EVALUATION OF TRAFFIC PREDICTION BASED ACCESS CONTROL USING DIFFERENT VIDEO TRAFFIC MODELS IN 3G CDMA HIGH SPEED DATA NETWORKS

by

Yat Hong Chan
Bachelor of Applied Science in Computer Engineering
University of Waterloo 2000

MASTER OF ENGINEERING PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINERRING

In the
School of Engineering Science

SIMON FRASER UNIVERSITY

Spring 2006

APPROVAL

Name: **Yat Hong Chan**

Degree: **Master of Engineering**

Title of Project: **Evaluation of Traffic Prediction Based Access Control Using Different Video Traffic Models in 3G CDMA High Speed Data Network**

Examining Committee:

Chair: **Dr. Ivan Bajic**
Assistant Professor, School of Engineering Science, Simon Fraser University

Dr. Stephen Hardy
Senior Supervisor
Professor, School of Engineering Science, Simon Fraser University

Dr. Tejinder Randhawa
Supervisor
Adjunct Professor, School of Engineering Science, Simon Fraser University

Date Defended/Approved: Feb. 21/06

ABSTRACT

The evolution of 3G Code Division Multiple Access (CDMA) network towards higher data rates is through the introduction of High Speed Downlink Packet Access (HSDPA) enhancement to the existing 3rd Generation Partnership Project (3GPP) standards. In this paper, an access control protocol is proposed for an integrated voice, video and non real-time data traffic on the forward link (cell-site to mobile). The protocol involves predicting the residual capacity available for the HSDPA traffic. This paper evaluated the performance of three video traffic models in predicting the number of data packets that could be scheduled at the next time slot. All three video traffic models exploit the frame properties of Motion Picture Experts Group (MPEG) traffic. The traffic models are based on Markovian, Autoregressive (AR) and two-sided Markov Renewal Model (TSMR) processes. Simulation results are obtained using Network Simulator 2 (ns2) with the Enhanced UMTS Radio Access Network Extensions (EURANE) using real video traces. The performance of the proposed estimation schemes are compared with estimation scheme using static guard margin. Findings of this paper can be used to improve the downlink performance of non-real time data traffic in the presence of MPEG video traffic in 3G CDMA networks.

ACKNOWLEGMENT

I would like to express my gratitude to Dr. Stephen Hardy and Dr. Tejinder Randhawa for their valuable guidance, supervision and support throughout my graduate studies. It has been an honour for me to be a member of the New Media Innovation Center (Newmic) affiliated with Simon Fraser University.

I would also like to thank Dr. Ivan Bajic for being the reader of this report and for his valuable advices and comments.

I would like to thank my employer for the financial support and my colleagues who had given their support and advice that are essential to the successful completion of the project.

And finally, I want to express my gratitude to my family and friends, especially Patricia Ma for their support and encouragement. Without them, I would not be able to accomplish so much in this research project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Acronyms	Description
ACF	Autocorrelation Function
AR	Autoregressive
ARQ	Automatic Repeat Request
AM	Acknowledged Mode
BER	Bit Error Rate
BLER	Block Error Rate
BS	Base Station
CBR	Constant Bit Rate
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CQI	Channel Quality Indicator
CRC	Cyclic Redundancy Check
DCH	Dedicated Channel
DCT	Discrete Cosine Transform
DL	Downlink
DSCH	Downlink Shared Channel
FACH	Forward Access Channel
FDD	Frequency Division Duplex
FER	Frame Error Rate
GoP	Group of Pictures
HARQ	Hybrid Automatic Repeat Request
HO	Handover
HSDPA	High Speed Downlink Packet Access
HS-DSCH	High Speed DSCH
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
Iub	Interface between Node B and RNC
LRD	Long Range Dependence
MAC	Medium Access Control
MPEG	Motion Picture Expert Group
MSE	Mean Square Error
NS2	Network Simulator 2
OTCL	Object Tcl
PC	Power Control
PDU	Protocol Data Unit
PHY	Physical Network Layer
QoS	Quality of Service
RLC	Radio Link Control

RNC	Radio Network Controller
RT	Real Time
RTT	Round Trip Time
SDU	Service Data Unit
SRD	Short Range Dependence
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
STD	Standard Deviation
TCL	Tool Command Language
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TES	Transform-Expand Sample
TSMR	Two-Sided Markov Renewal
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunication System
UTRAN	UMTS Terrestrial Radio Access Network
VBR	Variable Bit Rate
WCDMA	Wideband CDMA

1 INTRODUCTION

In the past few years, 3G CDMA networks have gained substantial growth despite setbacks due to technical difficulties in early deployment. According to reports from Global Mobile Suppliers Association [1] and the CDMA Development Group [2], there are over 200 million subscribers of WCDMA and CDMA2000 combined worldwide and the rate of 3G adoption is accelerating. It is forecast that, in a few years, the subscribers of 3G networks will overtake those of the 2G networks. 3G networks have already gained enormous popularity in Asia and Europe, where North America is following up closely. The attraction of 3G networks is not only provision of higher voice traffic capacity solving the problem of the already congested 2G networks; but they also introduce the possibility of new types of services such as video over cellular networks and high-speed data access. The integration of voice, video and data services creates technical challenges to the cellular service providers as well as opportunities to expand existing business in the saturated voice market.

There are many other proposals for providing triple play services using wireless connection. However, most of them are still in drawing board, while the few with working prototypes have limited mobility and reliability compared to the cellular network. Using the cellular network for universal wireless access has a clear advantage. It can leverage the network infrastructure built over the years that provides complete coverage in most of the populated areas. The technology has proven itself in trial demonstration of early adopters such as South Korea and Hong Kong. In the near future, the cellular network is the obvious choice for the consumers in mobile wireless access with integrated voice, video and data services.

Although 3G networks provide much higher capacity, users have found ways to utilise the newly added capacity in the system. The scarcity of the available radio frequency spectrum is always the major limiting factor in the system. Therefore, efficient allocation of the bandwidth among users and different types of services is the key to improve the network performance. Unlike 2G networks, which mostly carry homogeneous voice traffic, 3G networks carry various type of traffic with different quality of service (QoS) requirements. Voice and streaming video traffic are very delay sensitive and require delivery in real time. The bandwidth requirement of voice traffic is well understood and is assumed statistically stable after aggregation. The bandwidth requirement of voice traffic can vary and depends on the content. Data services such as text messages, web browsing, music downloads have less stringent delay, thus they are classified as non-real time traffic, which can be delivered with a lower priority in the system. Therefore finding the optimal balance in bandwidth allocation between real time and non-real time traffic is very important in providing a reliable integrated service to the end users.

To achieve efficient bandwidth allocation between real time and non-real time traffic, the network has to apply an access control scheme to maintain the total system interference and power consumption within the operation limit. In previous research in CDMA systems, most proposed admission control schemes [3][4] are mainly focused on supporting voice and data services. Notable exceptions are the access control schemes based on video traffic prediction proposed in [5][6]. The idea is based on the recognition that the bandwidth of voice and video traffic fluctuates with statistical significance. By applying a traffic model to the real time traffic in the current time slot, the system can predict the residual capacity in the next time slot, thus it can optimize the scheduling of non-real time data transmission. The above two proposals are based on 2G IS-95B CDMA network, using basic discrete-state continuous time Markov chain model on the video traffic model and focused only on the uplink channels.

In this thesis, the works in [5][6] are extended to apply the same access control scheme to 3G CDMA networks and with a focus on the downlink channels, as video and data traffic are mostly asymmetrical downlink traffic. Recent developments in the downlink high-speed data channel and the advantages of using the shared downlink channel are examined. This report will apply the proposed access control schemes to schedule data traffic in the high-speed shared downlink channel. This report also looks into recent developments in video traffic models, in order to search for better traffic models to replace the Markov chain used in the previous proposals to predict the residual capacity. In this report, three different types of traffic models exploring different statistical characteristics of MPEG video traffic are investigated and compared in simulation. They are Markov, Autoregression (AR) and Two-Sided Markov Renewal model (TSMR) based traffic models respectively.

This report consists of eight chapters. Chapters 2 to 5 give an overview of research works in integrated voice/video/data services in 3G networks, video encoding scheme and MPEG traffic models. Chapter 6 describes the proposed residual capacity estimation and access control scheme. In chapter 7, the simulation model is introduced using NS2 to simulate a single cell network. In chapter 8, the performance obtained by applying different traffic models in the prediction scheme is evaluated and analyzed using the simulation results. Finally, the conclusions and discussions of future work are presented in chapter 9.

2 BASICS OF 3G CDMA

Compared to the 2G cellular network that was designed to carry mostly voice traffic, the 3G network is capable of handling multi-media traffic in addition to voice calls. [7]

Figure 1 shows the basic architecture of 3G cellular network. The core network manages the billing and location information of the subscribers, setup connection between the radio system and the fixed networks. The mobile switch centre (MSC) handles the switching of voice calls between the plain old telephone systems (POTS) and the roaming mobile stations in the network. The internet gateway routes data and multi-media traffic between the mobile stations and remote servers in the internet alongside with voice traffic.

