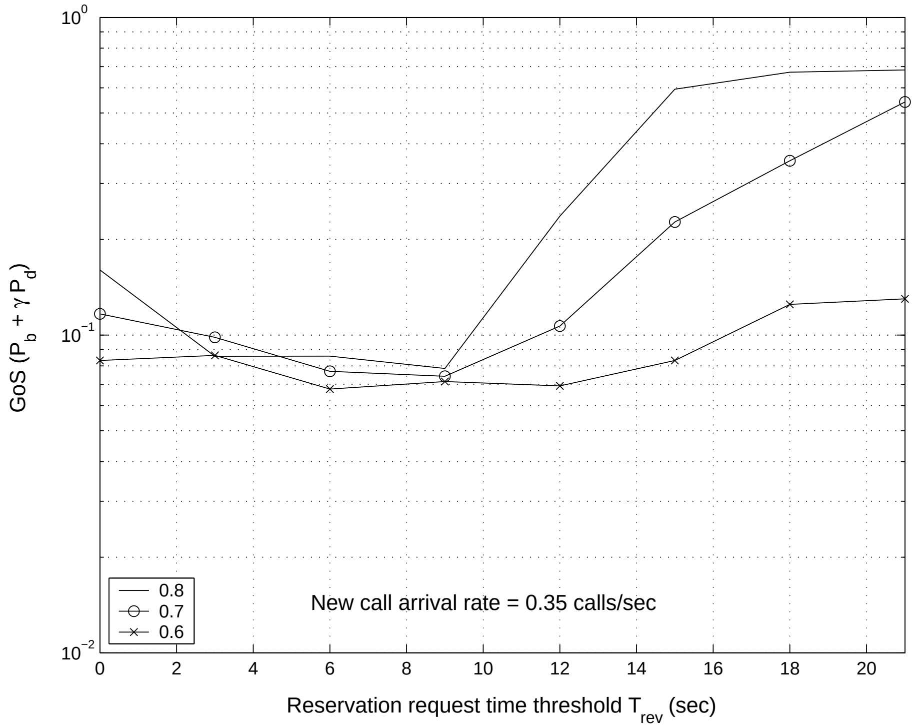

Figure 6.3: GoS performance vs. reservation-request time threshold with the ratio of the number of pedestrian to vehicle users being 0.8, 0.7 and 0.6

The performance of the reservation-request time threshold adaptation to the varying traffic load is shown as follows. As the traffic load varies with time, the required short-term GoS performance in terms of the estimated new call blocking and handoff call dropping rates is used to drive the dynamic determination of a proper reservation-request time threshold. Fig. 6.5 demonstrates the GoS performance. As stated above, here the time basis is mapped to the prescribed number of handoff call arrival events that happened in

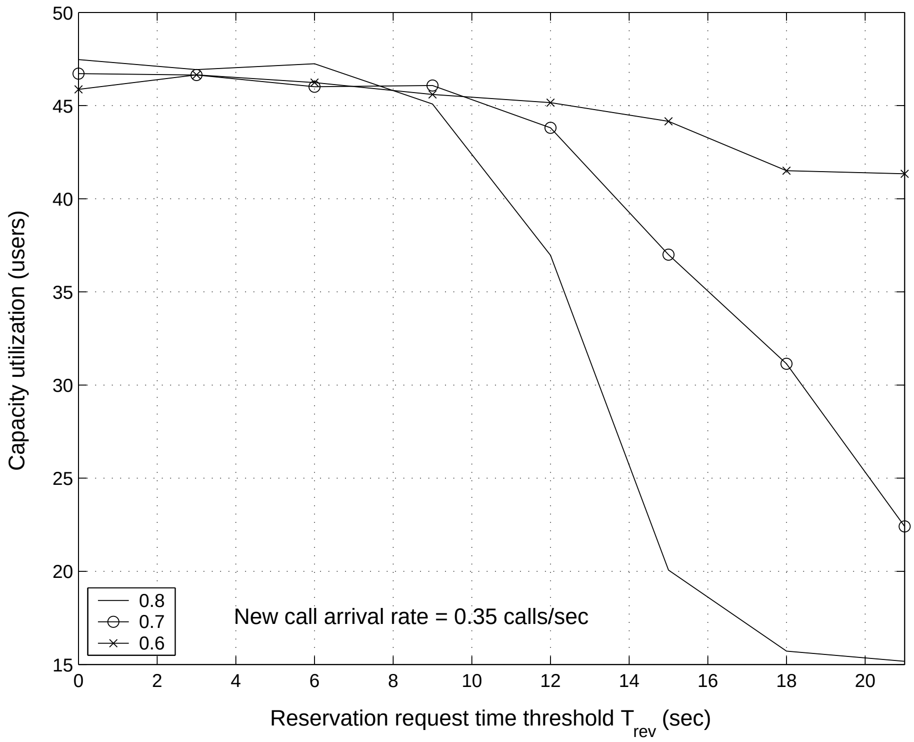

Figure 6.4: capacity utilization vs. reservation-request time threshold with the ratio of the number of pedestrian to vehicle users being 0.8, 0.7 and 0.6

a time duration. The new call arrival rate is changed from 0.3 calls/sec to 0.4 calls/sec. The ratio of the number of pedestrian to vehicle users is set to be 0.7. For demonstration purposes, we only show the curve with the parameter c of the reservation time adaptation strategy set to 1.5. As can be seen from the figure, with increasing traffic load in the system, the GoS performance increases as well, which indicates the accordingly increased new call blocking and handoff call dropping rates. The figure also demonstrates that the

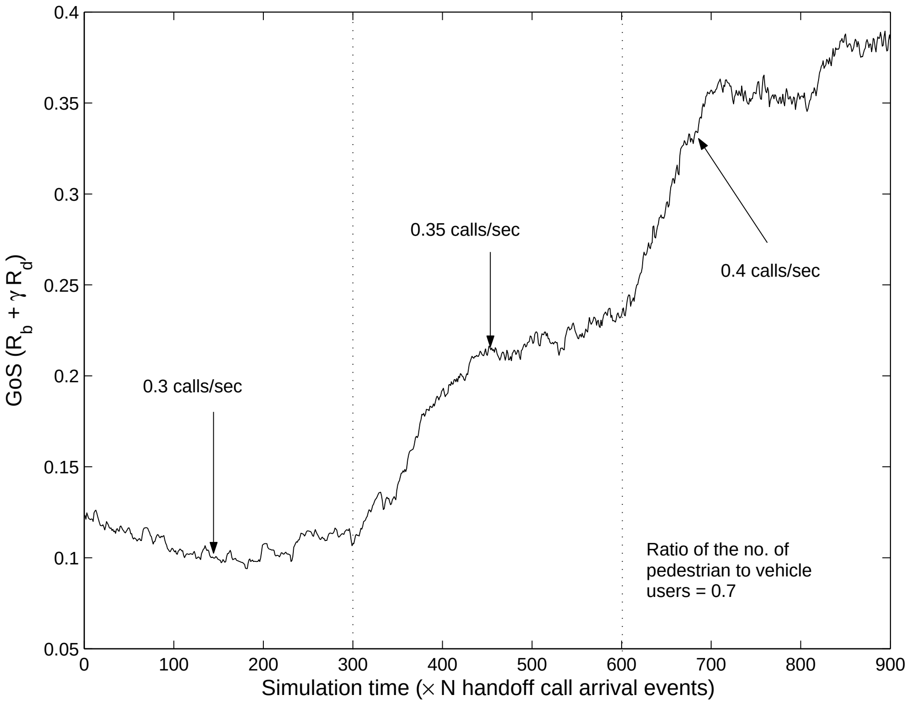

Figure 6.5: GoS performance vs. simulation time with the ratio of the number of pedestrian to vehicle users being 0.7 and $c = 1.5$

proposed adaptive threshold scheme can track both the traffic load variations and the burst call events. The reservation-request time threshold vs. traffic load is shown in Fig. 6.6. With increasing traffic load, the T_{rev} value increases as expected to reserve more resources for the increasing handoff call arrival events. To demonstrate the effect of different c values on traffic load adaptation, Table. 6.6 gives the mean and variance of the ratio of new call blocking and handoff call dropping rates R_b/R_d. It is seen that the mean of R_b/R_d is closely kept at the value of $\gamma = 10$. As expected, because of the slowly varying traffic load, a small c

