Traffic Prediction Based Access Control Using Different Video Traffic Models in 3G CDMA High Speed Data Networks

Yat Hong Chan, Tejinder Randhawa, Stephen Hardy

School of Engineering Science
Simon Fraser University
Burnaby BC Canada V5A 1S6
horace_chan@pmc-sierra.com, {tsr, rhardy}@sfu.ca

***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 evaluates 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, Autogressive (AR) and two-sided Markov Renewal Model (TSMR) processes. The performances 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.**

***Index Terms*— Code division multi-access, access control, dynamic estimation, residual capacity, traffic prediction, voice/data/video services, MPEG**

I. Introduction

In 3G CDMA networks, 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 improving 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 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 requirements, 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 [1][2] in CDMA systems, most proposed admission control schemes are mainly focused on supporting voice and data services. Notable exceptions are the access control schemes based on video traffic prediction proposed in [1][2]. The idea is based on the statistically significant fluctuation of bandwidth of voice and video traffic. By applying a traffic model to the real time traffic in the current time slot, the system can predict the residue 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 to model the video traffic and focus only on the uplink channels.

In this paper, the work in [1][2] is extended to apply the same access control scheme to 3G CDMA networks and focus on the downlink channels, since video and data traffic are mostly asymmetrical downlink traffic. Recent developments in the downlink high-speed packet access (HSDPA) of WCDMA and the advantage of using the shared downlink channel are considered. This paper applies the proposed access control schemes to schedule data traffic in the high-speed shared downlink channel. This paper also looks into recent development on video traffic models, to replace the Markov chain traffic model used in the previous proposals for residual capacity prediction. In this paper, three different types of traffic models exploring different statistical characteristic aspects of MPEG video traffic are investigated and compared in simulation. They are markov, Autoregression (AR) and two-sided markov renewal model (TSMR) based traffic model respectively. The performance of applying different traffic mode in the prediction scheme are evaluated and analyzed with the simulation result, followed by conclusions highlighting the main contributions of this work.

II. Traffic Models

A. *Voice Traffic*

An ON/OFF voice activity model is used to model the voice traffic.[1][2] In the ON state, the call utilizes a CDMA channel, and in the OFF state, no power is transmitted due to silence detection. We assume 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.

B. *Video Traffic*

In both of the previous works [3][4] on which this paper is based, the aggregated video traffic is modeled as a discrete-state continuous time Markov birth-death process. [5] 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 proposed video traffic models with better accuracy. This paper chose to investigate the application of the following three video models in traffic prediction based access control.

1) *Markov-based Model*

The first model uses a finite state Markov chain to model the MPEG video sequence.[6]-[8] 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:

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

where N_i is the total number of times that the system goes through states S_i, N_{ij} is the number of times that the system makes a transition to state S_j from S_i. 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:

$$Sz[k+1] = \left.\sum_{j}^{n} p_{ij} Sz_j\right|_{S_i = S[k]} \tag{2}$$

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.

2) *Auto Regression (AR) Model*

The second model uses second-order autoregressive (AR) process to model the MPEG video sequence [10]. 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 define 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 $\{a_k,\ 1 \le k \le p\}$ is a finite sequence with $a_p \neq 0$. Such a process is denoated 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 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 modeled independently. High correlation has been observed consistently in the split sequence since adjacent frames tend to have similar scenes and amounts of motion. 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 modeling the frame fluctuation. In this paper, we choose $\varepsilon(k) = \sigma^2$, the variance of the frame sequence, to give some extra margin to the AR model. Finally, the next frame size prediction of the I-frame, P-frame, and B-frame are combined together according to the GOP pattern.

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 [11]. 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 modeled 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 Fig. 1.