The radio system consists of radio network controllers (RNC), base stations (BS), and user equipment (UE). The RNC is responsible for the control of radio resources in the network. It controls one or many BS to handle power control, handover control, admission control, load control and packet scheduling. The BS is responsible for the air interface processing, which includes channel encoding, interleaving, rate adaptation and spreading, as well as soft handovers and inner closed-loop power control. The BS communicates with the UE over the radio interface. The radio traffic is terminated at the UE, which serves as the mobile station for the subscriber. An UE is not limited to cellular phone; it can also be 3G enabled personal digital assistance (PDA) or laptop.

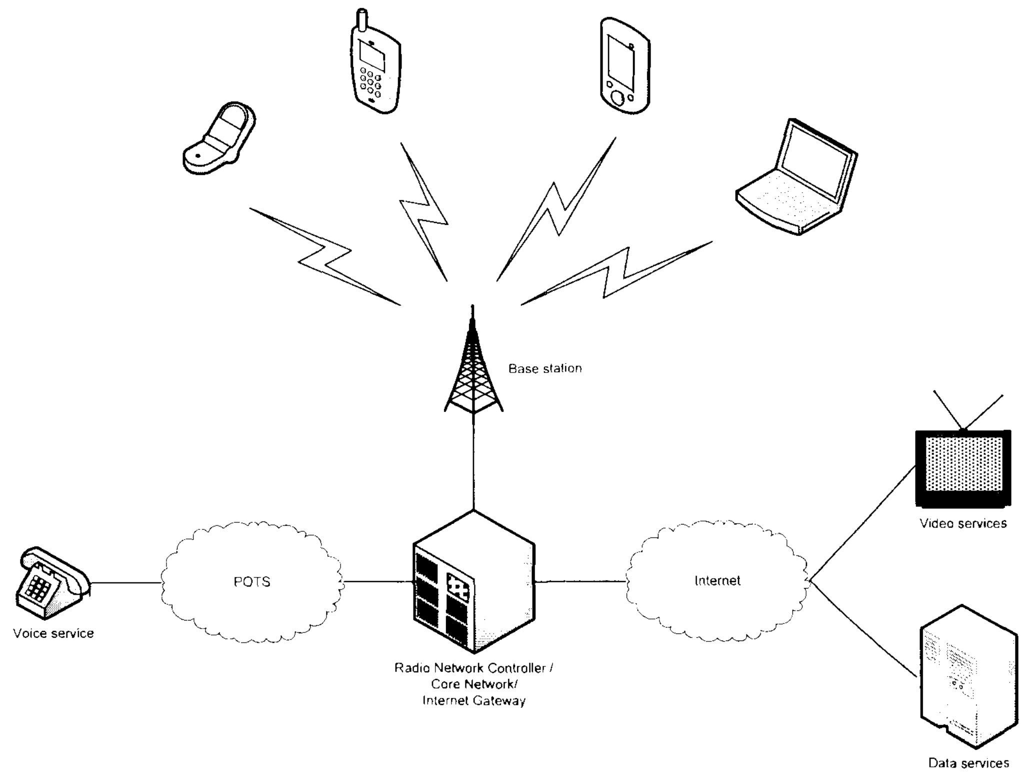

Figure 1 Basic 3G CDMA Network Architecture

The 3G network uses Code Division Multiple Access (CDMA) as the access scheme of the air interface. In CDMA, each user is assigned a spreading code to encode the user traffic. The receiver retrieves the data by using the same code sequence to decode the received signals. [8] Unlike 1G frequency division multiple access (FDMA) systems or 2G time division multiple access (TDMA) systems, CDMA system occupies the entire allocated spectrum, allowing a number of users to transmit at the same time without conflict. There is no absolute upper limit on the number of users a cell site can support in CDMA, which is known as soft capacity. There are two technical proposals for 3G network; they are known as wideband CDMA (WCDMA) by the European

Telecommunication Standards Institute (ETSI) of Europe and CDMA2000 by Telecommunications Industry Association (TIA) of USA [9]. In this paper, we focus on applying the traffic prediction based access control on WCDMA as there are more research papers available. Since the two standards support similar features with different physical layer implementations and different use of terminology, the findings of this paper can apply to both 3G proposals.

Earlier works in CDMA research assume the interference limited reverse link is the bottleneck of network performance. [9][10] However the new types of services introduced in 3G network, such as wireless internet and multi-media streaming, are expected to generate asymmetric traffic demand in favour of the downlink. Research in [14]-[17] had shown that the downlink performance is power limited, and there is a trade off between cell capacity and cell coverage. Therefore more efficient use of the maximum transmit power provided by the BS is very important to increase the throughput of the network.

3 HIGH SPEED DOWNLINK PACKET ACCESS (HSDPA)

Packet-based data service is one of the key elements of 3G networks and it is expected to contribute to a significant source of traffic load in the mobile cellular network. The characteristic of traffic carried by the data services is very different from the conventional voice service. Data services originated from the internet are often asymmetrical, bursty and are tolerant to latency. In the original WCDMA specification, there exist several types of downlink radio channels [19]; however, they are not optimized to carry data services. The forward access channel (FACH) lacks fast closed loop power control and is limited to carry small amounts of data. The dedicated channel (DCH) is not efficient for bursty and high data rate services due to slow channel reconfiguration. The downlink shared channel (DSCH) enhanced trunking efficiency by time-multiplexing different users with fast channel reconfiguration time and packet scheduling procedure, but have limited support for very high data rate. In order to address the deficiencies of the original WCDMA specification in carrying downlink data traffic, the high-speed downlink packet access (HSDPA) is proposed as an extension to DSCH in order to enhance WCDMA towards higher data rate [18][19]. The HSDPA applies fast link adaptation, fast hybrid ARQ, fast scheduling and best-effort packet data to get better performance on downlink data traffic [20]. Due to the advantages of HSDPA over other downlink channels, it will carry most of the data traffic in WCDMA networks in the future. Therefore, we choose to apply the traffic prediction model on HSDPA packet scheduling in this paper.

4 BASICS OF MPEG VIDEO ENCODING

The video encoding standards of the Motion Picture Experts Group (MPEG) of the International Standards Organization (ISO) form the de facto format for video traffic in the Internet [21]. The standard has gone through many revisions with new features and enhancements since it was introduced. The original MPEG-1 standard is designed for video storage and playback on CD. The MPEG-2 standard is designed to support digital broadcasting of television and was later adopted by the DVD forum. The H.261 standard was developed by ITU to support video conferencing using similar techniques as in MPEG-1. The H.263 standard improves the compression of H.261 for very low bit rate applications. The core compression techniques used in H.263 were later incorporated into the MPEG-4 standard. MPEG-4 has two different core codecs, the MPEG-4 Visual/Part 2 and MPEG-4 Part 10/H.264. The MPEG-4 Visual standard has a very flexible syntax to support encoding of natural and synthetic visual sequence with multiple layers of video objects. The H.264 standard is designed to supporting efficient and robust coding of rectangular natural video frames. The H.264 standard has been adopted by many cell phone makers as the format to carry video traffic over the cellular network. In this paper, we apply the traffic prediction model to H.264 video traces. Although the emphases of each MPEG coding scheme are different, all standards in the MPEG family are block-based compression schemes following the same principle. The findings in this paper are also relevant to other MPEG standards and commercial codecs derived from these standards, such as Real Video and Windows Media.

The main principle in MPEG video coding is exploiting the spatial, temporal and statistical redundancy in the video stream for compression. The MPEG codec uses lossy compression, where subjective redundant elements in the video are removed without significantly affecting the viewer's perception of visual quality. There are three frame types: intra-coded (I), inter-coded (P), and bi-directional coded (B) frames in a MPEG video sequence. The frames between two I frames form a Group of Pictures (GoP) structure as illustrated in Figure 2.

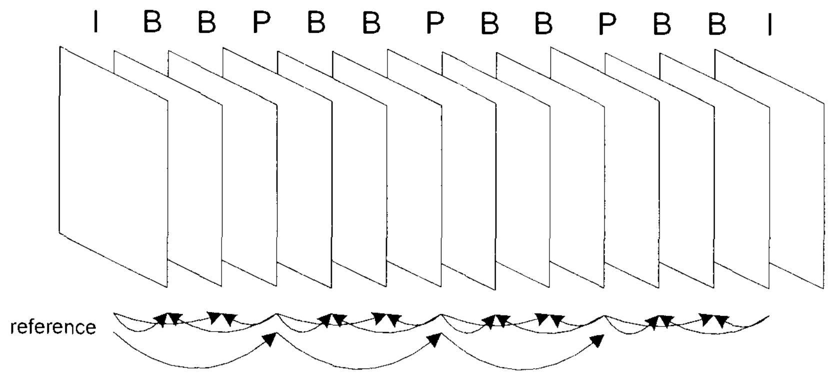

Figure 2 MPEG GoP Structure

The I frames are intra-frame coded using discrete cosine transform (DCT) and quantization to reduce the frame size. The P and B frames are inter-frame coded using motion estimation and motion compensation between successive video frames. Both P and B frames forward reference to preceding I or P frames, where the B frames also backward reference to succeeding I or P frames. The I, P and B frames form a GoP structure together with header information. The MPEG codec then uses arithmetic coding or variable-length coding to compress the combined GoP bit stream.