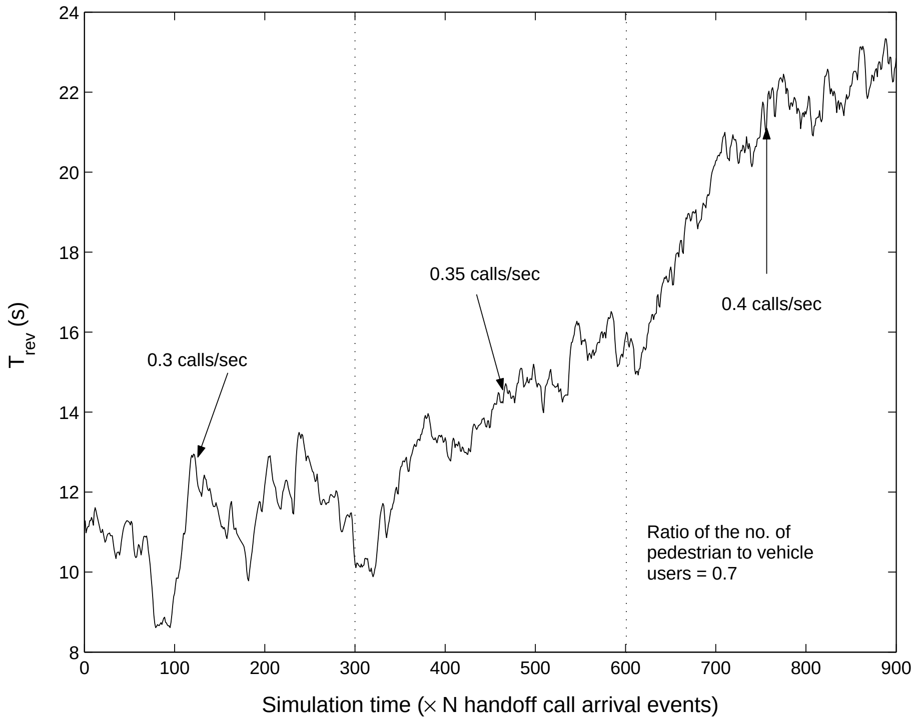

Figure 6.6: Reservation-request time threshold vs. simulation time with the ratio of the number of pedestrian to vehicle users being 0.7 and $c = 1.5$

value performs better than a larger c value in our simulation. When $c = 1.5$, the deviation of R_b/R_d from the objective $\gamma = 10$ achieves its minimum variance of 0.0725. Because the choice of c depends on many factors, such as traffic variation rate, GoS performance estimation, mobility prediction and so on, a reasonable c value can be determined based on a specific communications environment. A rather small variance for all cases in the simulation substantiates the viability of the proposed adaptive resource reservation scheme

Table 6.6: Mean and variation of R_b/R_d

c	1.1	1.5	2	3	4
Mean	9.9809	10.1060	9.9962	9.9307	9.8680
Variance	0.0938	0.0725	0.2512	0.6530	0.6350

to both user mobility and traffic load variations in handoff prioritization.

To focus on the proposed adaptive handoff priority scheme supporting multimedia applications, we simplify the simulation environment by not considering the proposed adaptive resource reservation scheme. It is assumed that the new video traffic load is fixed at 2 Erlang. The ratios of handoff traffic load to the new traffic load for voice and video are set to be 1/3 and 1/10, respectively. Figs. 6.7 and 6.8 show the new call blocking and handoff call dropping probabilities with the call admission regions given in Tables. 6.2 and 6.3. It can be seen that the new call blocking and the handoff call dropping probabilities are significantly reduced with the enlarged call admission region. Owing to the much higher transmission rate and more stringent transmission error rate requirement, video calls suffer relatively higher new call blocking and handoff call dropping probabilities compared with voice calls. It can also be seen that the ratio of new call blocking probabilities and handoff call dropping probabilities remains around 10.

Resource utilization in terms of the admitted voice and video calls vs. new voice traffic load is shown in Fig. 6.9. With increasing voice traffic load, the resource utilization for

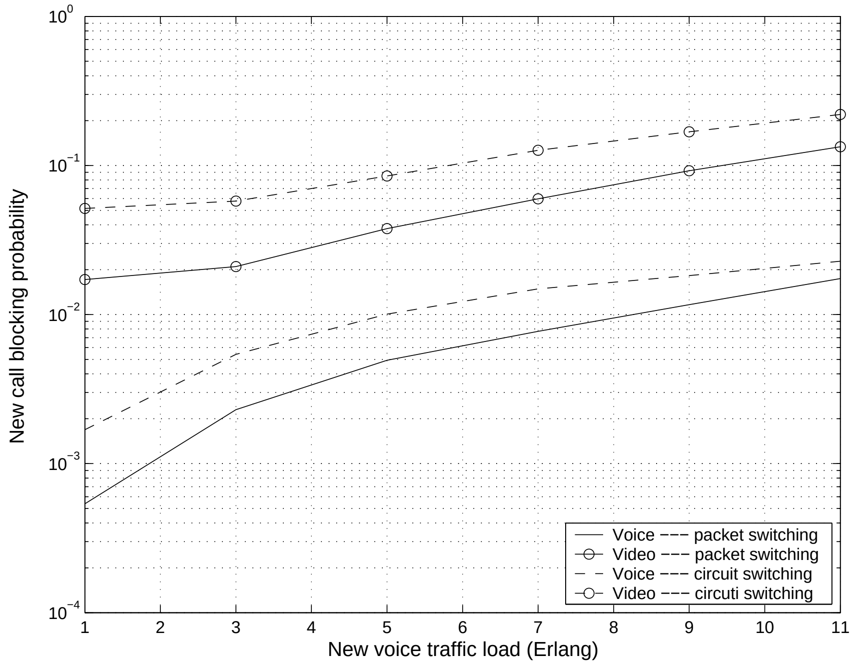

Figure 6.7: New call blocking probabilities with the packet switching and circuit switching transmissions

voice traffic increases accordingly, and that for video traffic decreases gradually. This is due to the increasing new call blocking and handoff call dropping probabilities for video calls. The figure also shows that with the enlarged call admission region, the video traffic resource utilization is increased.