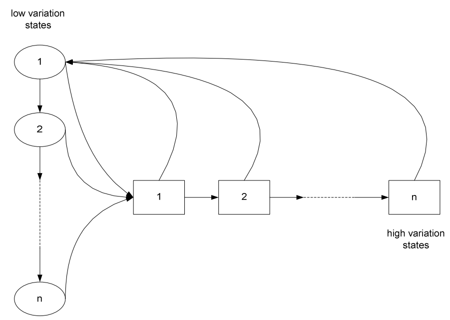


Fig. 1. State transition diagram for the backward recurrence series

First, the sequence of difference in frame sizes in the video is calculated. Each value in the sequence is identified into 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:

$$P_{i,i+1} = \frac{P(C_k \geq i+1)}{P(C_k \geq i)} \qquad (5)$$
$$P_{i,1} = 1 - P_{i,i+1}$$

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 difference is modeled as Gaussian i.i.d. process, it 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.

III. Residual Capacity Estimation and Access Control

In CDMA system, the maximum downlink power the base station can transmit in a single cell is ***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

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

where r_j and γ_j are the data rate and the required per bit transmit power of the jth user. Assume the required per bit transmit power 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

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

where *vo(n)*, *vd(n)* and *d(n)* are the consumed bandwidths of active voice, video and data users who transmit at the nth time slot respectively. $\Gamma_d(n)$ is defined as the ideal residual capacity for data users at the nth time slot. An outage event happens when the above inequality is violated. When the required transmit power is higher than the maximum power the antenna 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 successful 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.

1) *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 nth time slot as follows

$$\Gamma_d(n+1) = \Gamma_d(n) - \Delta(n) \qquad (8)$$

where $\Delta(n)$ is called a guard margin at the nth 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 [12].

2) *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 [12].

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

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

where P_{ij} are the transition probabilities of the Markov process and R_v is the data rate of voice packet 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 section 2. Assume $\boldsymbol{K_{vd}}$ video users in the system. Then the predicted video traffic *vd(n+1)* in the (n+1)th time slot is represented by

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

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

3) *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 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.

IV. Simulation Results

In this section, we present a detailed performance evaluation of the access control schemes proposed including: 1) statistic 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) models for video traffic. We also include 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 I shows the parameters used in the simulation and Table II lists the statistics of the selected video trace used in the simulation.

The main problem considered in this paper is to investigate which access control scheme and traffic model yield 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.

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 Fig. 1a and Fig. 1b. The first observation is as the number of video users increases, 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.

TABLE I
SIMULATION PARAMETERS

Parameters	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	9.6 kbps
Voice Call Activity Factor	0.4
Scenario 1 File Size	1M bytes
Scenario 2 Poisson Arrival Rate	0.5s
Scenario 2 Pareto Minimum Packet Size	2.5 kbytes
Scenario 2 Pareto Parameter	1.5

TABLE II
OVERVIEW OF FRAME STATISTICS OF THE VIDEO TRACE

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 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

The third observation is the performances 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 decreases as the aggregation of more video streams tends to smooth out the spikes in the frame size fluctuation.

2) *Scenario 2: Video and Internet Traffic*

The second scenario simulates the cell site with only video traffic and data users browsing the internet, which is modeled using the Poison Pareto Burst Process [13]. The simulation results of 1 to 10 video users and 10 data users in the cell site are given in Fig. 1c and Fig. 1d. The observations in this scenario are 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.

3) *Scenario 3: Mixed Voice, Video and Internet Traffic*

The third scenario simulates 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 Fig. 1e and Fig. 1f. Both the packet delay and retransmissions percentage stay mostly flat as the number of voice channels increases. 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.

V. Conclusion

In this paper, we have investigated the application of MPEG traffic models to traffic prediction based access control in the forward link of WCDMA HSDPA channel. Simulation is setup 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 term of data message delay and retransmission percentage. Among the three video models, the AR and TSMR model are superior to the Markov model. The TSMR model performs better in high video loads with HSDPA buffer full most of the time, where 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 and cause the total traffic to 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.