5 MPEG TRAFFIC MODELS

In both of the previous works [5][6] which this paper is based on, the aggregated video traffic is modelled as a discrete-state continuous time Markov birth-death process [22]. This video traffic model is one of the earliest works published in this area and is widely referenced. Since its publication, there have been many researches proposed video traffic models with better accuracy [23]-[25]. The proposed approaches for modelling video traffic can be roughly classified as follows:

a. Transform-Expand-Sample (TES) models

b. Non-linear adaptive neural network models

c. Markov-based models

d. Autoregressive (AR) models

e. Autoregressive Integrated Moving Average (ARIMA) models

f. Fractional Autoregressive Integrated Moving Average (FARIMA) models

g. Wavelet Models

Based on the accuracy and computation efficiency of the video traffic models presented in [23]-[25], this report chose to investigate the application of the following three video models to the traffic prediction based access control proposed in [5][6].

5.1 Markov-based Model

The first model uses a finite state Markov chain to model the MPEG video sequence.[26]-[28] The model isolates individual I, P and B frames into 3 sets of frame size data. Each set of data is represented by n states, where each state S quantizes the bandwidth requirement of each frame. The transition probability p_{ij} from S_i to S_j is estimated from the empirical data as follows:

$$p_{ij} = \left. \frac{N_{ij}}{N_i} \right|_{i,j=1..n} \quad (1)$$

where N_i is the total number of times that the system goes through state S_i, N_{ij} is the number of times that the system makes a transition to state S_i from S_j. Since a Markov chain has the memoryless property, the size of the next I, P, and B frame $Sz[k+1]$ is predicted using the previous I, P and B frame state $S[k]$ as follows:

$$Sz[k+1] = \left. \sum_j^n p_{ij} Sz_j \right|_{S_i = S[k]} \quad (2)$$

where Sz_j is the frame size at state S_j. The predicted frame size of the next I, P and B frame sequence is then combined following the underlying MPEG GOP sequence. (for example IBBPBB…)

5.2 Auto Regression (AR) Model

The second model uses second-order autoregressive (AR) process to model the MPEG video sequence [29][30]. The AR model can estimate the short-range dependence (SRD) nature in the autocorrelation function of the frame sequence. AR process is a linear system with input $\{s(t)\}$ and output $\{y(t)\}$, where t is the discrete time. The finite AR process is defined by

$$y(t) = \sum_{k=1}^{p} a_k y(t-k) + s(t) \tag{3}$$

where $\{s(t)\}$ is an uncorrelated process with zero mean and variance σ^2, and $\{a_k, 1 \leq k \leq p\}$ is a finite sequence with $a_p \neq 0$. Such a process is denoted by AR(p) and p is called the order of the AR process. There are a number of methods to estimate the parameters for an AR process given $\{y(t)\}$. In this paper, the parameters are estimated using the Yule-Walker estimation implemented in Matlab with $p = 2$.

In the AR model, the video sequence is split into I-frame, P-frame and B-frame sequences and each sequence is modelled independently. High correlation has been observed consistently in the split sequence due to adjacent frames tending to have similar scenes and amounts of motions. The frame size of the next frame in the sequence $Sz[k+1]$ is predicted as follows:

$$Sz[k+1] = a_1 Sz[k] + a_2 Sz[k-1] + \varepsilon(k) \tag{4}$$

where $\varepsilon(k)$ is the error function which models the frame fluctuation. In this paper, we choose $\varepsilon(k) = \sigma^2$, the variance of the frame size sequence, to give some extra margin to the AR model. Finally, the next frame size predictions of the I-frame, P-frame, and B-frame are combined together according to the GOP pattern.

5.3 Two-sided Markov Renewal (TSMR) Model

The third video traffic prediction model is called the two-sided Markov Renewal model (TSMR), which models the variation of the video traffic using a modified Markov-renewal process [31]. The Markov states in the process are classified into two groups: low-variation states correspond to small changes in adjacent frame size, and high-variation states correspond to a significant change in frame size. The difference in frame size within each Markov state is modelled to match both the autocorrelation structure and marginal distribution function. This model is used explicitly for prediction and a two-sided backward recurrence time series C_k is constructed as shown in Figure 3.

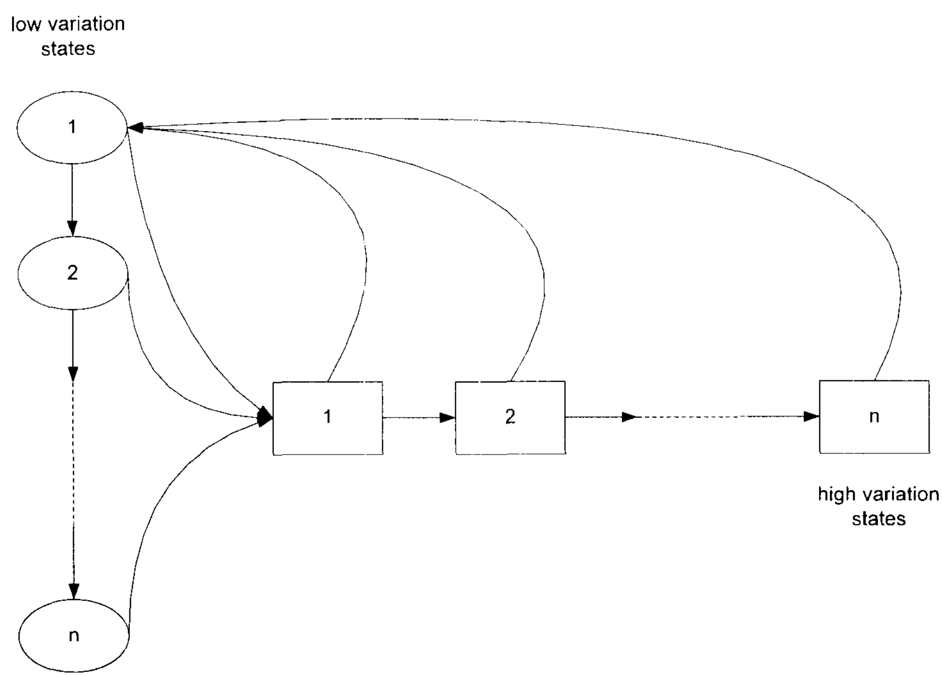

Figure 3 State transition diagram for the backward recurrence series

First, the sequence of difference in frame sizes in the video is calculated. Each value in the sequence is identified as a high-variation state if the absolute change in the adjacent frame size is larger than a predefined threshold; otherwise, the value is in low-variation state. Then the duration for the low-variation and high-variation states is counted from the two-sides of the TSMR sequence. The transition matrix **P** of the TSMR process is estimated from empirical data as follows:

$$\begin{aligned} P_{i,i+1} &= \frac{P(C_k \geq i+1)}{P(C_k \geq i)} \\ P_{i,1} &= 1 - P_{i,i+1} \end{aligned} \tag{5}$$

In forecasting the next state, the transition matrix **P** is referred to with the simplest prediction strategy, where the state with highest probability is selected. The difference between the current frame size and the next one is predicted and added back to the current frame size to generate the prediction of the next frame size. Since the frame differences are modelled as Gaussian i.i.d. process, they can be estimated by the sample mean of the frame difference sequence to minimize the mean square error (MSE) between the observed value and the forecast.

6 RESIDUAL CAPACITY ESTIMATION AND ACCESS CONTROL

6.1 Residual Capacity

In a CDMA system, the maximum downlink power the base station can transmit in a single cell is $\boldsymbol{W}$. We assume that that there are $\boldsymbol{K_{vo}}$ voice users, $\boldsymbol{K_{vd}}$ video users, and $\boldsymbol{K_d}$ data users accessing the forward link channel. The voice and video traffic are considered as real-time traffic and the data traffic is assumed non real-time traffic. Real-time traffic is delivered to the mobile station over the dedicated traffic channel with minimal delay. Non real-time traffic is delivered to the mobile station using the high-speed downlink shared channel. In a power limited CDMA downlink system, the transmission power assigned to K users is feasible if and only if

$$\sum_{j=1}^{K}(r_j\gamma_j) < W \tag{6}$$

where r_j and γ_j are the data rate and the required per bit transmit power of the j^{th} user. Assume the required per bit transmit powers for voice, video and data services are γ_{vo}, γ_{vd} and γ_d respectively. Then, the feasibility equation from above can be written as

$$\Gamma_d(n) \equiv d(n) < \frac{W - \sum \gamma_{vo} vo(n) - \sum \gamma_{vd} vd(n)}{\gamma_d} \tag{7}$$

where *vo(n)*, *vd(n)* and *d(n)* are the consumed bandwidths of active voice, video and data users who transmit at the n^{th} time slot respectively. $\Gamma_d(n)$ is defined as the ideal residual capacity for data users at the n^{th} time slot. An outage event happens when the above inequality is violated. When the required transmit power is higher than the maximum power the base station can provide, the transmit power is allocated to real-time traffic first, then the remaining transmit power is given to non real-time traffic. As a result, the data users will receive non real-time traffic below the required SNR that cannot be successfully decoded. The incorrectly decoded received data is not discarded in hybrid-ARQ with soft combining scheme implemented in HSDPA. Instead, the received signal is stored and soft combined with the later retransmissions of the same information bits. The combined signal effectively increases the received SNR, increasing the likelihood of a successful decoding of the information bits.