The ratio of the number of low quality video calls over the total number of video calls in the system is shown in Fig. 6.10. It can be seen that the low quality video call occupancy increases with voice traffic load. For the maximum traffic load in the simulation, the ratio

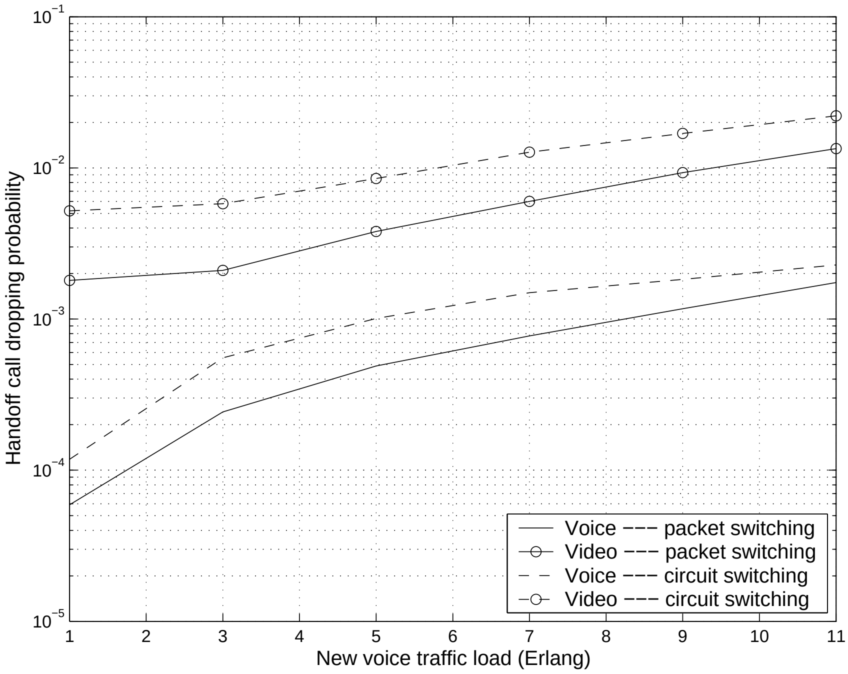

Figure 6.8: Handoff call dropping probabilities with the packet switching and circuit switching transmissions

obtained with the proposed packet switching wireless transmission scheme is 0.07, which significantly outperforms the ratio obtained with the traditional circuit switching wireless transmission scheme, which is 0.17. This result can be reasonably accepted with user satisfaction.

Figs. 6.11 and 6.12 show the new call blocking and the handoff call dropping probabilities with the proposed adaptive handoff scheme and the non-adaptive handoff scheme, respectively. In the non-adaptive handoff scheme, the transmission rate of video calls is

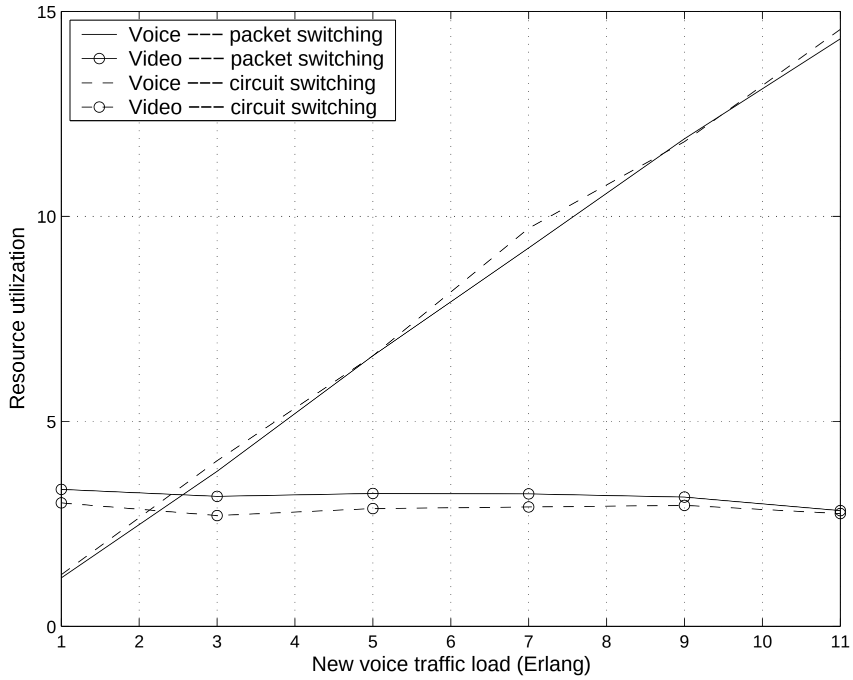

Figure 6.9: Resource utilization with the packet switching and circuit switching transmissions

fixed during call holding time. Because the transmission rate of rate-adaptive video calls is reduced to accommodate handoff calls, both handoff voice calls and handoff video calls are prioritized with much lower dropping probabilities compared to the new call blocking probabilities. Owing to the admission of more handoff calls into the system, the performance of new call blocking probabilities of both voice and video calls is slightly reduced compared to those obtained with the non-adaptive handoff schemes.

Figs. 6.13 - 6.15 show the effect of different prescribed P^*_{WTE} values on performance

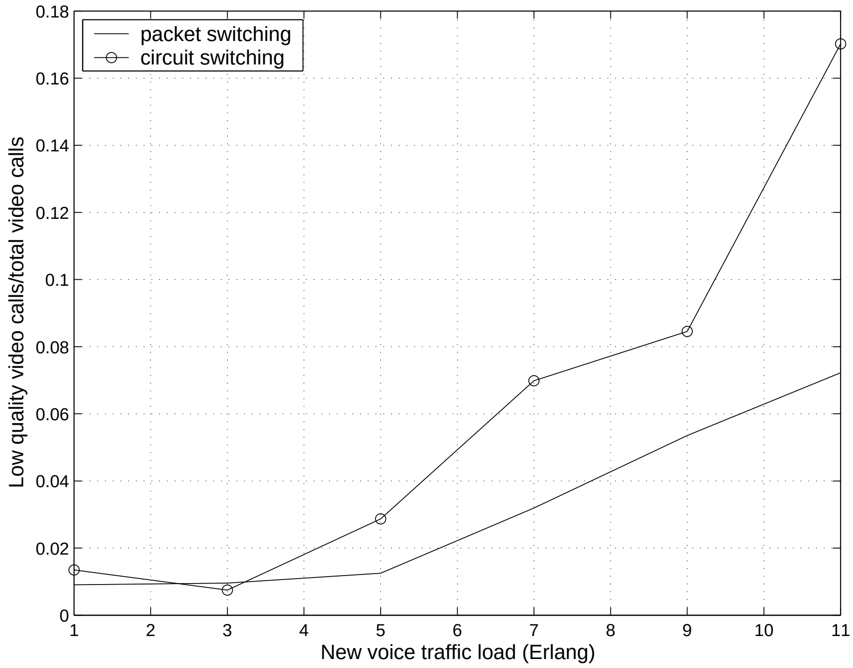

Figure 6.10: Ratios of low quality video calls over total video calls with the packet switching and circuit switching transmissions

comparisons. As expected, the performance in terms of new call blocking probability, handoff call dropping probability, and the ratio of the low quality video calls over the total ongoing video calls obtained with a less stringent P^*_{WTE} value 10^{-4} significantly outperforms that obtained with a more stringent P^*_{WTE} value 10^{-5}.

Performance evaluation of the proposed adaptive resource reservation scheme and adaptive handoff priority scheme supporting multimedia applications is also investigated via simulations. As mentioned in Chapter 1, resource reservation for wideband video calls