References

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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

[6] 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

[7] 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

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

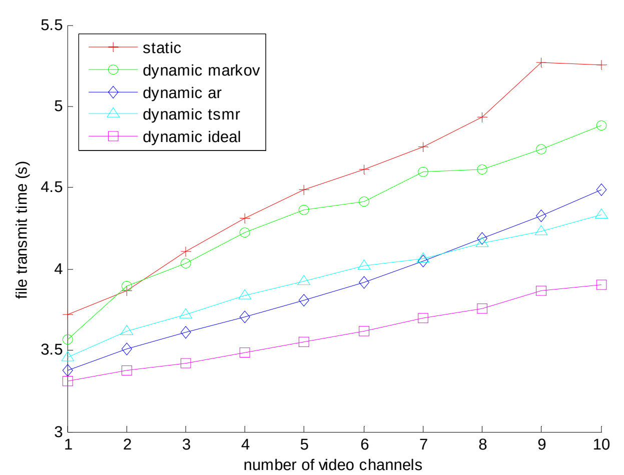
[9] D.Liu, E.I.Sara "Nested Auto-Regressive Processes for MPEG-Encoded Video Traffic Modeling", *Circuits and Systems for Video Technology, IEEE Transactions on*, Volume 11, Issue 2, Feb 2001 Page(s):169 - 183

[10] 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

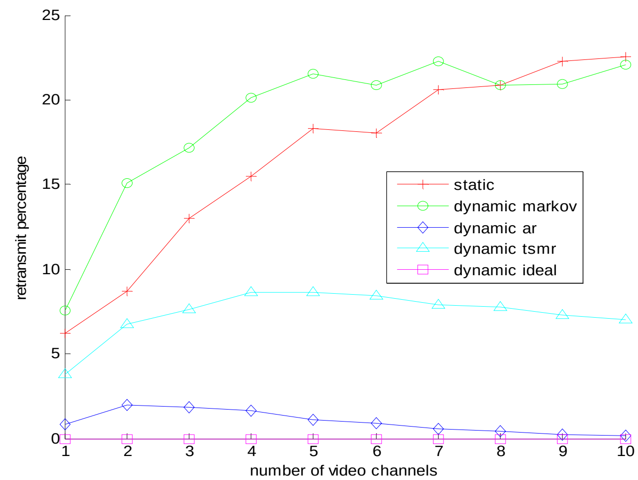
[11] W.Lee, H.Mehrpour "TSMR – A New Statistical Model for MPEG-Coded Video", *Networks*, 2002. ICON 2002. *10th IEEE International Conference on*, 27-30 Aug. 2002 Page(s):77 - 82

[12] 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

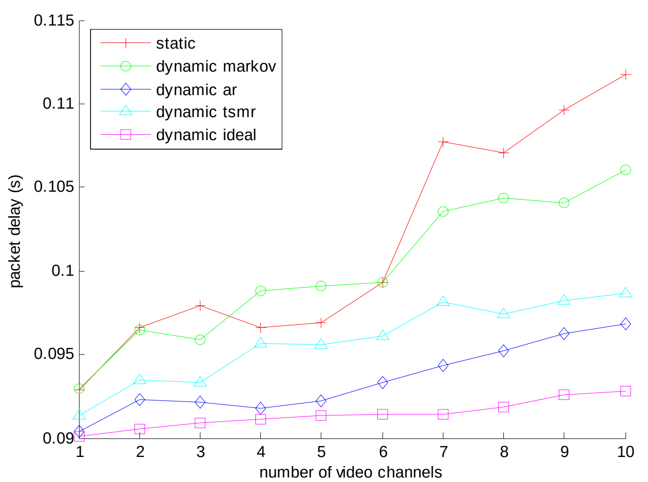
[13] M.Zukerman, T.D.Neame, R.G.Addie "Internet Traffic Modeling and Future Technology Implications", INFOCOM 2003. *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*.IEEE Volume 1, 30 March-3 April 2003 Page(s):587 - 596 vol.1