6.2 Static Residual Capacity Estimation

To reduce the outage event due to imperfect estimation, the estimated residual capacity $\Gamma_d(n+1)$ for data users at the $(n+1)^{th}$ time slot is commonly predicted less than the ideal residual capacity $\Gamma_d(n)$ at the n^{th} time slot as follows

$$\Gamma_d(n+1) = \Gamma_d(n) - \Delta(n) \tag{8}$$

where $\Delta(n)$ is called a guard margin at the n^{th} time slot.

In the static estimation scheme, the guard margin $\Delta(n)$ is statically set as a certain percentage of the maximum transmit power regardless of the time slot [32].

6.3 Dynamic Residual Capacity Estimation

In the dynamic estimation scheme, the guard margin $\Delta(n)$ is dynamically calculated based on the traffic load of voice and video services [32].

The voice traffic load at $(n+1)^{th}$ time slot is predicted with the voice traffic model presented in [5][6] . An ON/OFF voice activity model is used to model the voice traffic. In the ON state, the call utilizes a CDMA channel, and in the OFF state, no power is transmitted due to silence detection. The model assumes that the ON and OFF periods are independently and exponentially distributed with transition rate of μ (from ON to OFF) and λ (from OFF to ON). The voice source model is assumed to be a conventional discrete-time Markov process, as illustrated in Figure 4.

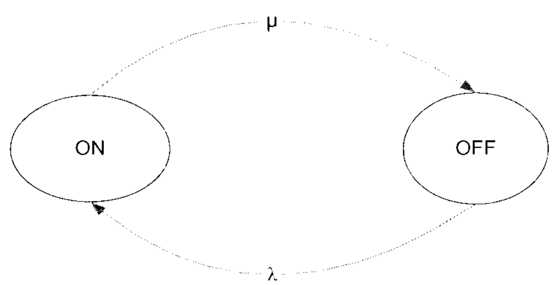

Figure 4 Voice Traffic Model

For $\boldsymbol{K_{vo}}$ voice users in the system and a$v(n)$ active voice users in the n^{th} time slot, the predicted voice traffic $vo(n+1)$ in the $(n+1)^{th}$ time slot is given by

$$vo(n+1) = \sum_{i=av(n),j=1}^{Kvo} P_{ij} R_v j \tag{9}$$

where P_{ij} are the transition probabilities of the Markov process and R_v is the data rate of voice packets per channel per time slot.

For the prediction of the video traffic load, we have chosen to implement and evaluate the three video traffic models introduced in chapter 5. For $\boldsymbol{K_{vd}}$ video users in the system. Then the predicted video traffic $vd(n+1)$ in the $(n+1)^{th}$ time slot is represented by

$$vd(n+1) = \sum_{j}^{K_{vd}} f(vd_j(n), vd_j(n-1) \ldots vd_j(0)) \tag{10}$$

where $f(.)$ is the prediction algorithm of the Markov model, AR model or TSMR model depending on the setup.

6.4 Access Control

In the HSDPA extension to the UMTS network, only up to one data user transmits a packet during the slot duration. Hence, the residual capacity, the data rate of data packet that could be scheduled at the $(n+1)^{th}$ time slot is estimated to be $\Gamma_d(n+1)$. When more than one data user is active in the same cell site, the base station will select the user to transmit based on the Channel Quality Indicator (CQI) of the mobile station using different scheduling schemes. To limit the variables in this research, we assume the users are uniformly distributed in the cell site and we assume perfect power control, so that the CQI of all data users are equal, therefore the data user in the next time slot is selected in a round-robin fashion.

7 SIMULATION MODEL

7.1 Simulation Setup

The simulation of a single cell WCDMA network is implemented in network simulator 2 (ns2)[33] with the Enhanced UMTS Radio Access Network Extensions (EURANE)[34]. The ns2 software is a popular academic research tool in the area of networking and communication. The software is a free open-source platform developed for simulation of communication networks and supports different network protocols and architectures via add-on modules mostly available in public domain. The ns2 version used in this project is version 2.28, which was the latest release when the project began. The ns2 software can be compiled to run on both Linux and Cygwin installed under Microsoft Windows XP environment.

The EURANE module for ns2 was developed within the Simulation of Enhanced UMTS Access and Core Network (SEACORN) project to evaluate the effects of proposed enhancements to the UMTS standard. The module was developed as a single cell WCDMA environment with support of the three functional nodes that make up the radio access network, User Equipment (UE), Base Station (BS) and the Radio Network Controller (RNC). The module supports Radio Link Control (RLC) protocol operating in Acknowledged Mode (AM) and Unacknowledged Mode (UM). The Medium Access Control (MAC) layer supports random access channel (RACH), forward access channel (FACH), dedicated channels (DCH) and the High-speed Downlink Shared Channel (HS-

DSCH) of High Speed Downlink Packet Access (HSDPA) extension to WCDMA. The EURANE version used in the project is release 1.09, which only worked with ns 2.26. This project contributes back to the ns2 community in two ways. We ported and released a patch of EURANE 1.09 to work with ns 2.28. We also submitted bug fixes to EURANE that were later incorporated into releases 1.10 and 1.11.

Figure 1 shows the simulation environment implemented with ns2 and the EURANE extension.

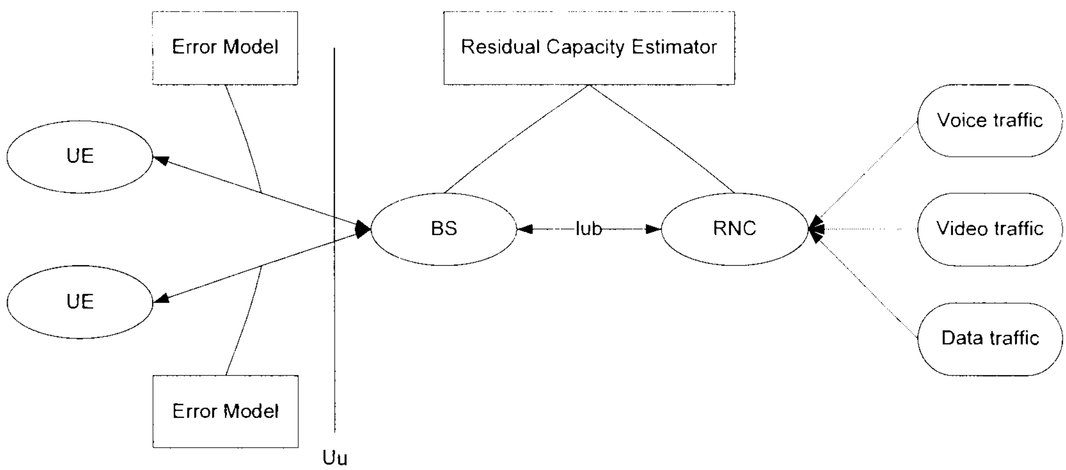

Figure 5 UMTS nodes in ns2

The simulation is setup to measure the performance within the cell site. Therefore the voice, video and data traffic are sourced into RNC directly to minimize unnecessary delay over the wire network. In ns2, agents represent endpoints where network-layer packets are constructed or consumed. We have built two new traffic agents to represent real-time traffic and non-real-time traffic. The real-time traffic agent handles voice and video traffic, where the non-real-time traffic agent handles the data traffic. Both traffic agents collect the sequence number, send timestamp, and receive timestamp of the two packet types into simulation trace files respectively for post-process analysis of the simulation result. All traffic agents read from pre-generated traces to generate traffic in RNC and sink the traffic received in UE after capturing the simulation result of each packet. The real-time traffic trace captures the frame size in each time slot as well as the frame size prediction in the next time slot. The advantage of using pre-processed traffic and prediction traces over generating the traffic and prediction in real-time is that the implementation of simulation model is less complex and the simulation runs faster. The only drawback of this approach is that extra memory and disk space are required by the simulator to store the pre-generated traces for all the simulation runs.

We implemented the residual capacity estimator (RCE) on top of the HSDPA MAC of EURANE to schedule non-real time data traffic. The RCE monitors the real-time traffic entering the cell site, calculates the residual capacity using (7) with the help of frame size prediction sent by the real-time traffic agent. The RCE uses the access control schemes outlined in chapter 6 together with the calculated residual capacity to determine the bandwidth and transmission power allocated to non-real time data traffic in the next time slot.