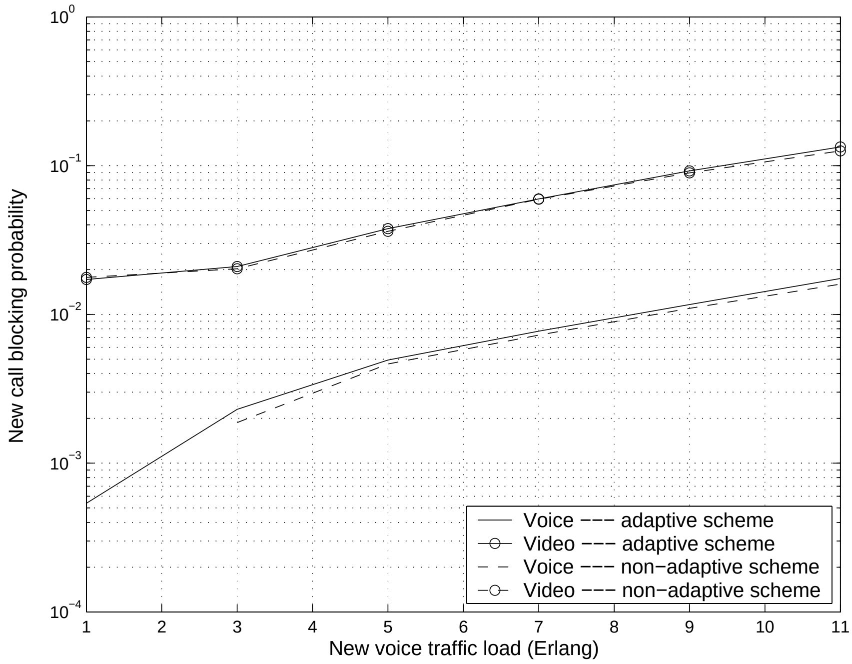

Figure 6.11: New call blocking probability with the adaptive priority and non-adaptive handoff schemes

will result in inefficient utilization of radio resources and degrade system performance. So, in the simulation, resource reservation for a potential video handoff call will only provide low-quality service. With a successful resource reservation, and if after a successful handoff there is still enough spare resource left, the low-quality video call will be provided with high-quality service. Therefore, the ratio of the number of low-quality video calls over the total number of video calls in the system can be kept at a low value. Also in this case, the GoS function defined in (2.7) in Chapter 2, and the new call blocking and handoff call

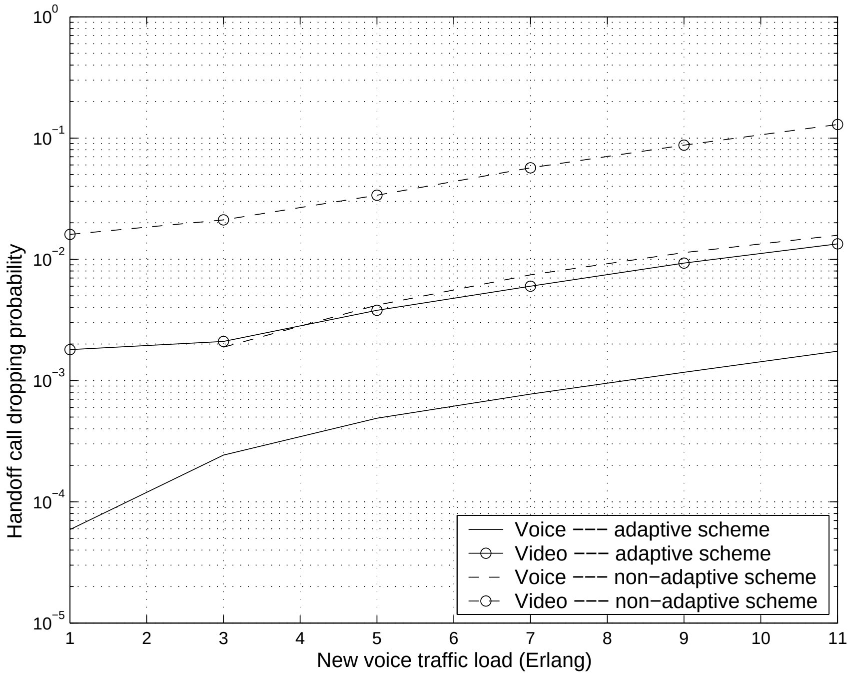

Figure 6.12: Handoff call dropping probability with the adaptive priority and non-adaptive handoff schemes

dropping rates that drive the adjustment of the reservation-request time threshold, need to be modified as:

$$\mathrm{GoS} = R_b + \gamma R_d, \tag{6.4}$$

and

$$\begin{cases} R_b = \upsilon_{voice} R_{b_{voice}} + \upsilon_{video} R_{b_{video}} \\ \\ R_d = \omega_{voice} R_{d_{voice}} + \omega_{video} R_{d_{video}}. \end{cases} \tag{6.5}$$

Here υ_{voice} and υ_{video} are the ratio of the number of new voice and new video call arrivals

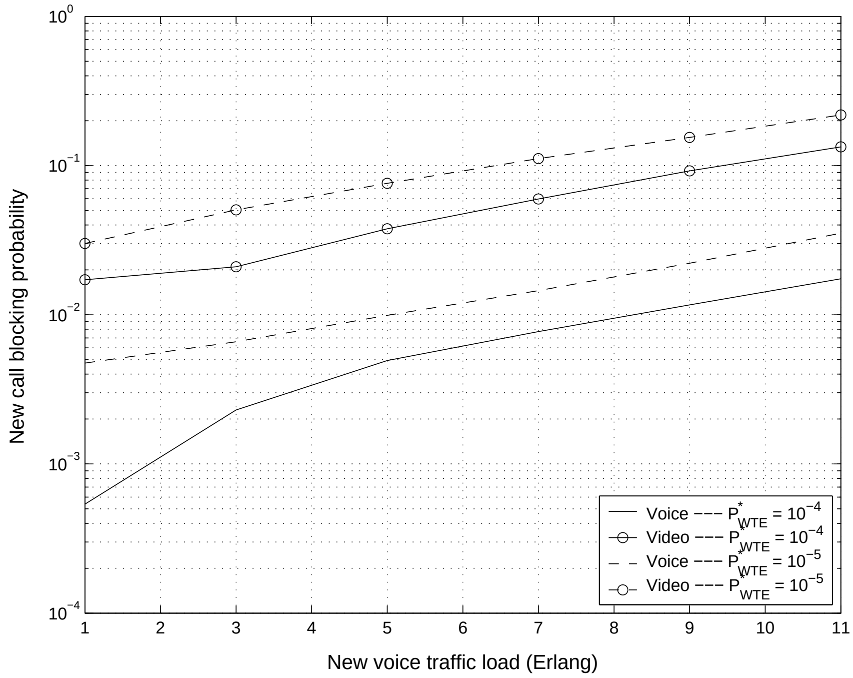

Figure 6.13: New call blocking probability with P^*_{WTE} being 10^{-4} and 10^{-5}

over the total new calls arrival events, and ω_{voice} and ω_{video} are the ratio of the number of handoff voice and handoff video call arrivals over the total handoff call arrival events, respectively. The definition of R_b and R_d in this form combines new call blocking and handoff dropping events from both the voice and the video calls.

Table. 6.7 demonstrates the GoS performance with the new voice call arrival rate being 0.12, 0.125 and 0.13 calls/sec. The new video call arrival rate is fixed at 0.12 calls/sec. The ratio of the number of pedestrian to vehicle users is set to be 0.7 and the call admission region is derived with P^*_{WTE} equal to 10^{-4}. Table. 6.7 also shows the significant GoS

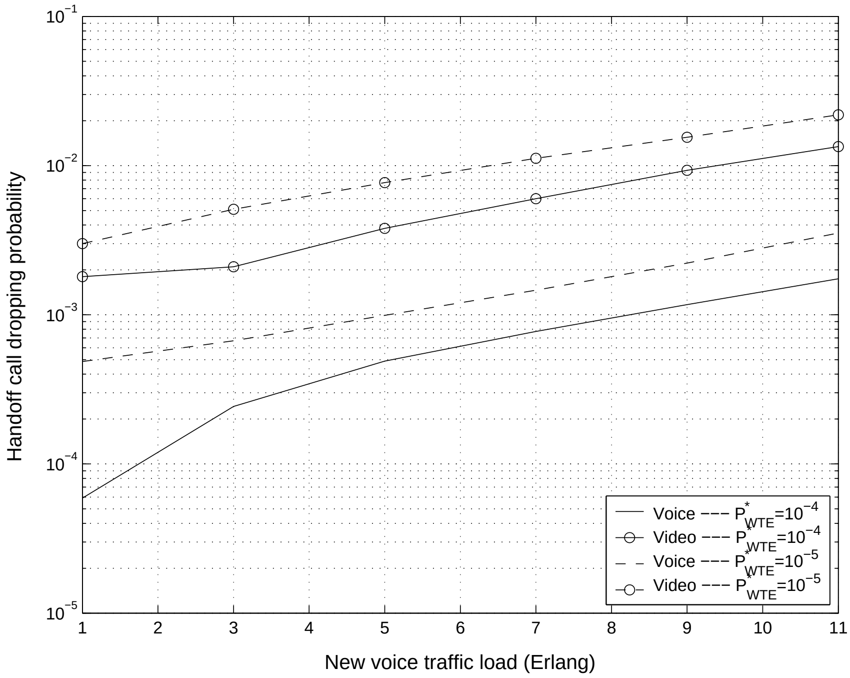

Figure 6.14: Handoff call dropping probability with P^*_{WTE} being 10^{-4} and 10^{-5}

performance improvement as compared with the other two schemes.