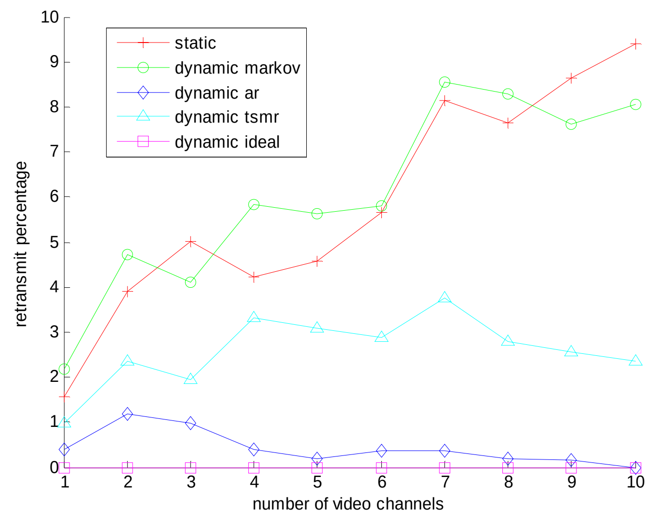
(a)



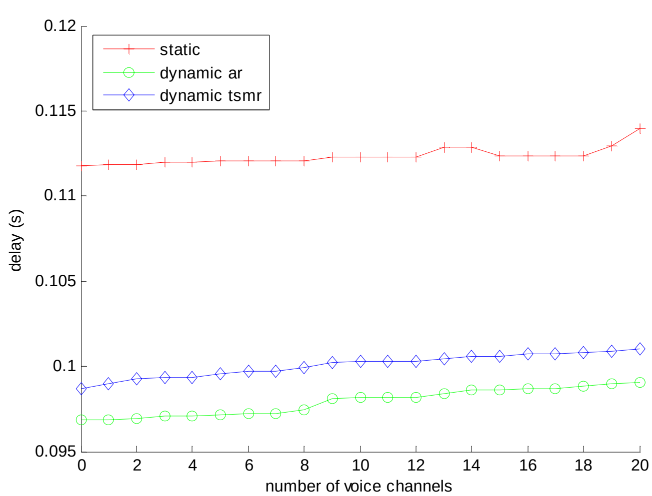
(b)



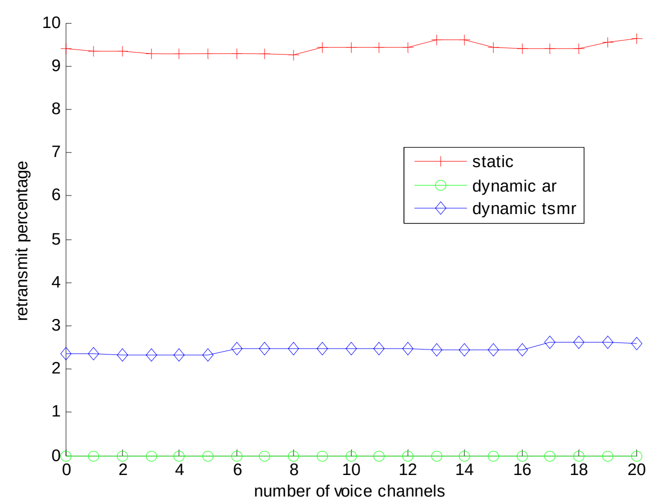
(c)



(d)



(e)



(f)

Fig. 1. (a) File transmission time vs number of video channels in scenario 1 (b) Retransmission percentage versus number of video channels in scenario 1 (c) Packet delay vs number of video channels in scenario 2 (d) Retransmission percentage versus number of video channels in scenario 2 (e) Packet delay vs number of voice channels in scenario 3 (d) Retransmission percentage versus number of voice channels in scenario 2