The original error model of EURANE determines when to drop a frame in the time slot based on pre-generated Signal to Noise Ratio (SNR) trace of each UE. In our simulation, the focus is to investigate the improvement obtained from using video traffic model in estimating residual capacity available for non-real time traffic, thus we made some assumptions on the error model to limit the scope of the research. From chapter 2, we know that the uplink capacity is interference limited, unlike the downlink capacity which is power limited. Therefore, for the purpose of this research, we can assume there is perfect power control. As long as there is enough transmission power, the received signal at UE will meet the SNR requirement. The transmit power required at the BS is the minimum required received power at UE times the path loss. Since this research is not concerned with UE mobility and path propagation model, for simplicity we assume that for each traffic type the transmit power per bit is the same for all UE. In our implementation, the error model will monitor the total frame size of all real-time and non-real-time traffic in a time slot. It will indicate a frame lost in non-real-time traffic if the total power required to transmit the traffic in the current time slot is greater than the maximum power supported by the BS.

We use Matlab to pre-generate the traffic traces before the simulation. The voice traffic trace is generated based on the voice model described in chapter 6.3. The data traffic trace is generated based on two simulation scenarios. The first scenario models the user initiating a large file transfer; the second one models internet traffic of multiple users using Poison Pareto Burst Process (PPBP) [35][36]. The characteristics of the video traffic traces, generated based on real MPEG video traces found in [37], will be further elaborated upon in the next section.

7.2 Video Traffic Trace

To evaluate the performance of the selected video traffic model, the simulation uses trace sequence from H.264 movie MPEG stream "Starship Troopers". The video stream is generated at a rate of 25 frames per second in the QCIF format. The uncompressed video sequence is in the common YUV 4:2:0 format with screen size of 176 x 144 pixels. The MPEG encoder uses a GOP size of 12, so resulting video stream has sequences of IBBPBBPBBPBB frames. There are 90,000 frames in the trace file with a total video duration of 60 minutes. Table 1 lists the statistics of the selected video trace and Figure 6 illustrates a typical view of the frame size record.

Video Source	Starship Troopers
Format	QCIF
Encoding	H.264
Length	60 minutes
Quantization	31
Frames	90000
Total Size	18026929 bytes
Min frame size	10 bytes
Max frame size	4962 bytes
Mean frame size	200.3 bytes
Mean bit rate	40059.84 bit/s
Mean I-frame size	859.99 bytes
Mean P-frame size	248.7 bytes
Mean B-frame size	99.61 bytes

Table 1 Overview of frame statistics of the video trace

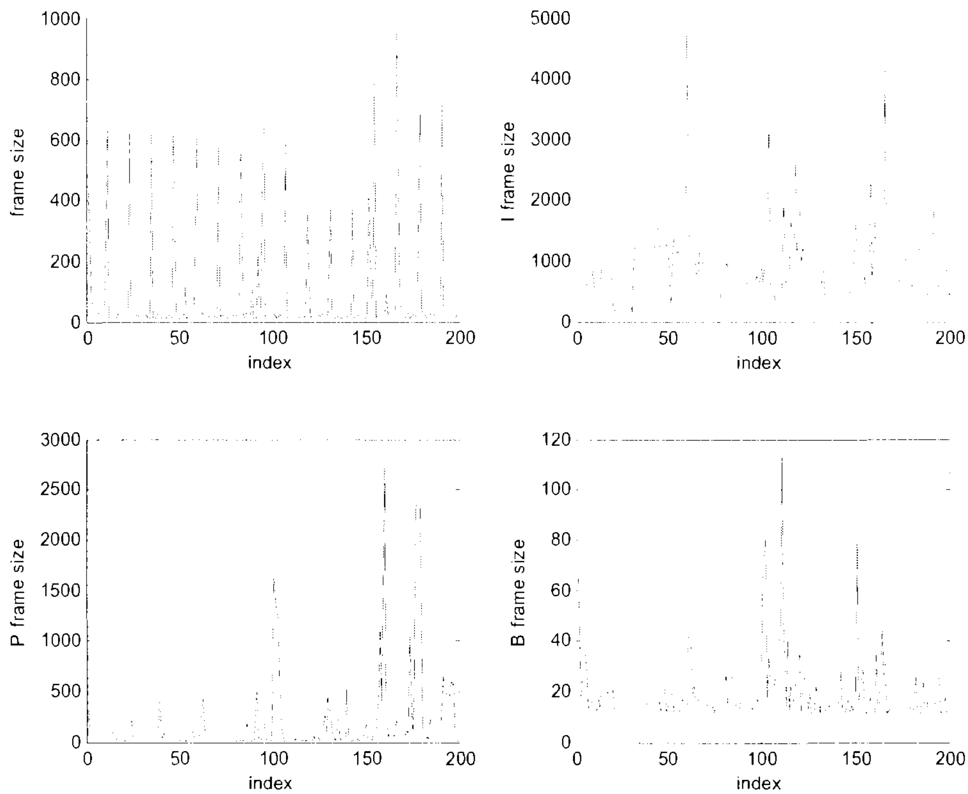

Figure 6 Frame size of the video sequence of Starship Troopers

The autocorrelation function (ACF) of the traffic trace is shown in Figure 7. The shape of the ACF of all frames follows the GOP pattern as expected. The ACF of the decomposed I-frame, P-frame and B-frame shows short lags exhibiting a relatively fast decay, which indicates that the traffic trace can be modelled by an AR model.

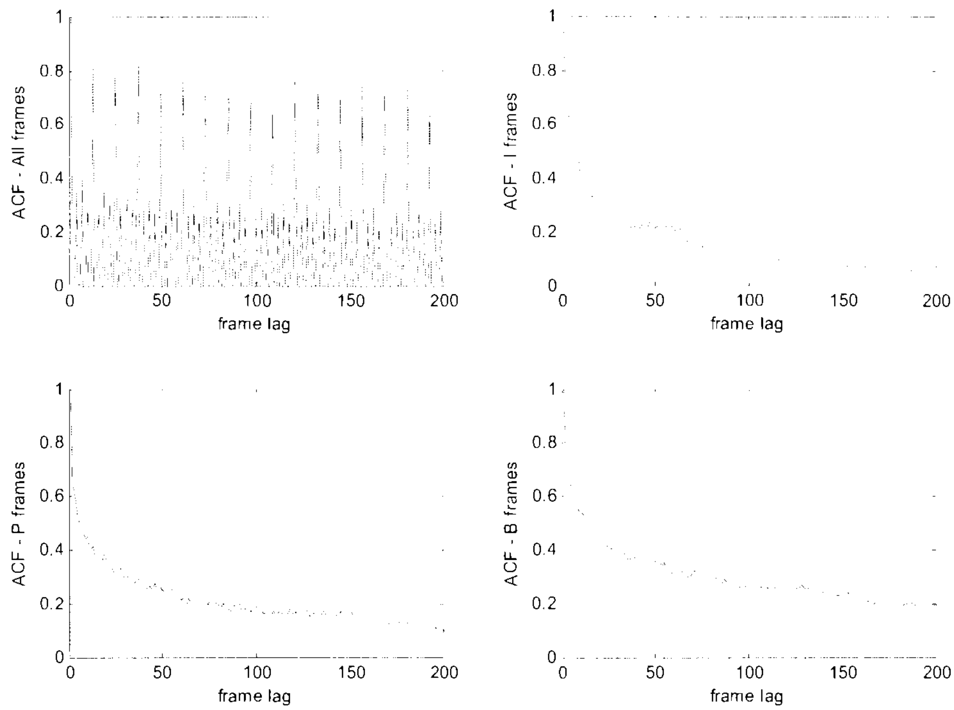

Figure 7 Autocorrelation function of the video trace

Figure 8 shows the phase portrait of lag one, derived from the size of I frames from the video traces. A phase portrait of lag k, is a plot of frame size at time n against frame size at time n-k for all values of n. The solid dots that fall outside the diagonal region in the diagram identify the high variations in the video. Although these variations look random in nature, they can be modelled by the TSMR process as described in chapter 5.3.