Table 6.7: GoS performance (the ratio of the number of pedestrian to vehicle users = 0.7, $P^*_{WTE} = 10^{-4}$)

$\lambda_{nvoice}(calls/sec)$	0.12	0.125	0.13
spec_RPSS	0.0879	0.0926	0.1089
spec_time	0.0741	0.0857	0.0985
comm_time	0.0442	0.0519	0.0627
improvement over spec_RPSS	49.7%	43.9%	42.4%
improvement over spec_time	40.4%	39.4%	36.4%

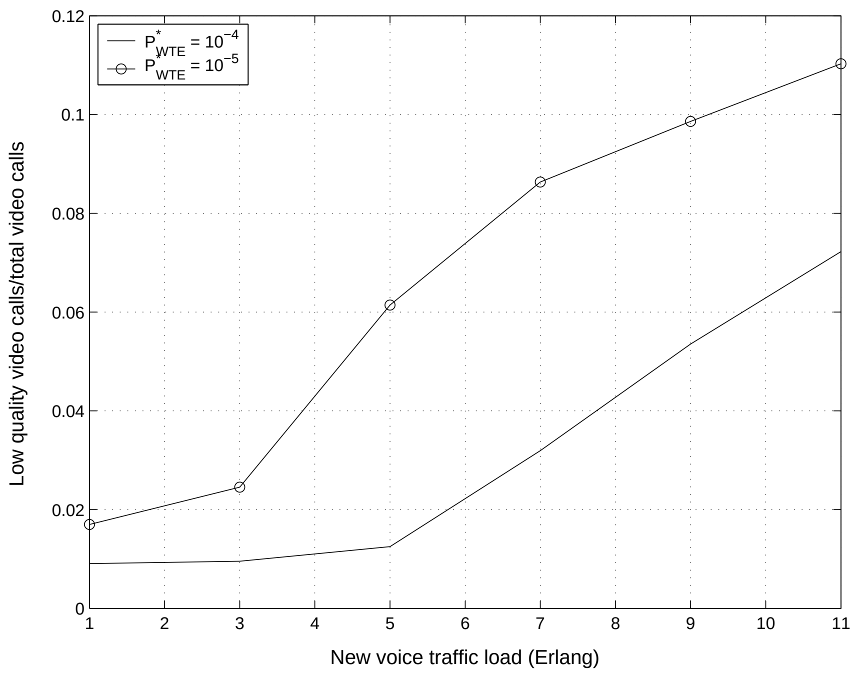

Figure 6.15: Ratio of low quality video calls over total video calls with P^*_{WTE} being 10^{-4} and 10^{-5}

Fig. 6.16 shows the GoS performance vs. different reservation-request time threshold with different traffic load. With the new voice call arrival rate increasing from 0.12 calls/sec to 0.13 calls/sec, the reservation-request time threshold that achieves minimum GoS performance increases from 1 to 3 s. This is because (1) high traffic load increases resource occupation, which reduces the possibility of a successful resource reservation; and (2) high traffic load introduces high handoff call arrival rate, which requires a large T_{rev} value to get the required resource ready.

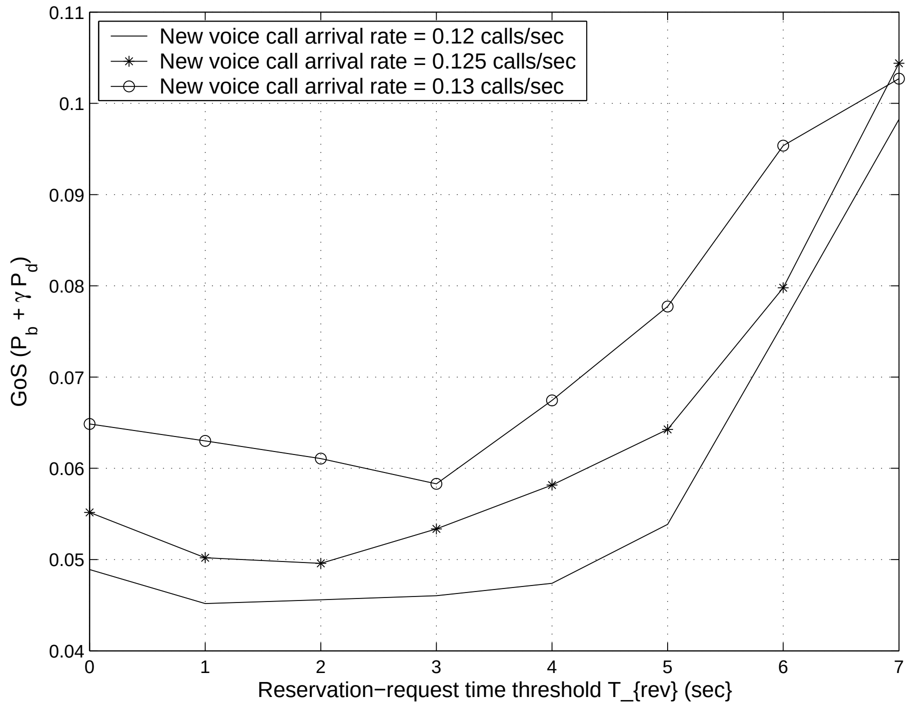

Figure 6.16: GoS performance vs. reservation-request time threshold with the new voice call arrival rate being 0.12, 0.125 and 0.13 calls/s

Fig. 6.17 shows the GoS performance adaptation to varying traffic load. The new call arrival rate is changed from 0.12 calls/sec to 0.14 calls/sec for demonstration purposes. Here only the curve with $c = 2$ is shown because in this case $c = 2$ achieves the minimum R_b/R_d variance of 6.0958. This value is much greater than 0.0725 obtained in a homogeneous system supporting voice traffic alone due to the fact that (1) resource reservation for wideband video calls in a target cell causes severe resource utilization degradation and accordingly causes much higher new call blocking and handoff call dropping events; (2)

complete resource sharing between narrow-band voice traffic and wideband video traffic will affect greatly on call admission decisions for each type of traffic, especially voice traffic performance is severely impacted by video traffic call admission control.

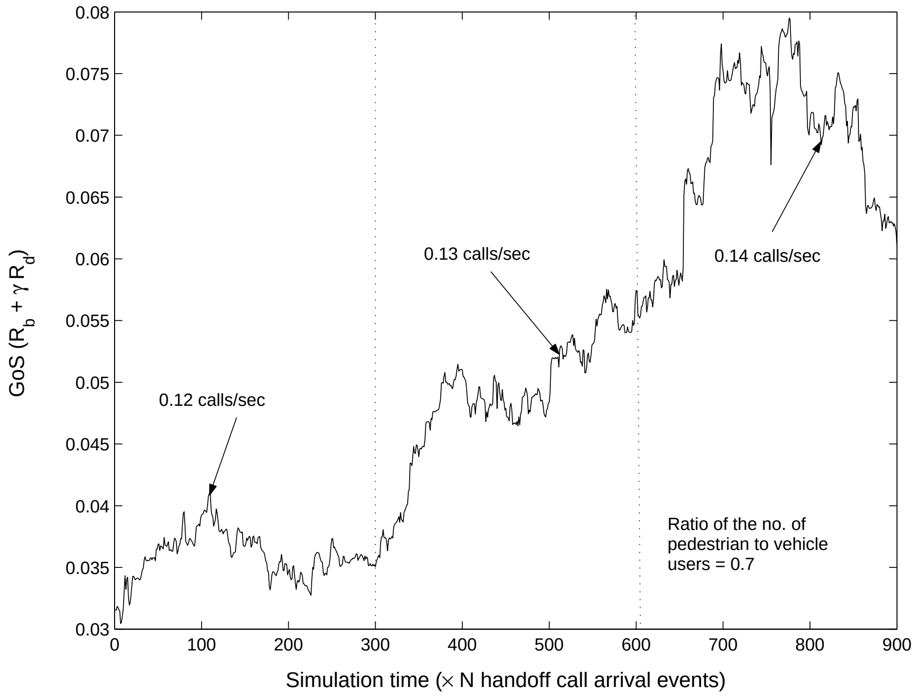

Figure 6.17: GoS performance vs. simulation time with the ratio of the number of pedestrian to vehicle users being 0.7 and $c = 2$

6.4 Summary

Numerical results have been presented in this chapter to demonstrate the performance of the proposed adaptive resource reservation scheme and the proposed handoff priority scheme supporting multimedia applications in a cellular CDMA system. The results show that:

- the cell capacity derivation considers the effect of imperfect power control and distinguish three kinds of interference;
- the proposed packet scheduling scheme can provide connections of different traffic types with different packet loss rate requirements and guarantee fairness among the connections of the same traffic type;
- different packet error rate due to wireless transmission and packet loss rate pairs have different effects on the call admission region derivation;
- a loose wireless transmission error constraint can increase cell capacity and result in better multiplexing gain at the link layer where multiple users share the same transmission media;
- compared with other resource reservation schemes in the literature, the proposed adaptive resource reservation scheme based on the reservation-request time threshold

not only can achieve better GoS performance, but also can adapt to both user mobility randomness and traffic load variation;

- compared with non-adaptive handoff schemes, service adaptation of rate-adaptive video calls to the available resources can improve handoff call dropping probability, while keeping the ratio of the number of low-quality video calls over the total number of video calls low to maintain service satisfaction.

Chapter 7

Conclusions and Future Work

7.1 Summary of contributions

A novel adaptive resource reservation scheme and a handoff priority scheme supporting multimedia applications have been proposed and evaluated. The schemes are based on the uplink cell capacity estimation and call admission region derivation with a new packet scheduling scheme for a cellular CDMA system by jointly taking the physical layer, the link layer, the network layer characteristics, the traffic characteristics, and the user mobility characteristics into account. The proposed schemes achieve effective and efficient handoff priority and radio resource utilization, which can both guarantee user satisfaction and benefit service providers. The main contributions of this thesis are listed as follows.

- Uplink cell capacity is properly estimated for a cellular CDMA network with the consideration of three types of interference from users connected to and power controlled by the reference BS, handoff users connected to the reference BS but power controlled by one of its neighboring BS, and users in the outer cells;
- An efficient packet scheduling and partial packets integration scheme is proposed to provide the required packet loss rate to different types of traffic and make efficient use of the limited radio resources;
- A call admission region for call admission control and handoff management is effectively derived, which guarantees the bit error rate requirement with the limited simultaneously transmitted packets and guarantees the packet loss rate requirement with the proposed packet scheduling scheme;
- The relation between the bit error rate due to wireless transmission and the packet loss rate due to link layer congestion is derived and the effect of the relation pair on call admission region is discussed;
- A user mobility model mimicking practical user movement is applied to the resource reservation scheme for accurate handoff prediction;
- An adaptive resource reservation scheme is developed whereby the introduced reservation-request time threshold can adapt to both user mobility and traffic variation to track

handoff attempts closely for better handoff performance and resource utilization.

- A pooling algorithm is proposed which makes the reserved resources a common pool for all the coming handoff calls to improve multiplexing gain, to eliminate false resource reservation and to simplify signaling process;

- An adaptive handoff priority scheme for multimedia traffic is proposed whereby the rate-adaptive high-quality video calls can be temporarily converted to low-quality service when a handoff call arrives and there are no spare resources available for handoff priority. Whenever there is resource release due to either call termination or call handoff to neighboring cells, the low-quality video call will try to improve service quality to keep the ratio of the number of low-quality video calls over the total video calls as low as possible.

7.2 Future work

Future wireless communication systems will have higher capacity, higher data transmission rates, more advanced services and flexible QoS provisioning, seamless handoff and user roaming among different networks. As this thesis focuses mainly on handoff initiation and handoff execution with immediate resource reallocation for realtime traffic, future research work on handoff management will also be extended to many other aspects.

- Development of inter-switch handoff algorithms and their performance evaluations. The inter-switch handoff is characterized by a long handoff delay which will have an effect on realtime services. To achieve a mean handoff delay that satisfies a prescribed delay tolerance is of great importance, and will entail a study of the manner in which inter-switch handoff is initiated and executed.

- As the Internet is becoming increasingly popular, IP may potentially become the universal network-layer protocol for both wireline and wireless communications. Handoff supports in mobile IP networks usually rely on rerouting algorithms, packet data delivery systems, etc.. To achieve fast handoff, a direct path from the old BS and the target new BS is established for packet forwarding. From the end correspondents' perspective, this kind of routing may not be optimal and sometimes the old BS might also become the bottleneck when many handoffs take place simultaneously. Therefore, optimal rerouting will gain more attention in future wireless Internet systems.

- In soft handoff, as two or more associated BS are transmitting the same copy of data to one MS, how to coordinate and reorder the transmission of multiple copies among the associated BSs, especially when the involved BSs belong to different systems, needs further investigation.

- As the cellular wireless networks are growing rapidly, other air interfaces, such as

Bluetooth, Wireless Local Area Network (WLAN), Ad Hoc, sensor network, satellite network have come into being and developed. It is expected that users may wish to access different systems according to the time, location, service availability, or other conditions. Therefore, internetworking among different kinds of wireless access networks, and the capability of seamless handoff among these networks will be of paramount importance. It needs standardized network architecture for regional, national and global services.

Bibliography

[1] *The cdma2000 ITU-R RTT candidate submission (0.18)*, June 1998.

[2] *ETSI UMTS Terrestrial Radio Access(UTRA) ITU-R RTT Candidate submission*, June 1998.

[3] *Global CDMA I: Multiband Direct-sequence CDMA System RTT System Description*, June 1998.

[4] *Global CDMA II for IMT-2000 RTT System Description*, June 1998.

[5] *Japan's Proposal for Candidate Radio Transmission Technology on IMT-2000:W-CDMA*, June 1998.

[6] A. Acharya, J. Li, A. Bakre, and D. Raychaudhuri. Design and prototyping of location management and handoff protocols for wireless ATM networks. In *1997 IEEE 6th*

International Conference on Universal Personal Communications Record, volume 1, pages 12–16, October 1997.

[7] M. Akar and U. Mitra. Soft handoff algorithms for CDMA cellular networks. *IEEE Transactions on Wireless Communications*, 2(6):1259–1274, November 2003.

[8] S. Ariyavisitakul and L. Chang. Signal and interference statistics of a CDMA system with feedback power control. *IEEE Transactions on Communications*, 41:1626–1634, November 1993.