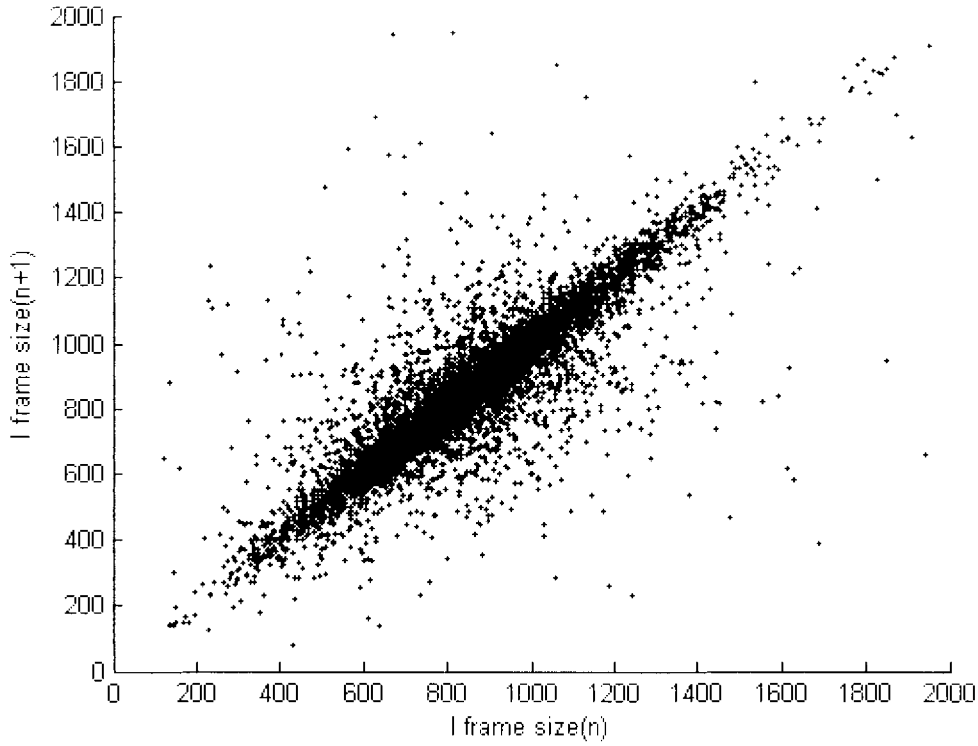

Figure 8 Phase Portrait of I frame size

8 RESULT ANALYSIS

In this section, we present a detailed performance evaluation of the access control schemes outlined in chapter 6 including: 1) static residual capacity estimation, 2) dynamic residual capacity estimation using a) Markov-based model, b) Auto Regression (AR) model, and c) Two-sided Markov Renewal (TSMR) model for video traffic. We also included dynamic residual capacity estimation using ideal video traffic prediction, which uses the exact frame size of the next frame, to give the upper bound performance of the access control schemes for reference. Table 2 shows the parameters used in the simulation.

Parameter	Value
TTI Length	20ms
Total BS Power	20W
Static Estimation Power Margin	1W
Maximum Bandwidth of the Cell	2.5Mbps
PDU size	40 bytes
Radio Link Control Protocol	AM
Iub delay	0.2ms
Effective Rate of 1st Retransmission	3/8
Effective Rate of 2nd Retransmission	¼
DCH Capacity	256 kbps
Voice Call Data Rate	9.6kbs
Voice Call Activity Factor	0.4
Scenario 1 File Size	1Mbyte
Scenario 2 Poisson Arrival Rate	0.5s
Scenario 2 Pareto Minimum Packet Size	2.5kbyte
Scenario 2 Pareto Parameter	1.5

Table 2 Simulation Parameters

The main problem considered in this paper is to investigate which access control scheme yields better network performance in the cell site. Accordingly, we use the measurements of the following two performance metrics in the simulation:

1. Packet delay or file transfer time: Time from sending a packet or a file to the RNC until the correct reception of the packet or file by the UE.

2. Percentage of retransmission: The number of retransmissions divided by the total number of PDUs sent in the HSDPA channel.

These metrics are shown as functions of real-time voice and video traffic loads in term of the number of voice and video users in the cell site. Each user's voice channel traffic is generated using the voice model described in chapter 6.3, and each user's video channel traffic is generated using the video traces described in chapter 7.2 beginning in an arbitrary frame number in the video sequence.

8.1 Scenario 1: Video Traffic with File Transfer

The first scenario simulates the cell site with only video users and data user downloading large files. The simulation results of 1 to 10 video users in the cell site are given in Figure 9 and Figure 10. The first observation is as the number of video users increase, the delay also increases due to more bandwidth being used up by the video traffic. The second observation is static residual capacity estimation has the worst performance, since it cannot anticipate the fluctuation in the bandwidth requirement of video traffic. Dynamic estimation using Markov model is slightly better than static estimation, but considerably worse than using prediction using the AR or TSMR models. The Markov model does not take the SRD nature of the video traffic into account, thus it tends to overestimate the residual capacity. As a result, there are more time slots exceeding the maximum power limit at the BS and causing higher retransmission percentage to the data traffic. The third observation is the performance of TSMR model and AR model are comparable. The AR model is slightly better in lower video loads, while the TSMR is slightly better in higher video loads. When there are more video users in the cell site, the retransmission percentage of the TSMR model decease as the aggregation of more video streams tends to smooth out the spikes in the frame size fluctuation.

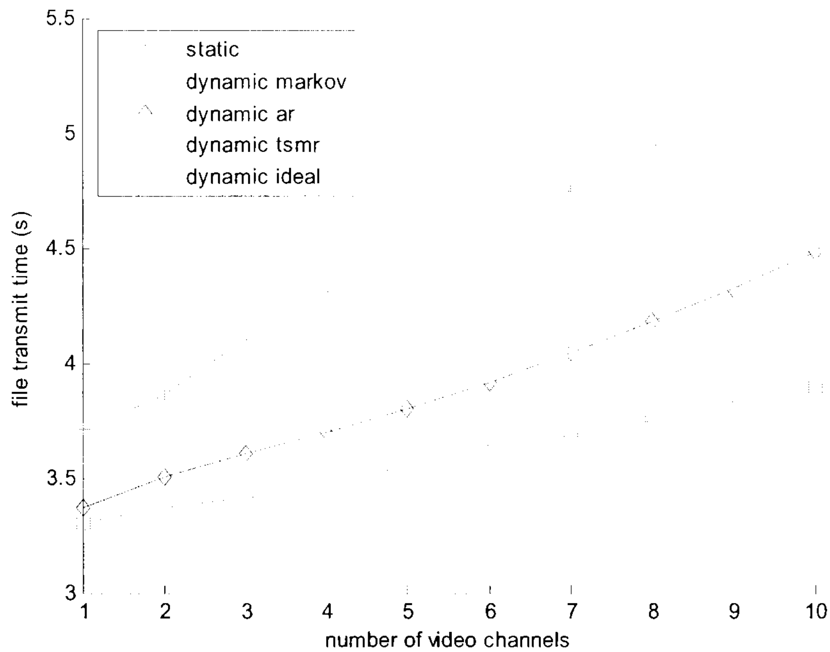

Figure 9 File transmit time versus number of video channels in scenario 1

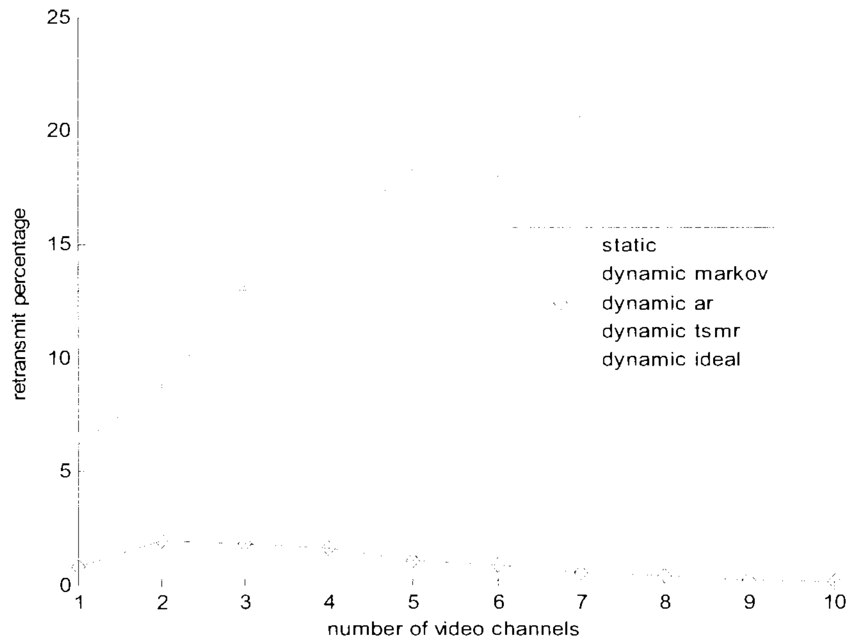

Figure 10 Retransmit percentage versus number of video channels in scenario 1

8.2 Scenario 2: Video and Internet Traffic

The second scenario simulates the cell site with only video traffic and data users browsing the Internet. The simulation results of 1 to 10 video users and 10 data users in the cell site are given in Figure 11 and Figure 12. The observation in this scenario is similar to the previous scenario. Static estimation has the worst performance, while the dynamic estimation with TSMR and AR model is significantly better. Unlike scenario 1, internet traffic is bursty in nature with low bandwidth consumption between the data bursts. Thus, the HSDPA channel buffer is not always full. Therefore, the more aggressive estimation by TSMR model cannot squeeze enough residual capacity to offset bandwidth lost to retransmission. As a result, the AR model has better performance.