[9] D. Ayyagari and A. Ephremides. Cellular multicode CDMA capacity for integrated (voice and data) services. *IEEE Journal On Selected Areas in Communications*, 17(5):928–938, May 1999.

[10] A. Baiocchi, F. Delli Priscoli, and F. Sestini. Improving the erlang capacity of a CDMA cellular network under bursty user mobility. In *1996 5th IEEE International Conference on Universal Personal Communications*, volume 1, pages 194–199, Sept. 29 1996.

[11] D. Calin. A probabilistic model for handling voice and data traffic in wireless networks. In *IEEE International Conference on Universal Personal Communications*, volume 1, pages 317–321, 1998.

[12] A. T. Campbell, J. Gomez, and A. G. Valko. Handoff in cellular systems. *IEEE Personal Communications*, 5:26–37, December 1998.

[13] M.-H. Chiu and M. Bassiouni. Performance evaluation of position-based channel reservation for handoff of cellular calls. In *IEEE Vehicular Technology Conference*, volume 3, pages 1799–1803, May 1999.

[14] M.-H. Chiu and M. A. Bassiouni. Predictive schemes for handoff prioritization in cellular networks based on mobilie positioning. *IEEE Journal On Selected Areas in Communications*, 18(3):510–522, March 2000.

[15] C. Cho, Y. Ko, and L.-K. Hyung. Fuzzy adaptive guard channel assignment strategy for handoff in PCS system. In *Fuzzy Systems, 1997., Proceedings of the Sixth IEEE Internationl Conference on*, volume 3, pages 1511–1517, July 1997.

[16] K. C. Chua, B. Bensaou, W. Zhuang, and S. Y. Choo. Dynamic channel reservation (DCR) scheme for handoffs prioritization in mobile micro/picocellular networks. In *Universal Personal Communications, 1998. ICUPC'98. IEEE 1998 International Conference on*, volume 1, pages 383–387, October 1998.

[17] K. C. Chua and S. Y. Choo. Probabilistic channel reservation scheme for mobile picl/microcellular networks. *IEEE Communications Letters*, 2:195–196, July 1998.

[18] X. Duan, Z. Niu, and J. Zheng. Capacity analysis of uplink and downlink in multimedia DS-CDMA systems based on constraint models. In *IEEE International Conference on Communications*, volume 1, pages 402–406, May 11 - 15 2003.

[19] H. G. Ebersman and O. K. Tonguz. Handoff ordering using signal prediction priority queueing in personal communication systems. *IEEE Transactions on Vehicular Technology*, 48(1):20–35, January 1999.

[20] Y. Fang, I. Chlamtac, and Y.-B. Lin. Channel occupancy times and handoff rate for mobile computing and PCS networks. *IEEE Transactions on Computers*, 47(6):679–692, June 1998.

[21] F. H. P. Fitzek, B Rathke, M. Schlager, and A. Wolisz. Quality of service support for real-time multimedia applications over wireless links using the simultaneous MAC-packet transmission (SMPT) in a CDMA environment. In *Proc. MoMuC 1998*, pages 367–378, October 1998.

[22] V. K. Garg and J. E. Wilkes. *Wireless and Personal Communications Systems*. Prentice Hall, 1996.

[23] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini. Error statistics of real-time power measurements in cellular channels with multipath and shadowing. *IEEE Transactions on Vehicular Technology*, 43(3):439–446, Auguest 1994.

[24] S. J. Grant and J. K. Cavers. A method for increasing downlink capacity by coded multiuser transmission with a base station antenna array. *IEEE Transactions on Communications*, 51(5):832–840, May 2003.

[25] M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *IEE Electronics Letters*, 27(23):2145–2146, November 1991.

[26] I. Han and D.-H. Cho. Group location management for mobile subscribers on transportation systems in mobile communication networks. *IEEE Transactions on Vehicular Technology*, 53(1):181–191, January 2004.

[27] S. Hara and R. Prasad. Overview of multicarrier CDMA. *IEEE Communications Magazine*, 35(12):126–133, December 1997.

[28] B. Hashem and E. S. Sousa. On the capacity of cellular DS/CDMA systems under slow Rician/Rayleigh-fading channels. *IEEE Transactions on Vehicular Technology*, 49(5):1752–1759, September 2000.

[29] H. Homa and A. Toskala. *WCDMA for UMTS:Radio Access for Third Generation Mobile Communications.* John Wiley & Sons, 2000.

[30] F. Hu and N. K. Sharma. A novel CAC mechanism for guaranteeing QoS of multimedia traffic in next-generation wireless and mobile networks. In *IEEE International Conference on Communications*, volume 2, pages 417–421, June 11 - 14 2001.

[31] Q. Huang, S. Chan, and K.-T. Ko. Dynamic bandwidth sharing for handoff traffic in multimedia mobile cellular networks. *Electronics Letters*, 39(10):801–802, May 2003.

[32] V. Huang and W. Zhuang. Qos-oriented packet scheduling for wireless multimedia CDMA communications. *IEEE Transactions on Mobile Computing*, 3(1):73–85, January-February 2004.

[33] C.-L. I and R.D. Gitlin. Multi-code CDMA wireless personal communications networks. In *IEEE International Conference on Communications*, volume 2, pages 1060–1064, 1995.

[34] 3GPP2 IS-2000.2. *Physical Layer Standard for cdma2000 Spread Spectrum Systems.*

[35] 3GPP2 IS-2002.3. *Medium Access Control (MAC) Standard for cdma2000 Spread Spectrum Systems.*

[36] 3GPP2 IS-2002.4. *Signaling Layer 2 Standard for cdma2000 Spread Spectrum Systems.*

[37] 3GPP2 IS-2002.5. *Upper Layer (Layer 3) Signaling Standard for cdma2000 Spread Spectrum Systems.*

[38] J. W. Chang, D. K. Sung . Adaptive Channel Reservation Scheme for Soft Handoff in DS-CDMA Cellular Systems. *IEEE Transactions on Vehicular Technology*, 50(2):341–353, March 2001.

[39] M. G. Jansen and R. Prasad. Capacity, throughput, and delay analysis of a cellular DS CDMA system with imperfect power control and imperfect sectorization. *IEEE Transactions on Vehicular Technology*, 44(1):67–75, February 1995.

[40] S. K. Jayaweera and H. V. Poor. Capacity of multiple-antenna systems with both receiver and transmitter channel state information. *IEEE Transactions on Information Theory*, 49(10):2697–2709, October 2003.

[41] D. K. Kim and D. K. Sung. Characterization of soft handoff in CDMA systems. *IEEE Transactions on Vehicular Technology*, 48:1195–1202, July 1999.

[42] D. K. Kim and D. K. Sung. Capacity estimation for a multicode CDMA system with SIR-based power control. *IEEE Transactions on Vehicular Technology*, 50(3):701–710, May 2001.

[43] J. Kim, Y.-J. Choi, and S. Bahk. Interference-based capacity analysis in CDMA cellular systems. In *IEEE Wireless Communications and Networking*, volume 2, pages 1412–1417, March 2003.

[44] Y. C. Kim, D. E. Lee, B. J. Lee, Y. S. Kim, and B. Mukherjee. Dynamic channel reservation based on mobility in wireless ATM networks. *IEEE Communications Magazine*, pages 47–51, November 1999.

[45] K. Lee. Supporting mobile multimedia in integrated services networkings. *Wireless Networks*, 2(3):205–217, 1996.

[46] W. Li and A. S. Alfa. Channel reservation for handoff calls in a PCS network. *IEEE Transactions on Vehicular Technology*, 48:95–104, January 2000.

[47] Y.-B. Lin, A. R. Noerpel, and D. J. Harasty. The sub-rating channel assignment strategy for PCS handoff-offs. *IEEE Transactions on Vehicular Technology*, 45(1):122–130, February 1996.

[48] T. Liu, P. Bahl, and I. Chlamtac. Mobility modeling, lacation tracking, and trajectory prediction in wireless ATM networks. *IEEE Journal On Selected Areas in Communications*, 16(6):922–936, August 1998.

[49] S. Manji and W. Zhuang. Power control and capacity analysis for a packetized indoor multimedia DS-CDMA network. *IEEE Transactions on Vehicular Technology*, 49(3):911–935, May 2000.

[50] H. Nie and P. T. Mathiopoulos. Reverse link capacity analysis for cellular CDMA systems employing successive interference cancellation. In *International Conference on Communication Technology Proceedings*, volume 2, pages 770–773, April 2003.

[51] C. Oliveira, J. B. Kim, and T. Suda. An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks. *IEEE Journal on Selected Areas in Communications*, 16:858–874, August 1998.

[52] L. Ortigoza-Guerrero and A. H. Aghvami. A prioritized handoff dynamic channel allocation strategy for PCS. *IEEE Transactions on Vehicular Technology*, 48(4):1203–1215, July 1999.

[53] F.-N. Pavlidou. Mixed media cellular systems. *IEEE Transactions on Communications*, 42:848–853, February/March/April 1994.

[54] F.-N. Pavlidou. Two-dimensional traffic models for cellular mobile systems. *IEEE Transactions on Communications*, 42:1505–1511, February/March/April 1994.

[55] Q.-A.Zeng and D. P. Agrawal. Modeling and efficient handling of handoffs in integrated wireless mobile networks. *IEEE Transactions on Vehicular Technology*, 51(6):1469–1478, November 2002.

[56] R. G. Brown. *Smoothing, Forcasting and Prediction of Discrete Time Series.* Prentice-Hall Inc, December 1964.

[57] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore. Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks. *IEEE Journal on Selected Areas in Communications*, 17:1270–1283, July 1999.

[58] S. Boumerdassi, A.-L. Beylot. Adaptive Channel Allocation for Wireless PCN. *Mobile Networks and Applications*, 4:111–116, 1999.

[59] C. U. Saraydar and A. Yener. Capacity enhancement for CDMA systems through adaptive cell sectorization. In *IEEE Wireless Communications and Networking Conference*, volume 3, pages 1139–1143, September 21 - 24 1999.

[60] C. U. Saraydar and A. Yener. Adaptive cell sectorization for CDMA systems. *IEEE Journal On Selected Areas in Communications*, 19(6):1041–1051, June 2001.

[61] H. Sari, F. Vanhaverbeke, and M. Moeneclaey. Increasing the capacity of CDMA using hybrid spreading sequences and iterative multistage detection. In *IEEE Vehicular Technology Conference*, volume 2, pages 1160–1164, September 19 - 22 1999.

[62] M. Schwartz. *Broadband Integrated Networks.* Pretice Hall PTR, Upper Saddle River, 1996.

[63] X. Shen, J. W. Mark, and J. Ye. Mobile location estimation in cellular networks using fuzzy logic. In *Vehicular Technology Conference, 2000, IEEE VTS-Fall VTC 2000. 52nd*, volume 5, pages 2108–2114, September 2000.

[64] D. Shiung and J.-F. Chang. Enhancing the capacity of DS-CDMA system using hybrid spreading sequences. *IEEE Transactions on Communications*, 52(3):372–375, March 2004.

[65] T. Shu and Z. Niu. Uplink capacity optimization by power allocation for multimedia CDMA networks with imperfect power control. *IEEE Journal On Selected Areas in Communications*, 21(10):1585–1594, December 2003.

[66] W.-S. Soh and H. S. Kim. Dynamic bandwidth reservation in hierarchical wireless ATM networks using GPS-based prediction. In *IEEE Vehicular Technology Conference*, volume 1, pages 528–532, 1999.

[67] A. H. Solana, A. V. Bardaji, and F. C. Palacio. Capacity analysis and performance evaluation of call admission control for multimedia packet transmission in UMTS WCDMA system. In *2003 IEEE Wireless Communications and Networking*, volume 3, pages 1550–1555, March 16 - 20 2003.

[68] F. Vanhaverbeke, M. Moeneclaey, and H. Sari. Increasing cdma capacity using multiple orthogonal spreading sequence sets and successive interference cancellation. In *IEEE*

International Conference on Communications, volume 3, pages 1516–1520, April 28 - May 2 2002.

[69] V. V. Veeravalli and O. E. Kelly. A locally optimal handoff algorithm for cellular communications. *IEEE Transactions on Vehicular Technology*, 46(3):603–609, Auguat 1997.

[70] A. J. Viterbi. *CDMA:Principles of Spread Spectrum Communication*. Addison-Wesley Publishing Company, April 1995.

[71] J. Voigt and G. P. Fettweis. Influence of user mobility and simulcast-handoff on the system capacity in pico-cellular environments. In *IEEE Wireless Communications and Networking Conference*, volume 2, pages 712–716, Sept. 21 - 24 1999.

[72] S.-R. Yang and Y.-B. Lin. Performance evaluation of location management in UMTS. *IEEE Transactions on Vehicular Technology*, 52(6):1603–1615, November 2003.

[73] T. Zhang, E. Berg, J. Chennikara, P. Agrawal, J.-C. Chen, and T. Kodama. Local prediction resource reservation for handoff in multimedia wireless IP networks. *IEEE Journal on Selected Areas in Communications*, 19(10):1931–1941, October 2001.

[74] W. Zhuang, B. Bensaou, and K. C. Chua. Adaptive quality of service handoff priority scheme for mobile multimedia networks. *IEEE Transactions on Vehicular Technology*, 49:494–505, March 2000.

[75] W. Zhuang, K. C. Chua, and S. M. Jiang. Measurement-based dynamic bandwidth reservation scheme for handoff in mobile multimedia networks. In *IEEE International Conference on Universal Personal Communications*, volume 1, pages 311–315, 1998.

[76] M. M. Zonoozi and P. Dassanayake. User mobility modeling and characterization of mobility patterns. *IEEE Journal On Selected Areas in Communications*, 15(7):1239–1252, September 1997.

Author's Publications

A list of author's publications:

[1] J. Zhang, J. W. Mark and X. Shen. An adaptive handoff priority scheme for wireless MC-CDMA cellular networks supporting multimedia applications. Accepted by *IEEE GLOBECOM 2004*.

[2] J. Zhang, J. W. Mark and X. Shen. An adaptive handoff priority scheme for wireless MC-CDMA cellular networks supporting realtime multimedia applications. To appear in the *Computer Communications Journal, special issue on Performance Issues of Wireless LANs, PANs, and Ad Hoc Networks.*

[3] J. Zhang, J. W. Mark and X. Shen. An adaptive resource reservation strategy for handoff in wireless CDMA cellular networks. *Canadian Journal of Electrical and Computer Engineering, IEEE Canada*, Vol. 29, no. 1/2, pp. 77-83, January/April 2004.

[4] J. Zhang, J. W. Mark and X. Shen. Joint packet- and call-level soft handoff in CDMA wireless cellular networks. In *14th IEEE proceeding of Personal, Indoor and Mobile Radio Communications, IEEE PIMRC*, vol. 2, pp. 2485-2489, September 2003.

[5] J. Zhang, J. W. Mark and X. Shen. A novel resource reservation scheme for handoff in CDMA wireless cellular networks. In *proceeding of IEEE Wireless Communications and Networking Conference, IEEE WCNC*, vol. 3, pp. 2069-2074, March 2003.

[6] J. Zhang, J. W. Mark and X. Shen. Adaptive handoff priority scheme for MC-CDMA multimedia wireless cellular networks. *2002 China Wireless Congress*, Hangzhou, China, 15-17 October 2002.