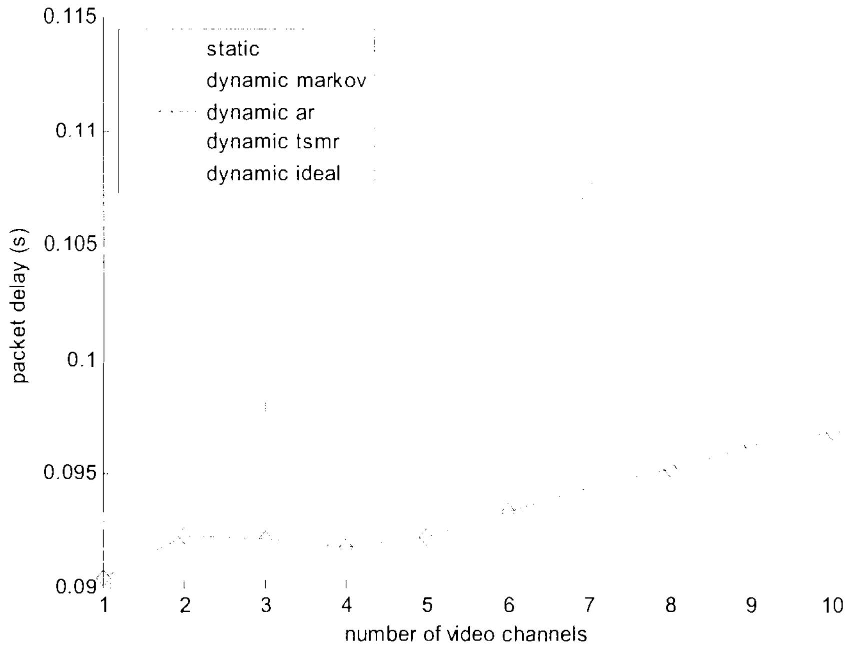

Figure 11 Packet delay versus number of video channels in scenario 2

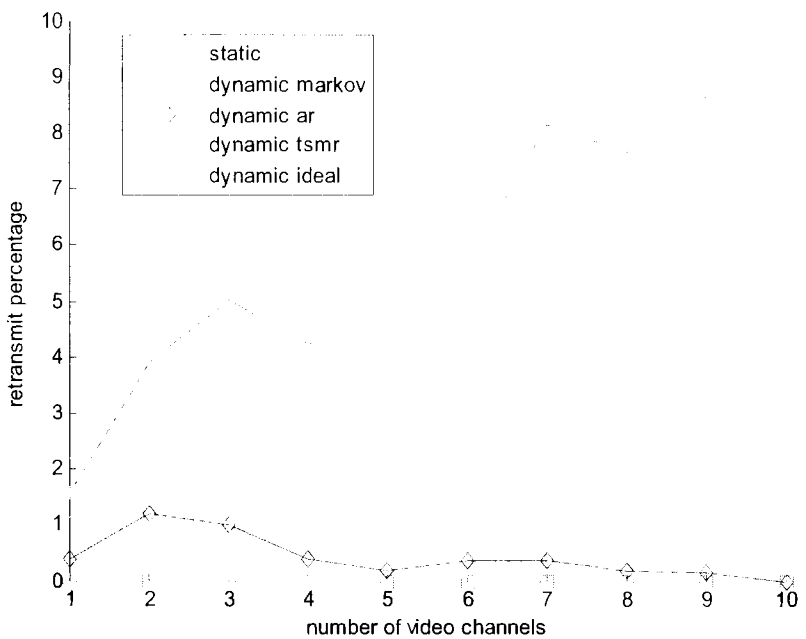

Figure 12 Retransmit percentage versus number of video channels in scenario 2

8.3 Scenario 3: Mixed Voice, Video and Internet Traffic

The third scenario simulates a cell site with mixed voice, video and data users browsing the internet. The simulation results of 0 to 20 voice users, 10 video users and 10 data users in the cell are given in Figure 13 and Figure 14. Both the delay and retransmission percentages stay mostly flat as the number of voice channels increase. The small amount of bandwidth consumed by the voice traffic can be neglected in the presence of bandwidth hungry video traffic. The performance of static estimation is worse than dynamic estimation, and prediction using the AR model is better than using the TSMR model, which agrees with the observation from the previous two scenarios.

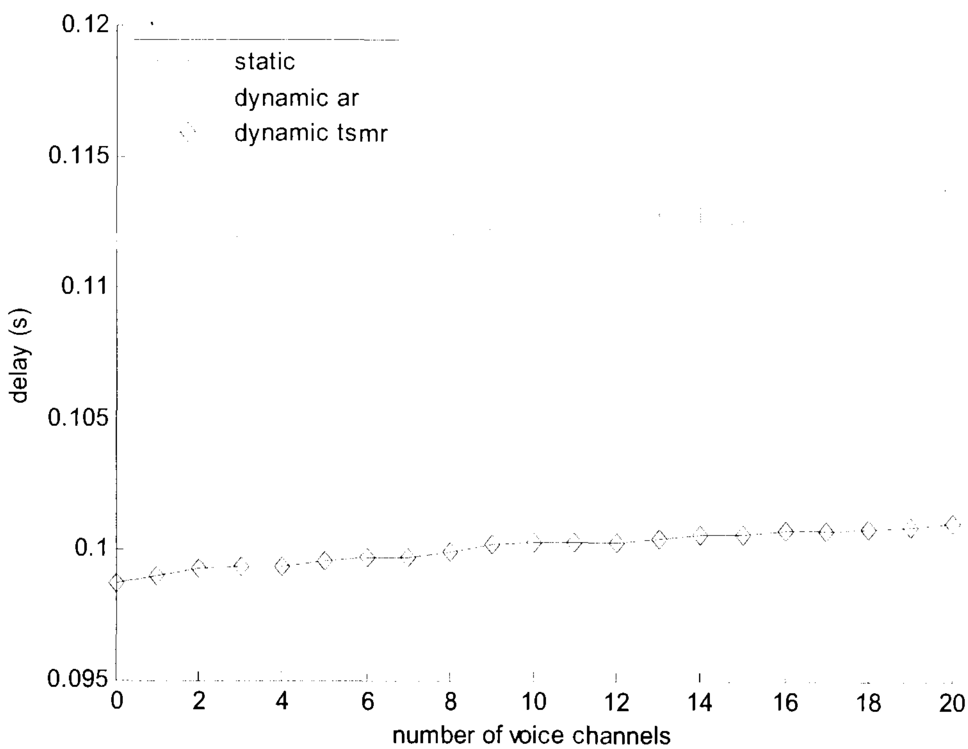

Figure 13 Packet delay versus number of voice channels in scenario 3

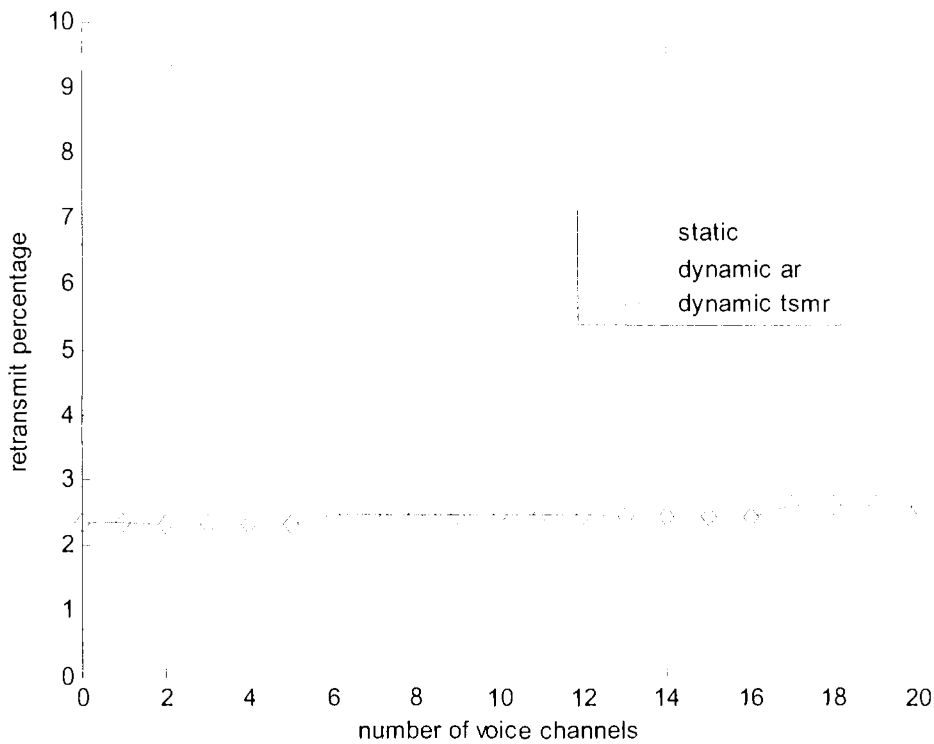

Figure 14 Retransmit percentage versus number of voice channels in scenario 3

9 CONCLUSION

In this thesis, we have investigated the application of MPEG traffic models to traffic prediction based access control in the forward link of WCDMA HSDPA channel. Simulation was used to evaluate the performance of static residual capacity estimation and dynamic residual capacity estimation using Markov, AR and TSMR video traffic model. It is concluded that dynamic estimation based access control outperforms static estimation based access control in integrated voice, video and data traffic in terms of data message delay and percentage of retransmission. Among the three video models, the AR and TSMR model are far superior to the Markov model. The TSMR model performs better in high video loads with HSDPA buffer full most of the time, whereas the AR model performs better with bursty internet traffic. When the video model is too aggressive in reclaiming the residual capacity, it may overestimate the available bandwidth for the data traffic. When this happens, the total traffic will exceed the transmission power at BS and bandwidth is wasted due to retransmission. Depending on the nature of non real time data traffic, in order to reduce the delay of data traffic, it has to strike a balance between scheduling more data into each time slot and preventing over scheduling in each time slot. It is also concluded that in the integrated voice, video and data CDMA system, we can neglect the voice channels in traffic prediction as video channels consume most of the bandwidth for real time traffic.

In order to improve the performance of non-real time traffic in CDMA network, future extensions to the work include the following areas:

1. Verify that the proposed traffic prediction based access control also works in the CDMA2000 systems with 1xEV-DV extension to cover both 3G standards.

2. Investigate the performance of traffic prediction using other MPEG models mentioned in chapter 5, such as the ARIMA model, FARIMA model, TES model, Wavelet model, etc.

3. Extend the traffic prediction based access control scheme to include a new class of real time traffic, video game traffic [38], in addition to the voice and video traffic. As playing on-line video games over the wireless network is getting more popular, in the near future it is predicted that traffic generated from video game applications will consume a significant amount of bandwidth in the CDMA network.

REFERENCES

[1] Global Mobile Supplier Association "Global Mobile Dialogue", *Issue 7, Nov 2005*

[2] CDMA Development Group "The 3G Evolution", *Oct 2005*

[3] A.Sampath, J.M.Holtzman, "Access Control of data in Integrated Voice/Data CDMA Systems: Benefits and Tradeoffs", *IEEE Journal on Selected Areas in Communications, Volume 15, Issue 8, Oct. 1997 Page(s):1511 – 1526*

[4] C.Comaniciu, N.B.Mandayam "Delta Modulation Based Prediction for Access Control in Integrated Voice/Data CDMA Systems", *IEEE Journal on Selected Areas in Communications, Volume 18, Issue 1, Jan. 2000 Page(s):112 -122*

[5] T.S.Randhawa, R.H.S.Hardy "Traffic Prediction based Access Control in Wireless CDMA Networks Supporting Integrated Services", *Wireless Communications and Networking Conference, Volume 3, 16-20 March 2003 Page(s):1561 - 1567 vol.3*

[6] J.W.So "Adaptive Traffic Prediction Based Access Control in Wireless CDMA System Supporting Integrated Voice/Data/Video Services", *IEEE Communications Letters, Volume 8, Issue 12, Dec. 2004 Page(s):703 – 705*

[7] J. Korhonen "Introduction to 3G Mobile Communications", *Artech House, 2001*

[8] A.J.Viterbi *"CDMA Principles of Spread Specturm Communication", Addison-Wesley, 1995*

[9] N.Parameshwar, R. Rajagopalan "A Comparative Study of CDMA2000 and W-CDMA", *Wireless Communications and Systems, 1999 Emerging Technologies Symposium,12-13 April 1999 Page(s):15.1 - 15.9*

[10] K.S.Gilhousen, I.M.Jacobs, "On the Capacity of a Cellular CDMA System" *IEEE Transactions on Vehicular Technology, Volume 40, Issue 2, May 1991 Page(s):303 - 312*

[11] W.C.Y.Lee, D.Lee "CDMA System Capacity Analysis", *Conference on Communications, Computers and signal Processing, 2001. PACRIM. 2001 IEEE Pacific Rim, Volume 1, 26-28 Aug. 2001 Page(s):22 - 26 vol.1*

[12] C. Cordie, S. Ortega "On WCDMA Downlink Multiservice Coverage and Capacity", *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th, Volume 4, 7-11 Oct. 2001 Page(s):2754 - 2758 vol.4*

[13] A. De Hoz, C. Cordier "W-CDMA Downlink Performance Analysis", *Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th, Volume 2, 19-22 Sept. 1999 Page(s):968 - 972 vol.2*

[14] L.Zhuge, V.O.K.Li "Forward Link Capacity in Multi-service DS-CDMA Systems", *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE, Volume 1, 25-29 Nov. 2001 Page(s):609 - 613 vol.1*

[15] W.Choi, J.Y.Kim "Forward-Link Capacity of a DS/CDMA System with Mixed Multirate Sources", *Vehicular Technology, IEEE Transactions on, Volume 50, Issue 3, May 2001 Page(s):737 - 749*

[16] H.Fu, J.S.Tompson "Downlink Capacity Analysis in 3GPP WCDMA Networks System", *3G Mobile Communication Technologies, 2002. Third International Conference on, 8-10 May 2002 Page(s):534 - 538*

[17] K.Hiltunen, R.D. Bernardi "WCDMA Downlink Capacity Estimation" *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st, Volume 2, 15-18 May 2000 Page(s):992 - 996 vol.2*

[18] T.E.Kolding, F.Frederiksen, "Performance aspects of WCDMA systems with high speed downlink packet access (HSDPA)", *Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th, Volume 1, 24-28 Sept. 2002 Page(s):477 - 481 vol.1*

[19] S.Parkvall, E.Dahlman "The High Speed Packet Data Evolution of WCDMA", *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on, Volume 2, 30 Sept.-3 Oct. 2001 Page(s):G-27 - G-31 vol.2*

[20] A.Furuskar, S.Parkvall "Performance of WCDMA High Speed Packet Data", *Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th, Volume 3, 6-9 May 2002 Page(s):1116 - 1120 vol.3*

[21] I.E. Richardson "H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia", *Wiley, 2003*

[22] B.Maglaris, D.Anastassiou "Performance Models of Statistical Multiplexing in Packet Video Communications", *Communications, IEEE Transactions on Volume 36, Issue 7, July 1988 Page(s):834 – 844*

[23] P.Makrov, H.Mehrpour "Statistical Modeling of MPEG Coded Video", *Networks, 2001. Proceedings. Ninth IEEE International Conference on 10-12 Oct. 2001 Page(s):38 - 43*

[24] O.Lazaro, D.Girma, J.Dunlop "A Study of Video Source Modelling for 3G Mobile Communication Systems", *3G Mobile Communication Technologies, 2000. First International Conference on (IEE Conf. Publ. No. 471), 27-29 March 2000 Page(s):461 - 465*

[25] A.D.Doulamis, N.D.Doulamis "Recursive Non Linear Models for On Line Traffic Prediction of VBR MPEG Coded Video Sources", *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, Volume 6, 24-27 July 2000 Page(s):114 - 119 vol.6*

[26] M.X.Li, A.D.Men, B.Z.Yuan "Traffic Model for MPEG compatible video service", *Signal Processing Proceedings, 1998. ICSP '98. 1998 Fourth International Conference on 12-16 Oct. 1998 Page(s):831 - 836 vol.1*

[27] W.Zhou, S.Ramakrishnan "Bandwidth Estimation for Multiplexed Videos Using MMG-based Single Video Traffic Model", *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE, Volume 6, 1-5 Dec. 2003 Page(s):3564 - 3568 vol.6*

[28] N.Ansari, H.Liu "On Modeling MPEG Video Traffics", *Broadcasting, IEEE Transactions on, Volume 48, Issue 4, Dec. 2002 Page(s):337 - 347*

[29] D.Liu, E.I.Sara "Nested Auto-Regressive Processes for MPEG-Encoded Video Traffic Modeling", *Circuits and Systems for Video Technology, IEEE Transactions on, Volume 11, Issue 2, Feb 2001 Page(s):169 - 183*

[30] Q.Zhang, C.Lin "An Accurate Scene-Based Traffic Model for MPEG Video Stream", *Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on, Volume 1, 14-17 Dec. 2003 Page(s):228 - 231 Vol.1*

[31] W.Lee, H.Mehrpour "TSMR – A New Statistical Model for MPEG-Coded Video", *Networks, 2002. ICON 2002. 10th IEEE International Conference on 27-30 Aug. 2002 Page(s):77 - 82*

[32] X.Wang, R.Ramjee "Adaptive and Predictive Downlink Resource Management in Next-Generation CDMA Networks", *Selected Areas in Communications, IEEE Journal on, Volume 23, Issue 6, June 2005 Page(s):1219 – 1232*

[33] K. Fall, K. Vardhan, "The ns Manual", *http://www.isi.edu/ns, Apr 2005*

[34] N.Whillans "End-to-end network model for Enhanced UMTS", *SEACORN, Oct 2003*

[35] M.Zukerman, T.D.Neame, R.G.Addie "Internet Traffic Modeling and Future Technology Implications", *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE Volume 1, 30 March-3 April 2003 Page(s):587 - 596 vol.1*

[36] P.Giacomazzi, L.Musumeci, G.Verticale "Performance of Web-Browsing Services over the WCDMA-FDD Downlink Shared Channel", *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE, Volume 6, 25-29 Nov. 2001 Page(s):3514 - 3518 vol.6*

[37] P.Seeling, M.Reisslein, B.Kulapala "Network Performance Evaluation Using Frame Size and Quality Traces of Single-Layer and Two-Layer Video: A Tutorial", *IEEE Communications Surveys and Tutorials., Volume 6, No. 3, Third Quarter 2004 Page(s):58-78*

[38] C.Heyaime, V.K.Prabhu "Traffic Models for Video Game Applications", *Broadband Communications for the Internet Era Symposium digest, 2001 IEEE Emerging Technologies Symposium on 10-11 Sept. 2001 Page(s):28 - 32*